



Designing beyond the Product – Understanding Activity and User Experience in Ubiquitous Environments

European Conference on Cognitive Ergonomics 2009

30.9.–2.10.2009, Helsinki, Finland

ECCE 2009 – European Conference on Cognitive Ergonomics

Designing beyond the Product – Understanding Activity and User Experience in Ubiquitous Environments

Edited by

Leena Norros, Hanna Koskinen, Leena Salo & Paula Savioja

Organised by

VTT



ISBN 978-951-38-6339-5 (soft back ed.)

ISSN 0357-9387 (soft back ed.)

ISBN 978-951-38-6340-1 (URL: <http://www.vtt.fi/publications/index.jsp>)

ISSN 1455-0873 (URL: <http://www.vtt.fi/publications/index.jsp>)

Copyright © VTT 2009

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 3, PL 1000, 02044 VTT

puh. vaihde 020 722 111, faksi 020 722 4374

VTT, Bergsmansvägen 3, PB 1000, 02044 VTT

tel. växel 020 722 111, fax 020 722 4374

VTT Technical Research Centre of Finland

Vuorimiehentie 3, P.O. Box 1000, FI-02044 VTT, Finland

phone internat. +358 20 722 111, fax + 358 20 722 4374

Revised electronic version, November 2009.

Edita Prima Oy, Helsinki 2009

Preface

The European Conference on Cognitive Ergonomics is a yearly activity of the European Association of Cognitive Ergonomics. The 2009 meeting is hosted by VTT Technical Research Centre of Finland. ECCE 2009 focuses on the topic “Designing beyond the product – Understanding activity and user experience in ubiquitous environments”. The topic challenges the cognitive ergonomics community in several respects.

The first challenge is the focus on design. While we may agree that human factors / ergonomics is basically a science of design, the full meaning and consequences of this characterisation of the role of our discipline needs thinking and clarification. Attention should then be devoted on issues concerning the nature of knowledge we consider valid, the methods we use, the way we understand the relationship of human and technology, and also the societal purposes for which our knowledge is applied.

A further challenge is that the target of cognitive ergonomic design is widening from singular products to systems, environments, and services. This is much due to the advanced possibilities to apply information technology in these systems and environments. It is to be expected that our notions concerning cognitive phenomena and human action will be focused towards the environment and the distributed systems in which people are involved. At the same time the design of technologies needs understanding of the neural and physical bases of human behaviour.

Finally, we may notice that one of the fundamental topics of philosophy and psychology, “experience”, has taken its place in everyday design vocabulary. It is an interesting challenge for the cognitive ergonomics community to support in deepening our joint understanding of this notion in design. All these issues are discussed extensively by our knowledgeable keynote speakers and discussed by invited experts from the industry. The presentations of the conference participants will in more specific and detailed ways elaborate the themes and bring up new ones. We hope that the scientific and social yield of the congress will be abundant and that it fulfils the expectations of our colleagues and guests.

I wish to thank all the members of the international programme committee and the organisers of the doctoral consortium for their professional work and kind help in designing the congress programme. The local organising committee at VTT deserves a special thank for its great efforts in taking care of the practical realization of the conference and the proceedings. Our final thanks go to VTT and our sponsors whose support has made the conference possible.

September 2009

Leena Norros

President of the ECCE2009 Conference

International Program Committee

Roland Akselsson, Sweden
Carl-Martin Allwood, Sweden
Françoise Anceaux, France
Natalie Bonnardel, France
Guy Boy, USA
Willem-Paul Brinkman, The Netherlands
Djose Juan Canas, Spain
Francoise Detienne, France
Vincent Grosjean, France
Gudela Grote, Switzerland
Geert de Haan, the Netherlands
Austin Henderson, USA
Jean-Michel Hoc, France
Victor Kaptelinin, Sweden
Kari Kuutti, Finland
Jari Laarni, Finland
Anneli Leppänen, Finland
Laura Militello, USA
Anna-Lisa Osvalder, Sweden
Antti Oulasvirta, Finland
Nicolas Marmaras, Greece
Antonio Rizzo, Italy
Pertti Saariluoma, Finland
Tjerk van der Schaaf, Belgium
Marcin Sikorski, Poland
Patrice Terrier, France
Gerrit C. van der Veer, the Netherlands
Willemien Visser, France
John Wilson, UK
William Wong, UK

Local Organizing Committee at VTT Finland

Leena Norros
Hanna Koskinen
Leena Salo
Paula Savioja

Contents

Preface	3
International Program Committee	4
Keynote Presentations	9
The Dynamics of Experience: A Search for What Matters 11 <i>John M. Flach</i>	
User and Designer: Affair, Partnership or Nurturance 19 <i>Turkka Keinonen</i>	
Session 1: Developing the Conceptual Basis of HSI Design	27
Perspectives on the Design Process: From a Focus on Artefacts to Practices 29 <i>Victor Kaptelinin, Liam Bannon</i>	
Challenges to Cognitive Systems Engineering: Understanding Qualitative Aspects of Control Actions 37 <i>Morten Lind</i>	
Mental Contents in Analyzing Presence in Human Technology Interaction: the Approach and Preliminary Results 45 <i>Pertti Saariluoma, José J. Cañas</i>	
Session 2: User Experience	51
Do Web Pages Have Personalities? 53 <i>Phil Turner, Susan Turner, Leanne Wilson</i>	
How Cognitive Ergonomics Can Deal with the Problem of Persuasive Interfaces: Is a Criteria-based Approach Possible? 61 <i>Alexandra Nemery, Eric Brangier, Steve Kopp</i>	
Early Adopters' Experiences of Using Mobile Multimedia Phones in News Journalism 65 <i>Tiina Koponen, Heli Väättäjä</i>	
Does Cognitive Style Affect Student Performance on a Web-based Course? 69 <i>Lamis Hammoud, Steve Love, Willem-Paul Brinkman</i>	
Developing a Rapid Questionnaire-based Metric for Experience of Form 73 <i>Ville Lehtinen, Riikka Hänninen, Ville Toivonen, Antti Oulasvirta, Sari Kuuva, Pertti Saariluoma</i>	
Session 3: Coping with Emergency Situation	77
Change Blindness and Situation Awareness in a Police C2 Environment 79 <i>Gabriela Mancero, William Wong, Martin Loomes</i>	
Collaborative Situational Mapping during Emergency Response 85 <i>Lucy T. Gunawan, Augustinus H. J. Oomes, Mark Neerinx, Willem-Paul Brinkman, Hani Alers</i>	
Analysis of Work Demands of Multi-agency Emergency Response Activity for Developing Information Support Systems 92 <i>Leena Norros, Robert Hutton, Patrick Grommes, Nick Colford, Marja Liinasuo, Paula Savioja</i>	
Decision Making with a Time Limit: The Effects of Presentation Modality and Structure 96 <i>Yujia Cao, Mariët Theune, Anton Nijholt</i>	

Posters	105
Levels of Automation and User Control in Use of an Automatic Turbine System	107
<i>Jonas Andersson, Anna-Lisa Osva</i>	
Use Errors and Usability problems in Relation to Automation Levels of Medical Devices	111
<i>Lars-Ola Bligård, Jonas Andersson</i>	
Mobile Air Ticket Booking	115
<i>Ivan Burmistrov</i>	
Towards Co-creation of Service Research Projects – A Method for Learning in the Network	119
<i>Minna Halonen, Katri Kallio, Eveliina Saari</i>	
Exploring the Use of Humanizing Attributions in User Experience Design for Smart Environments.....	123
<i>Minna Kynsilehto, Thomas Olsson</i>	
Human-Centred Design of Technical Infrastructures of Smart Environments.....	127
<i>Marketta Niemelä, Thomas Olsson, Eija Kaasinen</i>	
“Show me, how does it look now”: Remote Help-giving in Collaborative Design	131
<i>Dhaval Vyas, Gerrit van der Veer, Anton Nijholt, Dirk Heylen</i>	
Session 4: Acting in Computerised Everyday Environments	137
A Descriptive Model of Contextual Activities for the Design of Domestic Situations	139
<i>Pascal Salembier, Julie Dugdale, Myriam Fréjus, Yvon Haradji</i>	
Analysis of Activity in Domestic Settings for the Design Ubiquitous Technologies	145
<i>Germain Poizat, Myriam Fréjus, Yvon Haradji</i>	
Food for Talk: Photo Frames to Support Social Connectedness for Elderly People in a Nursing Home	147
<i>Margit Biemans, Betsy van Dijk</i>	
Disembedding Computers – Interfacing Ubiquitous Computers.....	154
<i>William Edmondson, Russell Beale</i>	
Session 5: Supporting Activity by Virtual and Augmented Technology	159
Augmented Reality as Means for Creating Shared Understanding.....	161
<i>Mirja Lievonen, Duska Rosenberg, Ralf Dörner, Guido Kühn, Swen Walkowski</i>	
Field Observations of Therapists Conducting Virtual Reality Exposure Treatment for the Fear of Flying.....	169
<i>Willem-Paul Brinkman, Guntur Sandino, Charles van der Mast</i>	
User Expectations for Mobile Mixed Reality Services: An Initial User Study.....	177
<i>Thomas Olsson, Pirita Ihamäki, Else Lagerstam, Leena Ventä-Olkkonen, Kaisa Väänänen-Vainio-Mattila</i>	
Session 6: Cognitive Analyses of Control Activity	185
The Orchestra: A Conceptual Model for Function Allocation and Scenario-based Engineering in Multi-agent Safety-critical Systems	187
<i>Guy A. Boy</i>	
Cognitive Control in Car Driving: The Case of Brain-injured Drivers	194
<i>Jean-Michel Hoc, Camilo Charron, Isabelle Milleville-Pennel</i>	
Distributed Cognition in Ship Navigation and Prevention of Collision	202
<i>Thomas Koester, Nikolaj Hyll, Jan Stage</i>	
Alarm Handling in Rail Electrical Control.....	210
<i>Nastaran Dadashi, Sarah Sharples, John R. Wilson</i>	

Session 7: Developing Usability Evaluation	219
Familiarity with Mobile Phones	221
<i>Phil Turner, Emilia Sobolewska</i>	
User Testing when Test Tasks are not Appropriate	228
<i>Sirpa Riihiahio</i>	
What Explains Differences in Users' Inclination to Appropriate Technology for Unconventional Purposes? A Preliminary Analysis	236
<i>Antti Salovaara, Sacha Helfenstein, Mikael Wahlström, Antti Oulasvirta</i>	
Developing Practical Tools for User Experience Evaluation – A Case from Mobile News Journalism	240
<i>Heli Väättäjä, Tiina Koponen, Virpi Roto</i>	
Session 8: Analysis and Design of Tools for Process Control	249
Good Practice Regarding Alarm System Design and Alarm Management Strategies for Complex Process Control Settings.....	251
<i>Anna Thunberg, Anna-Lisa Osvalder</i>	
Using Operating Procedures in NPP Process Control	258
<i>Leena Salo, Leena Norros, Paula Savioja</i>	
Improving Safety Through the Design of a New Function: Design Process Findings and Post-release Analyses.....	267
<i>Thomas Koester, Nikolaj Hyll, Jan Stage</i>	
Activity-driven Design Approach to the Novel Control Room Environments	275
<i>Hanna Koskinen, Jari Laarni, Leena Norros</i>	
Session 9: Activity Theory in Design	279
Cultivating and Enriching a Tool in Use – Case ‘Development Radar’	281
<i>Hanna Toiviainen, Hannele Kerosuo</i>	
The Qualitative Differences of the Effects of Technological Changes: Case Wood Procurement Process	289
<i>Arja Ala-Laurinaho, Marika Schaupp, Arto Kariniemi</i>	
Can System Innovations be Facilitated by Societal Embedding?	293
<i>Eveliina Saari, Sirkku Kivisaari</i>	
Strategic Partnership as a Design Challenge – Applying Design Competence to Facilitate Innovation-Driven Relationships and Activities in Public Organization	297
<i>Mervi Hasu, Tuuli Mattelmäki, Salu Ylirisku</i>	
Representing Changes in Work for Design: An Activity-theoretical Perspective.....	301
<i>Laura Seppänen, Arja Ala-Laurinaho</i>	
Session 10: HSI Design for Collaborative Work	305
Being Virtually Everywhere: An Exploration of Teachers' Multitasking in a Hybrid Ecology of Collaboration	307
<i>Rikard Harr, Victor Kaptelinin</i>	
Reconceptualising Clinical Handover: Information Sharing for Situation Awareness.....	315
<i>Stephanie Wilson, Rebecca Randell, Julia Galliers, Peter Woodward</i>	
Design Guidelines for Negotiation Support Systems: An Expert Perspective Using Scenarios.....	323
<i>Alina Pommeranz, Willem-Paul Brinkman, Pascal Wiggers, Joost Broekens, Catholijn M. Jonker</i>	
Session 11: Cognitive Studies in HSI Design	331
Effects of Repetitive Practice on Interruption Costs: An Empirical Review and Theoretical Implications	333
<i>Antti Oulasvirta, K. Anders Ericsson</i>	

Occlusion Technique – Valid Metrics for Testing Visual Distraction?.....	341
<i>Tuomo Kujala</i>	
New Interactions with Workflow Systems.....	349
<i>Ingo Wassink, Paul E. van der Vet, Gerrit C. van der Veer, Marco Roos, Betsy van Dijk</i>	
Session 12: Developing Tools for Design Activity	353
An Approach to Assess the Quality of Collaboration in Technology-mediated Design Situations	355
<i>Jean-Marie Burkhardt, Françoise Détienne, Anne-Marie Hébert, Laurence Perron, Stéphane Safin, Pierre Leclercq</i>	
Some Severe Deficiencies of the Input-output HCI-paradigm and Their Influence on Practical Design	363
<i>Kai Tuuri, Antti Pirhonen, Eve Hoggan</i>	
Designers, Users and a Shared Understanding of the Product: A Case Study with Product Designers.....	370
<i>Anna Mieczakowski, Patrick Langdon, P. John Clarkson</i>	
The Expert Community Staff: An Innovative Method for Capturing End-users' Needs	374
<i>J. M. Christian Bastien, Éric Brangier, Jérôme Dinet, Javier Barcenilla, Gabriel Michel, Robin Vivian</i>	
Session 13: Supporting Usage of Information Systems	379
User Interface Sketching to Improve Low and High Literacy User Information Seeking on Government and Social Service Websites.....	381
<i>Neesha Kodagoda, B. L. William Wong, Nawaz Khan</i>	
A Comparative Field Study of Four Crane Control Interfaces	389
<i>Joanna Bergström-Lehtovirta, Jussi Rämänen, Tiia Suomalainen, Antti Oulasvirta, Tuomo Kujala</i>	
Reasons for Under-use in Professional Software.....	393
<i>Anna Kämäräinen, Pertti Saariluoma</i>	
Anatomy of an Adaptive Scaffold for Analysts.....	397
<i>Stephanie Santosa, Judi McCuaig, Joseph MacInnes, Nathan Kronenfeld, William Wright</i>	
Session 14: Understanding Social Dimensions of Activity	401
Understanding Social Entities, Activity Coupling and Social Computing Affordances in Industrial Work Environments.....	403
<i>Heljä Franssila, Petri Mannonen</i>	
Investigating Accountability Relations with Narrative Networks	411
<i>Daniel Boos, Gudela Grote, Mikko Lehtonen</i>	
What Is Real Work? The Case of Process Control	416
<i>Alexandra Bidet</i>	
Doctoral Consortium	421
Designing Sustainable Tools for Living with the Aid of Mobile Technology in the Emerging Economy Contexts	423
<i>Karthikeya Acharya</i>	
Culture, Notation and Formalisation	427
<i>Laura Benvenuti, Gerrit C. van der Veer</i>	
Friendship Online: Users' Experiences of Relationships in Social Networks	430
<i>Rob Comber</i>	
Assessment of Usability Evaluation Methods.....	434
<i>Sirpa Riihiahho</i>	
User Experience of Mobile Systems and Services – A Case of Mobile News Journalism.....	438
<i>Heli Väättäjä</i>	
Is It Possible to Design for Aesthetic Experience without Damaging Agency? Mórna Ní Chonchúir	442
Author Index	446

Keynote Presentations

The Dynamics of Experience: A Search for What Matters

John M. Flach
 Department of Psychology
 Wright State University
 Dayton, OH, USA
 john.flach@wright.edu

ABSTRACT

The metaphysical assumptions of cognitive science are explored with the goal to develop a problem-centered science relevant to the design and engineering of effective products and technologies. Radical Empiricism is suggested as an ontological foundation for pursuing this goal. This ontology poses a single reality where mind and matter come together to shape experience. Peirce's triadic semiotic system is suggested as a framework for parsing this reality in ways that reveal important aspects of the dynamics of communication and control. Rasmussen's three analytic frames of the 1) Abstraction Hierarchy, 2) the Decision Ladder, and 3) Ecological Interface Design are suggested as important tools for achieving insight into the dynamics of this triadic semiotic system. These ideas are offered as a challenge to both scientists and designers to reassess the basic assumptions that guide their work. The hope is that by facing these challenges we can take the first tentative steps toward a coherent science of what matters.

Keywords

ontology, radical empiricism, semiotics, cognitive systems engineering, abstraction hierarchy, decision ladder, ecological interface design

ACM Classification Keywords

H. Information Systems; H.1 Models and Principles; H.1.2 User/Machine Systems.

INTRODUCTION

What must be admitted is that the definite images of traditional psychology form but the very smallest part of our minds as they actually live. [14]

This quote from James remains as true today as when it was first published more than 100 years ago. Despite enormous advances in our understanding of the biological mechanisms of the brain and nervous system and despite countless volumes of empirical research on human performance and human information processing, there still appears to be a huge gap between the science of cognitive psychology and the experiences of every day life. Thus, those who look to cognitive science for inspiration for

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

designing products and technologies that enhance the qualities of every day life and work are generally disappointed.

Over the years, there have been attempts to bridge the gap between cognitive science and work experience through an expanded portfolio of empirical research that recognizes the value of naturalistic field studies [13, 17] and design driven research [18]. And there is some evidence that people are recognizing and escaping from the false dichotomy of basic versus applied research to appreciate the value of a problem-centered science in the spirit of Pasteur [31]. Yet, despite these changes in attitude toward research, the gap and the associated incoherencies [21] remain.

Thus, I have come to the conclusion that the only way to close the gap is to return to first principles. To reconsider the basic ontological assumptions that set the foundation for cognitive science and, in fact, for science in general. In this paper, I would like to direct attention to an alternative ontology suggested by James [15], which has been largely ignored, but which I believe can help to close the gap and reduce the incoherence between cognitive science and cognitive engineering. From this foundation I would like to reconsider basic principles of semiotics, communication, and control [30, 39] and the overarching framework of Cognitive Systems Engineering articulated by Rasmussen [25] to see if it might be possible to assemble a more coherent story that better captures the every day experiences of life.

RADICAL EMPIRICISM

Metaphysics means only an unusually obstinate attempt to think clearly and consistently... as soon as one's purpose is the attainment of the maximum of possible insight into the world as a whole, the metaphysical puzzles become the most urgent of all. [14]

Most books on psychology and/or cognitive science begin with some consideration of the ontological options available as foundations from which to launch scientific explorations of human experience. Typically, three options are explored: Idealism, Materialism, and Dualism. Of these three, Idealism, the assumption that there is no basis for reality outside of the mind, is generally dismissed as antithetical to science. The other two are both offered as credible starting positions.

Materialism chooses the world of matter as the basis for all experience. This view dismisses 'mind' as an epiphenomenon whose ultimate 'causes' are based in the material substrates of biology, chemistry, and ultimately physics. I get a strong sense, that Materialism is the preferred ontological position of many scientists, including most cognitive scientists (many of who now prefer the label neuroscientist). The problem with this approach is that it invites a reductionist approach that leads the scientist further and further into the micro-structure

of brains, and further from the macro-level experiences of most interest to the designer and the cognitive engineer.

Despite the attractiveness of Materialism for many scientists, most keep an open mind to the possibility of Dualism. The Dualist ontology considers the possibility of two kinds of reality – one of mind and one of matter. This ontology is driven by observations of aspects of human experience that can't be easily explained using the formalisms of biology or chemistry. The term 'emergent' is often used to characterize these properties. Thus, the Dualist ontology suggests that there may be emergent properties and principles of mind that cannot be completely described in terms of or reduced to simple functions of matter. I believe that it is this plausibility of a separate reality of mind that attracts many to the social sciences. And in fact, I believe that despite a preference for Materialism, most Cognitive Science and in fact most of Western thought is based at least implicitly in a Dualistic ontology. Further, I would argue that this Dualism is a primary source of the incoherence and the ultimate gap that separates cognitive science from both the everyday experiences of life and from productive generalizations to design and engineering.

It is important to appreciate that the philosophical arguments that have dominated cognitive science over the last half a century have not been about ontology, but about epistemology. While the dualism of mind and matter has been generally accepted, there have been violent arguments about how these separate realities might interact – that is how the reality of mind can become aligned with or come to *know* the reality of matter. This is the problem of epistemology and the dominant debate has been between the Constructivists [e.g., 12] and the Realists [e.g., 9, 10]. The Realists argue that the reality of matter is directly and completely specified to the mind through constraints on perception and action. This perspective is typically associated with ecological approaches to psychology. The Constructivists, on the other hand, argue that the reality of matter is incompletely specified by the constraints on perception and action. Thus, this epistemological position argues that complete knowledge of the reality of matter can only be constructed by the mind from the inherently ambiguous cues available to perception and action using the tools of logic and inference. This perspective is typically associated with information processing approaches to psychology.

For most of my career, I have lived under the illusion that the three ontological positions outlined above were the only plausible options. And of these options, I found the Dualist ontology with a Realist epistemology the most comfortable stance for exploring problems like the visual control of locomotion [7, 8], graphical interface design [1, 2], naturalistic decision making [4], decision-aiding [3], and general issues associated with performance of human-machine systems [5]. Recently, however, I have discovered an alternative ontology that I find far more satisfying. This is James' [15] Radical Empiricism.

In order to appreciate this alternative ontology, it may help to first consider why the ultimate conclusion of one of the recognized 'fathers' of psychology has been completely ignored by modern cognitive science. Modern science tends to be dominated by an 'either/or' form of reasoning [28]. In this context, a phenomenon has to be *either* mind *or* matter. It can't be both. Thus, we are left with the option of dismissing one or the other as an epiphenomenon (idealism or materialism) or we have to include two exclusive sets or boxes one for things that are really and exclusively mind and the other for those things that are really and exclusively matter (dualism).

Radical Empiricism has been ignored because it does not fit this 'either/or' form of reasoning. James came to the conclusion that human experience is a single reality that is *both* mind (mental, subjective) *and* matter (material, objective). James concluded that the distinction or dualism between subjective and objective did not reflect distinct realities, but rather, reflected the different perspectives of the scientist. We call objective those properties of experience that tend to be invariant across wide changes in perspective and we call subjective those properties of experience that tend to vary with changing perspectives (or observers if you like). But these are not two distinct realities, but rather facets of or perspectives on a single reality of experience. Thus, Radical Empiricism was James' monistic alternative to the dualistic view that still dominates Western thought.

Since James, others have reached a similar conclusion. Most notably, Pirsig [23, 24] has introduced the term 'quality' to characterize the ultimate reality of experience. This *quality* is neither subjective nor objective, but is both! The biologist Rayner [28] suggests an inclusionary approach based on a *both/and* logic with the specific goal of integrating across the many dichotomies created by *either/or* thinking (e.g., subjective versus objective; aesthetics versus pragmatics; emotionality versus rationality). Also, the physicist Wheeler's [5, 38] metaphor of the surprise version of the twenty question game, illustrates an inference from quantum physics that suggests that the idea of an external nature, somewhere 'out there' apart from the observer, waiting to be discovered is an illusion. Nature is a single reality that emerges from the interactions of constraints some of which have been classically attributed to mind and other that have been attributed to matter.

Note that the choice of an ontological stance is 'meta' science. It is not right or wrong; true or false. One choice may be more or less useful for some specific purposes or it may be that one choice leads to more elegant and aesthetical stories. I suggest that Radical Empiricism is a choice that should be considered. And I want to offer the hypothesis that a shift to the ontological position of Radical Empiricism or to a Metaphysics of Quality may be a good first step toward a coherent science of what matters – a first step to bridging the gap between cognitive science and cognitive engineering.

SEMIOTICS, COMMUNICATION, AND CONTROL

Every model is ultimately the expression of one thing we think we hope to understand in terms of another we think we do understand [36].

Although our ontology takes a holistic attitude toward mind and matter, in the face of complexity we must ultimately become reductionists. That is, we need to find a way to parse the complexity into "chunks" or "long threads" that can ultimately be weaved into a coherent narrative. So, following the quote above, the thing that we hope to understand is the *dynamics of experience*, which according to the above ontological position, is both mind (mental) and matter (physical). So the next decision is to choose from the things we think we do understand that thing that will provide the best 'image' or 'language' for building a coherent model or narrative.

A good place to start is *semiotics*, where two perspectives have competed for dominance. The first approach, which set the stage for modern linguistics and cognitive science, is attributed to Saussure. This approach parses the semiotic problem into two components that reflect the signal (e.g., the

symptoms of a patient) and the interpretation (e.g., the diagnosis of the physician). This dyadic approach tends to frame semiotics as a problem of mind, of interpretation, with little consideration for the ecology that is typically the source of the signals (e.g., the actual health of the patient). Peirce [20], a contemporary and colleague of James, offered an alternative perspective. He suggested a triadic view of semiotics that includes the source of the signals (the situation or ecology), the signals (the representation or interface), and the interpreter (awareness or belief) as illustrated in Figure 1. It is important to note that the three elements identified by Peirce are components of a single semiotic system. Thus, these are three facets of a single reality or a single dynamic.

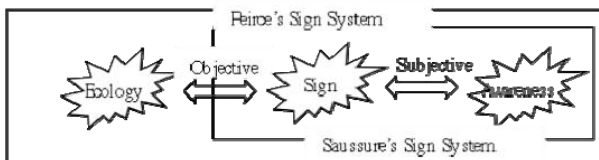


Figure 1. Alternative images of the semiotic system.

In describing the dynamics of experience within this semiotic system, Peirce introduced the logic of *hypothesis* or *abduction* as an alternative to more classical normative models of rationality (i.e., *deduction* and *induction*). Figure 2 illustrates this dynamic, which can best be described in the language of communication and control systems (e.g., Cybernetics, Information Theory, Control Theory, General Systems Theory, and Dynamical Systems Theory) that began developing some 50 years later as scientists and engineers began to explore the design of communication, control, and ultimately computing technologies.

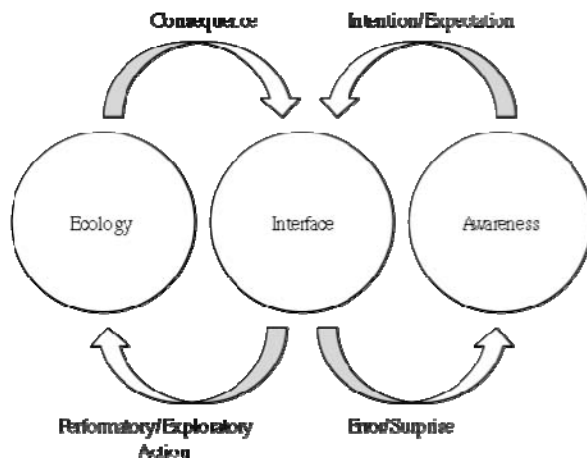


Figure 2. Dynamic coupling of perception (observation) and action (control).

The labels on the links in Figure 2 are taken from the language of communication and control systems. Note that the flow around this system can be viewed from the perspective of either a controller (action) or observer (perception) metaphor, and in fact both dynamics are operating simultaneously in a cognitive system (i.e., perception and action are coupled). The first terms in the paired labels of three of the links are associated with the servomechanism metaphor. With this metaphor *intentions* stimulate *performatory* actions, which in turn have *consequences* that can be fed back and compared with the intentions. The differences are termed '*errors*' and the servomechanism is typically designed to reduce these errors. That is, to stabilize around a steady state where the consequences match the intentions (i.e., error is zero).

The second set from the paired terms in Figure 2 is associated with the ideal observer metaphor. With this metaphor, *expectations* (or hypothesis) lead to *exploratory* actions or tests, which in turn have *consequences* that are fed back and compared with expectations. The differences are termed '*surprises*' and the observer is typically designed to reduce surprise. That is it is designed to stabilize around a steady state where the expectations match the consequences (i.e., surprise goes to zero).

The language of the observer metaphor matches best with the language Peirce used to describe abduction. However, it is clear that Peirce, James and other Pragmatists such as Dewey recognized the intimate connection between control and observation. From the pragmatic perspective, the ultimate test of a hypothesis is whether it leads to successful control – that is, whether it leads to satisfactory results relative to the motivating intentions.

In engineered systems, much of the observer problem (e.g., identifying the state variables) is solved, a priori, by the control system designer, who then designs the control system around the variables that have been identified. However, biological systems must simultaneously solve both the control problem and the observer problem. For example, the novice driver must discover the appropriate variables to attend to and the appropriate gains for linking perception and action while driving (e.g., how hard to press the brakes in order to stop at a specific point). This process of learning by doing often involves active hypotheses (tapping on the brakes) and adjustment of expectations based on the results of the active 'tests' of hypotheses. Hypotheses that lead to successful actions (e.g., comfortable braking) are retained. Hypotheses that lead to surprise (e.g., sudden jerks) are revised. This can be contrasted with inductive and deductive systems of rationality where 'truth' is evaluated relative to the forms of the arguments, independent of any actual correspondence with the ecology. With an abductive logic, there is no absolute 'truth,' there is simply the degree to which the action or belief leads to successful action (i.e., satisfies the intention).

Systems that are engineered to simultaneously solve the control and observation problems are called adaptive control systems. And natural systems where adaptive control processes are observed are typically called 'self-organizing.' The phase shifts often observed in these systems (e.g., gait transitions in animal locomotion) typically reflect how competing tensions between the observation and control demands are resolved by shifting from one locally stable organization to another.

Thus, Peirce's triadic semiotic model provides a logical partitioning of the reality of experience, and the languages of communication and control theories provided a theoretical calculus for describing the dynamic interactions among these elements. Together these set the foundations for Rasmussen's approach to Cognitive Systems Engineering.

COGNITIVE SYSTEMS ENGINEERING

The central issue is to consider the functional abstraction underlying control theory and to understand the implications of different control strategies on system behavior and design requirements. [25].

In approaching the problem of safety in the operation of nuclear power plants, Rasmussen realized that in order to improve the stability of control systems such as a nuclear power plant, it was essential to address the three elements in

Peirce's semiotic system: 1) to understand the constraints of the ecology – the work domain of nuclear power; 2) to understand the constraints on human awareness – the decision strategies of the plant operators; and 3) to understand the constraints on the representations – the design of the human-machine interfaces.

Ecology – Abstraction Hierarchy

Objects in the environment in fact only exist isolated from the background in the mind of a human, and the properties they are allocated depend on the actual intentions [25]

As we unpack the triadic semiotic system, it is important to keep in mind, as implied in this quote from Rasmussen, that the three facets are different perspectives on a single, unitary ontology or system. Thus, it is important to understand that the 'ecology' is not a distinct objective environment. It is perhaps more appropriate to think of this ecology as an Umwelt [35] or problem space [19]. As such, its properties are best conceived as affordances [9, 10] that reflect the functional constraints shaping the field of possibilities and the consequences for good or ill associated with the different capabilities and intentions.

Significant states within this problem space are the goal states corresponding with design and operator intentions. Significant dimensions of this space should reflect distinctive attributes of the process needed to characterize the goals and changes associated with motion through the state space with respect to these goals (e.g., masses, pressures, temperatures, etc. of the thermodynamic processes being controlled relative to process goals). For example, these would be the dimensions that might be used to characterize 'error' with regard to intentions and 'surprise' with regard to expectations in Figure 2.

In addition to the states of the problem space, it will also be important to consider the 'operators' [19] or constraints on action that determine what motions within the state space are possible and/or desirable in terms of associated costs (e.g., effort). This allows the identification of significant paths and functions within the problem space that correspond with physical laws and limits (e.g., mass and energy balances or maximum pressures) and functional values and regulatory constraints (e.g., optimal paths such as the minimum time path from an initial state to a goal state). In control theoretic terms the operators reflect the control interface to the plant dynamic and the functional values and regulatory constraints reflect properties of a 'cost functional' as required to formulate an optimal control model.

The essential challenge in specifying the ecology is to find a way to represent the 'deep structure' of the problem space. Rasmussen [25] proposed that this problem space be specified in terms of "goal-means or means-ends relationships in a functional abstraction hierarchy" (p. 14). He proposed a nested abstraction hierarchy of five levels with the top or dominant category reflecting the functional purpose of the system. Within this broad set of constraints, lower levels add more detailed constraints associated with physical laws and values; general functional organizations; types of physical components; and detailed physical arrangements and forms.

The Abstraction Hierarchy has sometimes been represented as a triangle with the Functional Purpose as the apex and the Physical Form level at the base. The pyramid reflects the increasing number of details or degrees of freedom at the lower levels due to the finer discriminations. However, I prefer to visualize the abstraction hierarchy as a categorical nesting, where the higher levels of abstraction set a fairly broad level of

categories that become a context for evaluating lower levels, where finer distinctions are introduced [6].

Thus, for example the physical laws, the general functions, the physical components, and the physical forms can all be classified with respect to their role as means for accomplishing specific purposes. Likewise, the general functional organization can be understood in the context of the physical laws that govern the process; the physical components can be understood relative to their roles in the general functional organization; and the physical details can be understood in the context of the physical components employed. In this way, higher levels of abstraction provide semantic categories for organizing the details at lower levels of abstraction. And the categories don't simply propagate from one level to the next, but they propagate all the way from top to bottom – with distinctions at the highest level suggesting important semantic distinctions for all levels below.

Details about the different levels and the relations across levels can be found elsewhere [25, 33]. However, for this paper we want to focus on one attribute of the Abstraction Hierarchy that can be puzzling, when viewed through the lens of the more conventional dualistic ontology of mind and matter. How is it that 'goals and intentions' can dominate physical laws? Conventional wisdom would suggest that intentions are dominated by or are subservient to the physical laws (Materialism) or are independent from them (Dualism). We can't choose to violate the laws of motion, or the laws of thermodynamics. In the conventional view of nature, the physical dimensions associated with physical laws are taken as fundamental properties of reality and the goals and preferences of humans tend to be treated as either derivative or independent properties.

However, in a monistic ontology of experience as reflected in James' Radical Empiricism or Pirsig's Metaphysics of Quality, the quality of experience is fundamental and the properties used to characterize the patterns of conventional physical laws are considered derivative properties. This is important to understand. The patterns of the physical laws can be described using extrinsic, observer independent dimensions such as Newton's dimensions of space and time. However, the *meaning* or *significance* of those patterns CANNOT be specified without reference to a goal or value system. Distance can be specified independent of any observer (or actor). But properties like too far, or too close, or close enough cannot be specified independent of some value system (e.g., the potential damage from a collision) and are typically only meaningful relative to some action capability (e.g., quality of brakes). I argue that properties such as too close (e.g., the imminence of collision) are the most fundamental or concrete properties of experience, and that the distance in terms of meters is a useful, but nevertheless a semantically arbitrary abstraction.

In the ontology of experience that we are proposing meaning is fundamental! Thus, the properties most important to the quality of experience are the most basic or the most dominant. And all other dimensions can only be understood in relation to those properties. Thus, although there may be room for much discussion about the best way to carve up the ecology into functional levels of abstraction, we believe that Rasmussen's choice to make the functional purposes the dominant or most basic level of description is exactly the right choice! Meaning can only be understood in the context of intentions and values.

Thus, we would amplify the opening quote from Rasmussen to leave out his qualifier ('in the mind of the human') to state that the real (i.e., ontologically most basic) properties of the

environment can only be defined in relation to the intentions and values of the semiotic system (i.e., human, animal, agent, or organization being studied)!

Note that to illustrate the different layers of abstraction we have used examples from technical systems (e.g., nuclear process control) to reflect the origins of Rasmussen's discoveries. However, I believe that the need to characterize the ecology generalizes to all types of experiential contexts, from highly technical (e.g., aviation and process control), to more social/biological (e.g., medical decision-making), and to more personal and idiosyncratic (e.g., libraries and consumer products). There can, however, be important differences in the nature of the dominant functional goals that shape performance. For example, Vicente [33] uses the distinction *correspondence*–driven versus *coherence*–driven to distinguish domains such as nuclear power from domains like libraries. In technical and safety critical systems the functional goals and constraints tend to be far more clearly specified and rationally justified. As we move into socio-biological systems and consumer products the functional goals and constraints can be increasingly ambiguous/tacit, diverse, and idiosyncratic; and far more tightly linked with emotional, spiritual, and aesthetic aspects of experience. However, I would argue that at the non-technical end of this spectrum, the ecological constraints remain an important and significant facet shaping the dynamics of the semiotic system. Such constraints are no less real and no less important, though they may be far more difficult for the scientist to discover and specify.

Awareness – Decision Ladder

... it is very important to analyze the subjective preferences and performance criteria that guide an operator's choice of strategy in a specific situation. Unless these criteria are known, it will not be possible to predict the strategy that an operator will choose, faced with a specific interface design [p. 23].

Rasmussen's view of the awareness facet of the semiotic system was strongly influenced by his early work on electronic troubleshooting [26]. Observations of the diagnostic processes of electronic technicians showed that the search for faults was often not consistent with conventionally rational hypothetical-deductive processes. Rather, the search process tended to be impulsive and highly context dependent. Often, when working on electronic equipment in the shop, the search involved high numbers of observations using simple procedures, rather than more normatively efficient processes guided by studying the functioning of the system. On the other hand, when asked to evaluate a safety critical system in the experimental reactor, where a mistake could be embarrassing and expensive, more formal analytic thinking was employed. Although, the observed search processes were often not optimal in terms of the number of tests, they did tend to work and they tended to minimize the cognitive demands. That is, they tended to rely on information that was directly accessible in perception-action loops, rather than to rely on more abstract forms of reasoning. In other words, the rationale of the search process tended to be more abductive and pragmatic.

The term that is typically used to characterize the pragmatic reasoning processes that Rasmussen and many others have observed is *heuristic* decision-making [11, 16, 32]. The gist of all of these observations is that humans take short cuts in the sense that they use search strategies that tend to reduce cognitive demands and generally work. However, these shortcuts typically violate general prescriptions of normative logical and economic models of rationality.

When compared and evaluated relative to the prescriptions of normative models of rationality the heuristics are often characterized as *biases* and these processes are typically attributed to limitations on cognitive abilities and resources. However, this perspective fails to appreciate that when grounded in the constraints of problem spaces, these short cuts can be the basis for smart, fast, and frugal decision-making [11, 32]. Rather than being symptoms of weakness, the heuristics can reflect humans' ability to leverage experience (i.e., expertise) in order to think productively about difficult problems [37]. Rather than relying on global context independent forms of reasoning, people tend to utilize problem-solving processes that leverage experience with local ecological constraints. This is postulated as the explanation why experts are able to 'see' good alternatives as one of the first options considered [17].

To illustrate the short cuts that expert problem solvers employed to solve complex problems, Rasmussen [25] introduced the Decision Ladder. Figure 3 shows a cartoon version of the Decision-Ladder illustrating that there is more than one way to solve a problem. With this representation the dark outer arrows illustrate the more formal, cognitively intensive path from a problem to a solution, while the inner paths depict short cuts for connecting perception and action. These shortcuts often significantly reduce the demands on formal reasoning (i.e., what Rasmussen termed *Knowledge-based* reasoning).

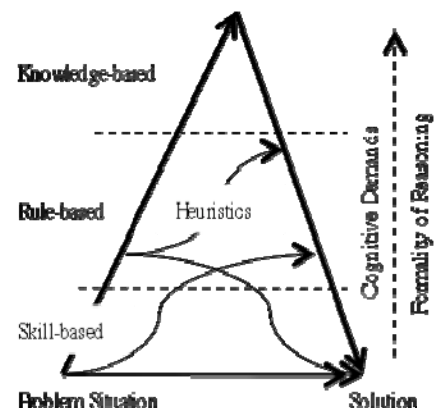


Figure 3. The Decision Ladder illustrates multiple paths from a problem to a solution.

Rasmussen distinguished two types of short cuts. *Rule-based* reasoning reflects short cuts based on explicit patterns or consistencies in the work domain. These short cuts typically reflect standard operating procedures or other learned rules that generally can be easily verbalized by the operators and that reflect typical solutions to common or routine situations.

The second form of short cut is termed *Skill-based*. Skill-based processes are based on invariant patterns that directly specify the links between perception, action, and consequences. Examples of such patterns are the optical invariants associated with control of locomotion [7, 8]. The skill-based short cuts typically result in automatic forms of behavior where the human is only tacitly aware of the patterns that are being leveraged. Thus, it is typically difficult for people to verbalize the basis for directly linking perception and action at this level.

Again, it is important to realize that the patterns that are the scaffold for both rule- and skill-based behaviors are not arbitrary, they are the products of experience and they typically reflect not only the constraints on awareness, but also the constraints within the ecology. The natural evolution of the

abductive, dynamical process illustrated in Figure 2 is toward a match between the constraints or structure in the ecology and the constraints or structure of awareness. The more experience that a person has in a specific ecology or work domain the less the rule-based and skill-based short cuts will appear as *biases* (i.e., overgeneralizations from other contexts), and the more they will appear as *resonance* with the deep structure of the problem space (i.e., ecologically rational solutions)! This process fits well with the abductive form of rationality that Peirce hypothesized as the basis for human reasoning.

This progression from a loose association between the constraints on awareness and the constraints in the problem ecology to increasingly tighter couplings is the development of expertise, skill, or more simply learning. In this context, the heuristics are not symptoms of information limitations, but a natural product of experience. They reflect an evolution from cognitively demanding, normative, formally rational *rote processes* toward more efficient, recognition-primed [17] ecologically grounded [11, 32], *smart cognitive mechanisms* [29]. Rasmussen often commented how often reasoning processes that appeared irrational when viewed from outside the context of a specific domain, would appear as elegant and efficient, once one came to appreciate how the deep structure of the problem domain was being leveraged.

Interface – Ecological Interface

... the goal is to make the invisible, abstract properties of the process (those that should be taken into account for deep control of the process) visible to the operator [27].

Research on problem solving by Gestalt Psychologists [e.g., 37] clearly demonstrates the impact that alternative representations can have on problem solving. A shift in the form of the representation can be the difference between confusion/uncertainty and clarity/insight. With their seminal papers on Ecological Interface Design (EID), Rasmussen and Vicente [27, 34] framed the challenge of interface design around these insights. The key challenge was to go beyond the observations of the Gestalt Psychologists to provide a prescriptive framework for designing effective representations as interfaces to complex work domains.

I believe Rasmussen and Vicente's answer to this challenge can be best appreciated from within the triadic semiotic system. The focus of EID is to specify the structure (e.g., constraints) of the work ecology as clearly as possible. One way to do this is to create a geometric representation that parallels the nested structure revealed by the Abstraction Hierarchy. That is, global properties (e.g., symmetries) of the interface should reflect the functional/intentional context. Relations within these global patterns should reflect the physical laws and organizational constraints on control (e.g., mass and energy balances; feedback relations). And finally, elements within the display should represent the physical details of the process (e.g., relative positions of valves and other components; values of specific variables such as flow rates, volumes, pressures, and temperatures).

Note that once again, mind and matter are intimately related in the interface. The prescriptions of EID will be relevant whether the interface refers to an external display (e.g., engineered interface in a Nuclear Control Room) or to an internal representation (e.g., the learned conceptual visualization of a physics expert). To be effective, the representation must correspond with structure in the functional work ecology and it must be 'as clear as possible.' That is, it must be presented in a way that matches the perceptual abilities, the intentions, and the expectations of the humans. The image of the Decision Ladder provides some guidance for the subjective side of this relation.

In addition to matching the constraints in the ecology, the representation should also leverage the heuristics as reflected in a decision ladder, in order to support rule- and skill-based interactions. This can reduce the demands on more formal knowledge-based forms of reasoning, which might be freed to engaged in more strategic planning and anticipation. In complex domains, consideration must also be given to supporting knowledge-based reasoning as well. It is in these domains where strategic and anticipatory awareness will have the highest payoff and where the inevitable challenge of unanticipated variability will require more formal analytic thinking.

However, I would argue that it is not sufficient to 'match' any heuristics that humans bring to the work situation. Rather, the *ultimate goal should be to shape the heuristics of the humans*. When we view the semiotic system as a resilient, adaptive, self-organizing control system – the goal is for the system to evolve in ways that drive both error and surprise as low as possible. That is, the intentions and expectations need to become increasingly congruent with the consequences. In other words, the heuristics that shape awareness should ultimately match the deep structure of the work ecology. Thus, an ecological interface can play an important role in the education of attention. Helping to make functionally significant structure in the ecology salient, so that it can shape the intentions, expectations, and heuristics such that awareness becomes well tuned to the functional demands of the ecology.

This approach comes into stark contrast with much conventional work on interface design and HCI framed in terms of the dyadic semiotic system. The prescription motivating the conventional work is often to match the operator's mental model, with little consideration to how that mental model corresponds with the deep structure of the work ecology (which sits on the other side of the interface – outside the dyadic semiotic system). Returning to the medical context of semiotics, the goal of these conventional approaches is to match the symptoms with the physician's conceptual model of disease, assuming that the model is valid. The validity of the model, itself, falls outside the dyadic semiotic system. For the EID approach, the ecological foundations of the conceptual model are included as part of the semiotic problem. Thus, interface designers are invited to work with the domain experts to explore and test alternative models of disease, so that testing the ecological validity of the representation (at least in a pragmatic sense) becomes an essential component of the design challenge.

Thus, the representation can be the hinge that determines whether the dynamics of experience converge to a stable solution where consequences align well with intentions and expectations; or whether the dynamic diverges in ways that lead to increasing confusion and surprise and ultimately to catastrophic levels of error.

SUMMARY

“a tale told by an idiot, full of sound and fury, signifying nothing.” Macbeth Quote (Act V, Scene V)

For most of my career, I have felt a deep degree of impatience for the philosophical debates that to me often seemed to be full of sound and fury and of no practical value. This led me to pursue what I considered to be the most concrete, the most practical aspects of human performance as the focus for my research. And this ultimately led me to the field of engineering psychology and the work of Jens Rasmussen. However, I am just beginning to appreciate the wisdom of Rasmussen's

insights. A dramatic step forward for me was the discovery of Peirce's semiotics and abduction, and then connecting the dots with the underlying metaphysics of James' Radical Empiricism [15], Pirsig's Metaphysics of Quality [23, 24] and Rayner's concept of Inclusionalism [28].

But ultimately, the discoveries that helped me to connect the dots between the philosophy, the science, and the design implications did not come from the literature, but from my own experiences of everyday life and from the opportunity to explore the experiences of drivers, pilots, surgeons, emergency room physicians, emergency medical and military commanders. I think this is the challenge that unites James, Peirce, Pirsig, and Rasmussen – a focus on the phenomenon of human performance *as it is lived*, rather than how it might be simulated in a mechanistic, logical machine. This leads to an appreciation of human cognition as an exquisite ability to successfully adapt to the constraints of complex ecologies, rather than as a collection of general, limited information processing mechanisms.

Appreciation of this exquisite ability, however, is difficult to achieve if the research frame does not include the ecological constraints as an integral component of the system of interest (i.e., the semiotic system). On the other hand, it can be impossible to see anything if one tries to see everything. The challenge is to reconcile an expansive ontology that spans both mind and matter, with the practical need to reduce or parse the phenomenon into manageable chunks. In my view, James and Pirsig provide the best guesses about such an expansive ontology. Peirce and Rasmussen provide the best guesses about how to parse that ontology into chunks consistent with the intuitions of communication and control theory (and more broadly dynamical systems theory). And, further, I suggest that these different perspectives fit together in a rather elegant way that might lead toward a problem-centered science motivated by a desire for both theoretical coherence and practical utility.

In sum, the hypothesis of this paper is that the vast gap between cognitive science and cognitive engineering and design (that we all struggle with) is a symptom of a dualistic ontology that places the constraints on products and technologies in one reality and the constraints on cognition (and emotion) in another. I suggest that this is a symptom of an exclusive-or (either/or) form of reasoning that is the source for many false dichotomies (e.g., subjective vs. objective; aesthetics vs. pragmatics; emotional vs. rational; mind vs. matter) that make it difficult to frame a coherent story about the dynamics of human experience. The dichotomy of most concern for this paper is the one between basic and applied science. Unfortunately many of our scientific research paradigms do not represent the practical constraints of natural environments and the resulting theories greatly underestimate the resilience of natural systems. When psychology frames its research questions around abstract general information processing mechanisms, without due consideration for the ecological problems of natural environments, it leads to a basic science that has little to offer and little to learn from the engineering and design community.

Ironically, neither the basic science community (with its emphasis on empirical methods and strong inference based on domino models of causality) nor the design community (with its emphasis on practical innovation and economic success) has much patience for metaphysics, which is typically dismissed as of academic interest only. But this is where both I and James most disagree with the conventional wisdom, we have both come to believe that the only way to close the gap between science and design, the only way to resolve the incoherencies in our theories is to go back to first principles and to reconsider the fundamental assumptions we have made about nature and

reality; the fundamental assumptions about mind and matter; the fundamental assumptions about what matters. At the end of the day I believe that the quality of our science, the quality of our products, and ultimately the quality of our lives is a joint function of both mind and matter!

Finally, let me make it absolutely clear that Radical Empiricism, Peirce's triadic image of the semiotic system, and Rasmussen's formalisms (Abstraction Hierarchy, Decision Ladder, and EID) are not the ends of the search for what matters. Rather, I suggest only that they are a good place to start the search and may provide a tentative direction for productive exploration in our joint search to discover what matters!

ACKNOWLEDGMENTS

I first want to give special thanks to the conference organizers for their kind invitation to present this keynote address. This provided a challenge for me to attempt to connect a wide range of influences that have shaped my experiences into a concise, coherent story. I doubt that I have achieved that goal, but I found this to be a very valuable exercise. Thanks also to the many teachers who have forced me to think about the philosophical and theoretical foundations for my choices and to the many students who have allowed me to kibitz over their shoulders as they explored the frontiers of human experience.

REFERENCES

- [1] Amelink, H. J. M., Mulder, M., van Paassen, M. M. and Flach, J. M. 2005). Theoretical foundations for total energy-based perspective flight-path displays for aircraft guidance. *International Journal of Aviation Psychology*, 15, 205–231.
- [2] Bennett, K. and Flach, J. M. 1992. Graphical Displays: Implications for divided attention, focused attention, and problem solving. *Human Factors*, 34(5), 513–533.
- [3] Brill, E. D. Jr., Flach, J. M., Hopkins, L. D. and Ranjithan, S. 1990. MGA: A decision support system for complex, incompletely defined problems. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(4), 745–757.
- [4] Dominguez, C. O., Flach, J. M., Lake, P., McKellar, D. and Dunn, M. 2004. The conversion decision: Knowing your limits and limiting your risks. In: Smith, K. Shanteau, J. and Johnson, P. (Eds.). *Psychological Explorations of Competent Decision Making*. Cambridge University Press. Pp. 7–39.
- [5] Flach, J. M., Dekker, S. and Stappers, P. J. 2008. Playing twenty questions with nature (The surprise version): Reflections on the dynamics of experience. *Theoretical Issues in Ergonomics Science*. 9:2, 125–145.
- [6] Flach, J. M., Schwartz, D., Bennett, A., Russell, S. and Hughes, T. 2008. Integrated constraint evaluation: A framework for continuous work analysis. In: Bisantz, A. M. and Burns, C. M. (Eds.). *Applications of Cognitive Work Analysis*. Mahwah, NJ: Erlbaum. Pp. 273–297.
- [7] Flach, J. M., Smith, M. R. H., Stanard, T. and Dittman, S. M. 2004. Collision: Getting them under control. In: Hecht, H. and Savelsbergh, G. J. P. (Eds.). *Theories of time to contact*. *Advances in Psychology Series*, Elsevier, North-Holland. Pp. 67–91.
- [8] Flach, J. M., Warren, R., Garness, S. A. Kelly, L. and Stanard, T. 1997. Perception and control of altitude:

- Splay & depression angles. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1764–1782.
- [9] Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- [10] Gibson, J. J. and Crooks, L. 1938. A field-analysis of automobile-driving. *American Journal of Psychology*, 51, 453–471.
- [11] Gigerenzer, G. and Goldstein, D. G. 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- [12] Gregory, R. L. 1974. *Concepts and mechanisms of perception*. New York: Charles Scribner & Sons.
- [13] Hutchins, E. 1995. *Cognition in the wild*. Cambridge, MA: MIT Press.
- [14] James, W. 1909. *Psychology*. New York: Henry Holt.
- [15] James, W. 1921. *Radical Empiricism*. New York: Longmans, Green, and Co.
- [16] Kahneman, D., Slovic, P. and Tversky, A. (Eds.). 1982. *Judgement under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
- [17] Klein, G. 2003. *Intuition at work*. New York: Doubleday.
- [18] Laurel, B. 2003. *Design Research: Methods and Perspectives*. Cambridge, MA: MIT Press.
- [19] Newell, A. and Simon, H. A. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- [20] Peirce, C. S. 1878. *Writings of Charles S. Peirce*. Bloomington, IN: Indiana University Press.
- [21] Percy, W. 1991. The fateful rift. The San Andreas Fault in the modern mind. In *Signposts in a strange land*. New York: Picador.
- [22] Piaget, J. 1973. *The child and reality*. New York: Grossman.
- [23] Pirsig, R. 1974. *Zen and the art of motorcycle maintenance*. New York: Harper Collins.
- [24] Pirsig, R. 1991. *Lila*. New York: Bantam.
- [25] Rasmussen, J. 1986. *Information processing and human-machine interaction*. New York: North Holland.
- [26] Rasmussen, J. and Jensen, A. 1974. Mental procedures in real-life tasks: A case study of electronic troubleshooting. *Ergonomics*, 17, 293–307.
- [27] Rasmussen, J. and Vicente, K. J. 1989. Coping with errors through system design: Implications for ecological interface design. *International Journal of Man-machine Studies*, 31, 517–534.
- [28] Rayner, A. 2006. *Inclusional Nature: Bringing Life and Love to Science*. <http://people.bath.ac.uk/bssadm/inclusionality/>.
- [29] Runeson, S. 1977. On the possibility of ‘smart’ perceptual mechanisms. *Scandinavian Journal of Psychology*, 18, 172–179.
- [30] Shannon, C. E. and Weaver, W. 1949. *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- [31] Stokes, D. E. 1997. *Pasteur’s Quadrant: Basic science and technological innovation*. Washington, D.C: Brookings Institute.

User and Designer: Affair, Partnership or Nurturance

Turkka Keinonen
 University of Art and Design Helsinki
 Hämeentie 135 C
 FIN-00560 Helsinki, Finland
 turkka.keinonen@taik.fi

ABSTRACT

The paper presents a series of conceptual elaborations on modelling the relationships between users and designers within user centred design (UCD). The elaborations are based on the allocations of control, creativity, initiative, and freedom of choice, between users and designers; on the commitment on user need satisfaction; and on the relative distances between the stakeholders in design. The discussion aims at introducing the idea of user-designer relationships as a feasible concept for mapping the field of UCD, and to underline the versatility of relationships that UCD community should be able to assume in its attempts to encounter with the users.

Keywords

user centred design, design collaboration

INTRODUCTION

Involving users into design is seen to be increasing in importance. At the same time the conception of the users themselves has become an issue. Sanders [1], for instance, has described the changes in the vocabulary designers use to refer to the people for whom they design: ‘Customers’ have been replaced by ‘consumers’, ‘users’, ‘participants’, ‘adapters’ and ‘co-creators’. And von Hippel [2] would probably add a ‘principal creator’. The trend is towards an increasingly active user.

Understanding the concept of a user is, indeed, a worth while effort. However, what is perhaps even more interesting and relevant from the point of view of User Centred Design (UCD) practice is to comprehend the relationships between designers and users. UCD is a dialogue and relationship between the representatives of the user and designer communities. Without the fragile relationship sustaining all its problems – e.g. conflicts in interests and problems with cross-disciplinary communication – the goals of collaboration including learning, contributing, challenging, stimulating, inspiring, etc. would not be achieved. Thus, the meaningful issue for designers, when it comes to users, is to see who they are with respect to design and designers – and vice versa.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

The intuitive idea of paying attention to user-designer relationships instead of the conception of users as such gets some theoretical support from the ecological approach to design [3, 4, 5, 6, 7, 8, 9, 10], which links technological systems and their development with social practices under one umbrella. It refers to an attitude of perceiving phenomena in human-technology systems from the point of view of relationships and interactions. According to the ecological approach, environments should be understood as practices that they enable, and their quality refers to the development potential of the human, social, physical and technical systems as a whole. Thus, the objects of design should not be understood as devices or even interactions, but as broader practices. Correspondingly, the ecological view to designing seems to suggest framing users and designers as parts of a single system and seeing users and designers as user-designer relationships with development potential.

This paper, however, does not aim at a faithfully following the ecology discussion. Instead, it aims at summarising, in the spirit of ecological design approach, the author’s attempts to grasp the nature of the user-designer relationships. The attempts draw from a range of discourses, and some are more side products of something else than the results of focused studies. The attempts might perhaps be best understood as a kind of designerly approach to design conceptual mappings. Nonetheless, the author trusts on them to make a worthwhile contribution on our understanding on user centred design by introducing the user-designer relationship as a feasible angle to map the field, and to illustrate the versatility of relationships that competent designers should be able establish with the user community. The attempts are the following:

- The first (Section: “Introtruction”) is a metaphorical presentation drafted for teaching purposes [11, 12]. It illustrates some of the roles that users and designers can assume in a design dialogue. The reason for developing the metaphors was to underline to students the range of possible ways to encounter users instead of stinking to a single method – perhaps the most fashionable one.
- The second (Section: “Design Contribution Square”) attempt is a more systematic and scholarly approach focusing on the allocation of the initiative and control in UCD called Design Contribution Square [13].
- Third (Section: “Need Rose”), the author presents a relationship mapping that is based on a conceptual exercise on user needs and moral responsibilities within UCD [14]. The aim of the exercise has been to discuss the motivational foundations of user involvement, which has a reflection on the types of relationships between users and designers.
- Finally in Section: “Third Wheel”, a new attempt is presented that is built on the recognition of the existence of a third partner in user-designer dialogue, namely the user researcher.

RELATIONSHIP METAPHORS

Recent ideas about users' involvement in design emphasise their active role, creativity and equal initiative with professional designers. This emphasis has time to time been seen to obsolete the ways designers have previously included users. However, we can also regard the trend as a cumulative process where new approaches complement, not replace, the older ones. Depending on the phase, goals and situational factors of design, the competent designers should be able to choose an appropriate way to approach the users. A mapping making the range of relationship models visible, helps in making the choice.

The author's first attempt to map the user-designer relationships was presented as a set of metaphors where designers and users were associated with different professional roles. These include the following seven relationships some of which were drawn from UCD literature, and some suggested by the author [11, 12]:

Engineer-designer and component-user. Typically designers have to use available components in their designs. The system has to be drafted so that the components fit and enable the system to meet its performance goals. The human operator can be seen as one of these components. Its dimensions can be obtained from human factors manuals, where human beings are presented the same way as in a component suppliers' catalogues with specific dimensions and performance capabilities. The data is generic and does not depend on what the components are used for. The metaphor of 'Engineer designer and component user' refers to designers' common habit [15] of regarding people as another type of component in a technical system.

Doctor-designer and patient-user. Technology becomes difficult to use, if it does not correspond to the capabilities and activities of the specific users who happen to be interacting with the product. Therefore it is not enough to be familiar with the generic user, but the focus of attention needs to shift from the generic to more specific interactions between narrower user segments and certain types of artefacts. The designer, who used to regard the user as a component, has to start looking at them as doctors diagnosing patients with their individual problems. The doctor isolates the symptoms, the reasons behind them and decides about the cure. A usability test is an example of the clinic where the symptoms are identified and medication prescribed. The patient-user is met as an individual or as a representative of a focused segment, but as one without expertise and initiative. The user just brings in the problems.

Student-designer and master-user. Anticipating and avoiding problems is more effective than resolving them. In order to anticipate design problems, the designers must be familiar with the practice they design for. With situated contextual observations, for instance, the designers learn the practice in the same way an apprentice learns from his master [22]. In Master-apprentice relationship the users are regarded as experts. There is a reason for every action taken by the user, and understanding the reason is important for the design. The designer, who used to be an engineer making use of the user and a doctor solving the user's problems, becomes a student who listens to and watches the master-user's behaviour to learn.

Coach-designer and athlete-user. According to participatory design, and more recent co-design, the users should be able to contribute and influence in the development of solutions affecting them. Thus, learning from users is not enough, but they have to be given an opportunity to propose improved solutions. Their problem framing and solving needs to be supported and they need to be provided with appropriate equipment to deal with design challenges. The role of the

design professionals changes again becoming a facilitator of design and interaction between a range of stakeholders. The designer-students become designer-coaches. They know the rules of the game and can plan the tactics, but they need the athlete-users to play.

Artist-designer and muse-user addresses those UCD approaches where information gathering and learning is secondary to creative stimulation, as was the case for instance with the inspiring Cultural Probes experimentations [16]. Users are active in contributing to design through different creative and creativity awakening exercises, but the interpretation of users' contributions into design suggestions remains the responsibility of the designers.

Chef-designer and customer-user. The designers can interpret user information in many ways, but it only crystallises and becomes visible when the concepts emerge. The productive designer can be seen as a chef, mixing new flavours in the kitchen. The chef has to taste the flavours himself in order to be able to develop them. The customer can suggest, approve or reject, but it is not his job or even within his ability to create new flavours [17].

Director-designer and character-user. Instead of only creating the designs, the designers also create new behaviours and the actors who act them out. The potential future user can be based on the user studies, but sometimes designers add characteristics that have not actually been observed. These stimulate ideas or make assumptions about emerging future behaviours. The designer takes creative liberties to behave like a theatre director turning a manuscript into a play. The users become fictional.

The seven relationships can be linked by seeing them as a continuum from the systematic generation and application of design knowledge towards increasingly creative processes. They can also be organised on a two-dimensional surface as shown in Figure 1, where the ends of the continuum meet and create a circle of relationships. Organised in this manner, the circle becomes a mapping, where its upper part represents a sphere of expert driven relationships where the control and power is in designers' hands, in the roles of engineer, doctor, director and chef, and the users' presence is passive and even representational. Towards the lower hemisphere the relationships develop towards more equal direction where users' roles become more active including school master, athlete and muse. Horizontally the circle can be seen as a graph the left side of which illustrates relationships generating solid knowledge foundations for design. The right hand side of the circle represents more inspirational relationships aiming at the production of designs without much emphasis on methodological rigor.

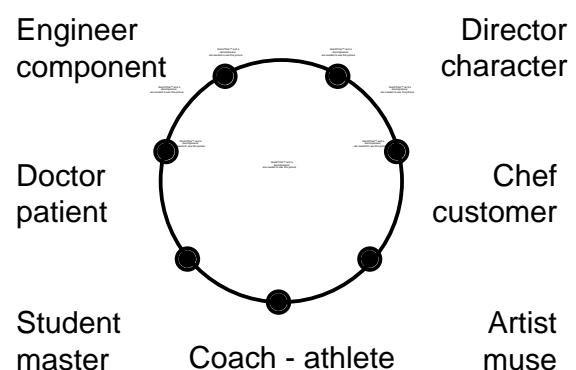


Figure 1. Relationship metaphors.

Even though the metaphors are perhaps illuminating and the mapping seems to have certain intuitive organizing power, the model is not very satisfactory for several reasons. First, it mixes the reasons and procedures of designer-user relationships in a rather coincidental manner. Second, it does not cover some designer-user relationships that have recently gathered a lot of attention. Third, the key concepts of the model, all of which are far from clear, have not been defined, but the model trusts on the readers to make personal appropriate interpretations. Thus, more scholarly models are needed.

DESIGN CONTRIBUTION SQUARE

The author's second attempt [13] to understand the user-designer relationships was built on the metaphor model, but aimed at a more systematic and rigorous concept definition and more inclusive mapping.

The second attempt was elaborated based on a premise that designers' and users' joint contribution defines the limits of UCD. It is build on an idea that the way people contribute to design is fundamental in characterizing the nature of user centred design. It is proposed that the stakeholders' level of activity with respect to design can be used as the single organizing principle for UCD approaches to give enough degree of freedom for giving the field a meaningful structure. The concept of activity refers to participants' interactivity, initiative and style of collaboration and contribution in design events. Activity is understood as a continuous dimension varying between inactive and proactive, but it is described by fixing three landmarks: A participant can contribute in an *inactive* manner with respect to design process, give *reactive* responses to design stimuli or take a *proactive* course of action.

Inactive participants do not contribute to a design process directly by interacting with other participants or by proposing solutions. However, they influence the decisions through other participants' active interpretations and contributions. Inactive participants may be physically present in design events or their contribution may be mediated and represented.

Reactive contribution refers to interactive design participation following a priori agenda and rules set for the design. Reactive participants contribute with the kind of knowledge that the

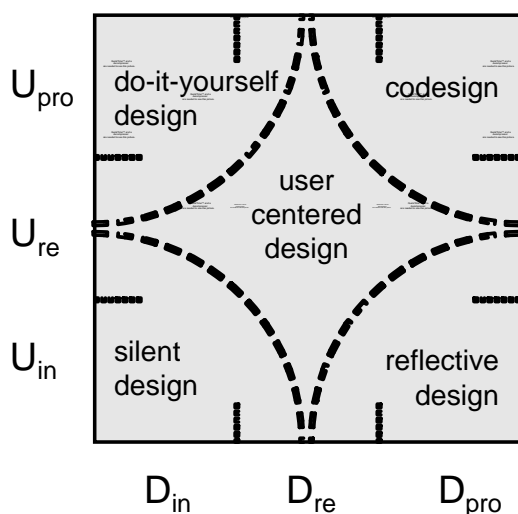


Figure 2. Design Contribution Square [13].

agenda recognizes and, thus, tend to be focused on issues that are known or expected to be relevant in solving the design problem. The reactive participants' roles do not include reconfiguring the way the design challenge is framed and approached, nor do they contribute with non-agenda related knowledge. While the agenda limits the contributions, it also enables focused and efficient progress along the track, and can be the preferred choice when speed and concentration are required.

In proactive participation stakeholders contribute to solving and framing design challenges and influence the procedures used to tackle them. The participants utilize their case and non-case-related expertise to create new connections and interpretations, which cannot directly be deduced from the event agenda. Importing relevant non-task related knowledge, reframing design challenges and adjusting approaches are skills which essentially separate creative design from routine problem solving [18, 19]. While communication in inactive and reactive collaboration is relatively unambiguous thanks to the stability of the interpretation schemas, in proactive mode it becomes a major issue, because new meanings need to be generated collaboratively during the design.

Problem solving and models of teamwork are essential angles in understanding contributions to design. The three contribution styles introduced above correspond to and, thus, get theoretical support from, Rasmussen and Jensen's [20] framework of skill, rule and knowledge based problem solving, and Engeström's [21] models of creative team collaboration as shown in Table 1.

Table 1. Comparison of design contribution modes, problem solving styles according to Rasmussen and Jensen [20] and team work practices according to Engeström [21].

Contribution	Inactive	Reactive	Proactive
Problem solving [20]	Skill-based	Rule-based	Knowledge-based
Teamwork style [21]	Coordination	Cooperation	Communications

For mapping the field of UCD it is necessary to allow both users and designers to adopt the whole range of contribution styles. A Design Contribution Square (DCS) in Figure 2 visually presents these combinations. Each location on DCS represents a unique type of user-designer relationship characterised by the level of activity of the partners. The contributions of other types of stakeholders could be added as 3rd or nth dimension.

In the bottom right corner proactive designers interpret users' reality and define their future without direct user contribution. Designers use documents about users or users' inactive real time participation as the basis of their problem framing and solving. They interpret user data and combine it with their previous knowledge and, when necessary, redefine the design space. They use methods that are developed to allow creative framing, sense making and *reflective design* discussions with the user data [e.g. 22, 23, 24, 25].

The upper left corner refers to situations where users take initiative in adjusting their environment applying *do-it-yourself design* means and tools [26]. Design wise interesting cases are the ones where users create improved versions of products, environments and systems that may have commercial or societal value. According to von Hippel [2] this is likely to happen when the most advanced users within specific practices start improving their equipment.

In the center of the model well-structured methods or rules are applied to guide design interaction, and designers and users contribute in a reactive mode. For instance, a usability test aiming at evaluating the quality of a prototype – rather than improving it – with its predefined task scenarios, measurements and participant roles is an example of an interaction with agenda that ties both users and designers. The interaction can lead to increased understanding about known challenges, but is unlikely to reveal anything unexpected. Even though many UCD scholars would argue for changing these situations towards a more proactive direction, there are reasons to accept the participants' relative passivity. First, evaluation processes benefit from following rigid agendas for comparable results. Second, problems with collaboration and communication are time consuming to solve, and thus agendas structuring design and focusing attention to relevant problems bring efficiency benefits.

Top right corner of DCS refers to proactive codesign, which is the most demanding form of interaction, because the exploration and redefinition of the problem takes place through a cross-disciplinary dialogue. Well-structured methods are not ideal for this, as they tend to propose a fixed agenda for the process and support only certain kinds of problem framings. Thus, more open approaches are needed where guiding is replaced by facilitation.

Silent design is a concept referring to design without explicit contribution of design professionals [27]. Correspondingly, silent UCD in lower left corner of Figure 2 refers to design without active participation of users or designers. One of designers' tasks is to produce decision-making material such as concepts and their justifications. When this material conveys the essentials of users' practices and designers' ideas, also silent design can lead to informed and appropriate solutions.

Design contribution square systematizes and brings a flavour of academic rigour to the metaphor approach. However, at the same time it omits much of what is important in mapping the motivational dimension of user-designer relationships. Consequently, a look at the objectives of user-designer relationship deserves some more attention.

NEED ROSE

The author came to elaborate the motivational basis of user-designer relationships as a side result of another question. For the purpose of understanding the connections of UCD and satisfying users' needs, the author developed two dimensions characterizing the fundamental need satisfying capability of UCD [14]. Later on, these were used to sketch a mapping of the objectives of user-designer interaction.

For isolating individuals' fundamental needs from their subjective desires, scholars recommend focusing on the harmful consequences of failing to satisfy the needs [28, 29, 30, 31]. The link between harmful consequences and fundamental need turns the attention of those designers, who want to concentrate on need satisfaction, to avoiding and solving problems. Based on this principle we can formulate a *protection dimension* of need satisfaction through design, and define it as designers' inclination towards protecting the prospective users of harm. Along the dimension designers, when capable of deciding between explicit or implicit alternatives, increasingly choose courses of action that reduce users' existing or potential harm and avoid causing them new harms.

The designers reduce harm rather than give priority to working for alternative design objectives and criteria possibly including goals such as increasing users' pleasure, providing new kinds of experiences, improving the profitability of the business, or penetrating new markets to name few.

The protection condition puts design into a defensive mode. While this might awake aversion among the design community in general, much of UCD has worked in this mode. The methodological and conceptual tools developed for usability engineering provide means for defending users against harm [32, 33, 34, 35, 36, 37]. More recently interaction designers have abandoned their faithfulness to user protection, even though that of user-centeredness would still be appreciated [38, 39, 40, 41, 42, 43]. Design for user experience, affective, pleasurable, and emotional design [40, 43, 44, 45, 46, 47] have taken more initiative to change the users' reality. Thus, we can see that the protection dimension is a relevant measure for UCD approaches.

Another line of thought concerning fundamental needs [30, 48, 49, 50, 51] is based on the idea of individual's capability to behave autonomously and in a non-impaired manner being herself able to choose what is best for her, and fulfilling the expectations related to her roles in society. Thus, satisfying users' needs in UCD seems to require taking up a holistic non-reducing view of the users allowing them to behave autonomously in the roles and situations that they have assumed including the design collaboration itself. Based on this we can formulate the *appreciation dimension*, which characterizes designers' inclination towards appreciating the non-reduced agency of the users.

Appreciation dimension is parallel idea to what the previous mappings, i.e. the metaphors and DCS, have been tangential. On one end of the dimension the users are reduced into quantitative generalizations and parts of system equal to machinery, while in another end they are regarded as emotional, creative, participating and initiative taking human beings utilizing advanced competences in demanding manners for their contextualized innovations.

It may be difficult to claim that the design approaches, which utilize reduced conceptions of the user would necessarily produce artefacts the use of which would compromise human autonomy, or correspondingly that the inclusion of 'augmented users' would lead to superior need satisfaction. However, reduced conceptions of users in design cannot model, and thus, justly enable design for a holistic user. If we understand the journey from Dreyfuss' [52] anthropometric models, Joe and Josephine staring at a radar display, to von Hippel's lead users creating innovative open source software as an accumulative development of UCD, we can probably agree that the possibilities of the discipline to satisfy the fundamental needs of being human have improved.

Protection and appreciation can be regarded as two milestones on a single trend of UCD development from 1980s to 2000s. However, if we assume that protection and appreciation are independent dimensions characterizing UCD community's conceptions of the objectives of its work and the users it serves, we can once again construct a two-dimensional space that allows us to deal with a range of orientations designers may follow in their attempts to satisfy users' needs (Figure 3).

Let protection dimension grow horizontally to the right and appreciation dimension vertically upwards. In the *protection* end designers commit themselves to safeguarding the users against harm caused as a consequence of interacting with an artefact. An opposite to protecting users would be abusing them, but this does not belong to normal design practice. However, it is easy to imagine a design project that leaves the users to cope with artefacts and the consequences of their use without much thought or effort to safeguard them. Designers can give priority to other objectives and neglect designing for harmless use. The resulting human-technology *confrontation* may be stimulating, motivating and give positive experiences, but also the opposite is possible, even likely.

At the *reduction* end of appreciation dimension users are considered components of socio-technical systems mechanically responding to the techno-oriented demands as discussed above. The users are not considered capable of influencing on the way human-technology relationships are composed. In the opposite, *augmentation*, end they are regarded as empowered competent actors creating what they regard worthwhile. The acknowledged competence of users and trust on their judgment allows the design community to allocate them initiative and responsibility to reinnovate practices and develop artefacts.

The appreciation-protection space cuts strategies to satisfy users' needs into quadrants. The most casual one in lower left corner of the Need rose in Figure 3 confronts and reduces the users. The design compromises satisfying users' needs, and the designers take advantage of the users' ignorance, flexibility, tolerance, and lack of choice. This kind of user *exploitation relationship* may lead to users' dissatisfaction, low efficiency and malfunctions. However, this may be difficult to notice because of users' compensating efforts, investments in education, and tolerating the frustrations due to, e.g., self-attribution [42].

In the reduction-protection corner, designers take action to protect the users from harm, but do it only in respects that the reduced conception of the user covers. The resulting artefacts should not harm users, but they don't either stimulate them, improve their skills, or enhance their practices. Not trusting users walls the design insiders responsible for the products from the outsiders. The outsiders are protected from the pain of facing the complexity, because the hard technology is *cushioned* to appear as soft, human and harmless. For several product categories, cushioning relationship may be the preferred solution. When the designers limited and reduced conception of users fails to reflect the real application of the products, the strategy stops working.

When users are simultaneously appreciated and confronted, they are regarded as competent and initiative taking agents capable of joining the project of developing practices, skills and new technologies. This relationship *excites* the users, stimulates exploration, modifications, and reinnovation. However, the relationship may lead to undesirable consequences of which the designers do not assume responsibility. The user is facing the complexity of the technology without cushioning layers and is herself responsible for making what seems relevant and desirable. The users is also responsible for the safety, social, financial, etc. risks involved in the technology and its use, and may fail in meeting the challenges.

Finally, the most user need oriented relationship combines the commitment to appreciating and protecting users. The resulting designs are build on the recognition of a variety of users' strengths, social and individual aspirations, and the designers take serious responsibility for guiding the users into a direction that protects them from harm. The technology is transparent

and approachable allowing the users to fundamentally influence on their environment and the ways technology supports their practices. The technology is designed so that the exploration and modifications do not lead to harmful results, any more serious than what is acceptable and necessary to support learning. The design aims at *nurturing* the users to create individual and social well-being.

THIRD WHEEL

The attempts above have regarded the user-designer relationship as a one-to-one type of direct bilateral dialogue. However, this may not be the full picture of the reality. We can identify plenty of other stakeholders aiming at participating the dialogue and making the relationship an essentially more complicated network. In the following we limit the discussion into one new partner namely the representative of the UCD community. This assumes the users and designs being representatives of their own communities that are characterised by different attributes than the UCD community. This person, let us call him or her a Researcher, may influence the user-designer dialogue in many ways and lead it to conclusions that are different from what the users or designers might have ended up without the third wheel. The topics of allocating expertise, creativity, responsibility, initiative, control, motivation, etc. discussed above need to be reconsidered along with the new member. There are several ways of conceptualising the roles the third wheel. One of these is to look at the closeness of the relationships between the partners. The idea of 'closeness' is not defined or elaborated in detail, but we trust on it being clear enough based on its everyday meaning, i.e. referring to the frequency of communication, the strength of empathy, mutual dependencies, etc.

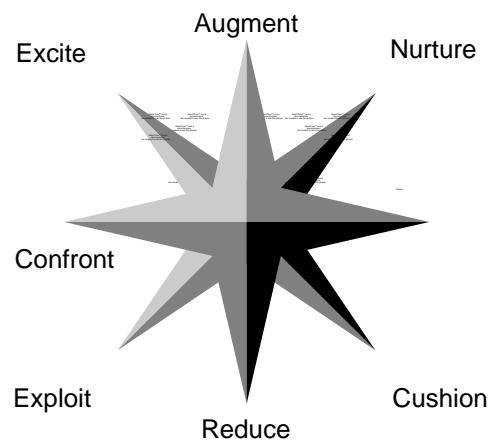


Figure 3. Need Rose describing the objectives of user involvement.

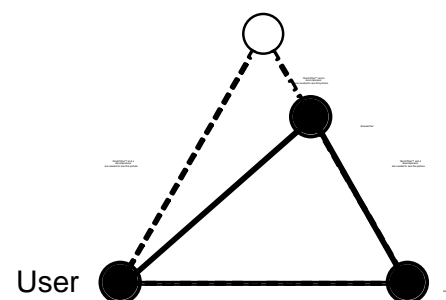


Figure 4. The third wheel.

We can imagine a triangle of relationships where the distances between users, researchers and designers are equally long and use the following notification for it: $U - R - D$. However, a perfect balance is probably unlikely. Consequently, we pay attention to less balanced settings (see Figure 4). First, the researcher may assume a position that is closer to the user than the designer, i.e. $(U - R) - D$. The researcher feels loyalty with the user, shares the user's values, organizes her work so that she becomes in closer contacts with the users than the designers. This kind of user-researcher relationship can be seen to create a new kind of sophisticated user who has the qualifications for analytical description, self-reflection and analysis of the practices of use. The user-researcher is capable of communicating the needs and wishes of the user community clearly and explicitly to the extent the research qualifications and the inside point of view allow. The user-researcher might challenge the lonely and distant designer to respond to the needs. However, the growing distance to the designer may also start to generate communication problems between design and user research and designers may start to dig themselves deeper into their old creativity trenches.

In contrary, if the researcher associates herself with the design, i.e. $U - (R - D)$, we can witness the birth of a designer-researcher. Research becomes design driven and design research driven. The designer-researcher would be loyal to the values of the design organization faithfully supporting its innovation aims. The research driven design might start "using the users" for the design community's goals that are not necessarily shared by the users.

The third model would associate users and designers leaving the user researchers into a more peripheral position, i.e. $(U - D) - R$. In this kind of situation the relationship between users and designers might be seen as an object of research or an entity to be improved or facilitated by the researchers. The researcher would probably be associated with a research community and affiliated to an academic institution. The researcher's position allows her to look at the user-designer relationship in a neutral and perhaps even strategic manner. However, the interests and contributions that these researchers represent may remain distant from the point of view of the practitioners, and user research become isolated academic discipline.

The user-researcher-designer relationships presented in the above explained manner can be connected to a model of UCD practices that divides design into three modes namely immediate, product and remote design [6, 53]. Immediate design refers to design that is highly responsive to users' current needs. It takes place where the activity is, on the site, and aims at solving the problem directly without withdrawing to more time consuming and generalizing product development mode. In addition to being immediate time- and location-wise, it should be immediate when it comes to the causes of design and the interaction with users: users' needs are the immediate reasons to which the design responds, rather than a global trend, business strategy, or technical opportunity. Even though the authors describe immediate design as a process where the designer is close to users, we can easily associate the $(U - R) - D$ relationship with immediate design. Immediate design links the user research activities directly to the use and users.

Remote design refers to design that aims at structural changes. Remote designers work for general solutions, principles, or understanding over individual contexts or implementations. They create conceptual, infrastructure, methodological, regulatory, competence, or resource-related foundations for others to develop products or local practices. When remote designers' conceptual work turns into more tangible design, the results are

models for generic platforms that will be adjusted before becoming useful for end-users. Remote designers' scope of interest in time and coverage is broader than that of immediate or product designers. It is distant from users' immediate needs in terms of time, location, reason, and status because its impacts incarnate much later and in modified appearance. The relationship of user researcher with users and design in remote design is of type $R - (D - U)$. When researchers associate their roles with product design the relationships would be of type $(D - R) - U$.

DISCUSSION

The paper has described four attempts of mapping user-designer relationships. The mappings are built on metaphors of control and creativity; on the freedom of activity in collaboration; on ideas about the need satisfying dimensions of UCD; and finally on the differences in the clones of relationship between stakeholders. The discussion shows that the history and present practice of UCD includes a range of different user-designer relationships. It has also become evident that the spectrum of the relationships appears different when we change our angle of observation. The design ideological point that the author wishes to make and underline is to avoid seeing the development of UCD as a simple trajectory which, depending on the agenda one subscribes, proceeds either towards increased research rigour or trust on end-users' competence and creativity.

Accepting the plurality of values, goals, means and types of relationships between users and designers in UCD, makes the scenery mistier, but good navigation tools help.

REFERENCES

- [1] Sanders, EB-N. 2005. Information, Inspiration and Co-creation. 6th International Conference of the European Academy of Design, March 29–31 2005, University of the Arts, Bremen, Germany.
- [2] Von Hippel, E. 2005. Democratizing Innovation. MIT Press.
- [3] Flach, J., Hancock, P., Caird, J. and Vicente, K. (Eds.). 1995. An Ecological Approach to Human Machine Systems I: A Global Perspective. Lawrence Erlbaum.
- [4] Vicente, K. J. 1999. Cognitive Work Analysis. Toward a Safe, Productive, and Healthy Computer- Based Work. Lawrence Erlbaum.
- [5] Norros, L., Kuutti, K., Rämä, P. and Alakärppä, I. 2007. Ekologisen suunnittelukonseptin kehittäminen [Development of Ecological Design Concept]. In: Kaasinen, E. and Norros, L. (Eds.). Älykkäiden ympäristöjen suunnittelu – Kohti ekologista systeemiajattelua [Design of Intelligent Environments – Towards Ecological System Approach]. Teknoliogiatoollisuus. Pp. 52–90.
- [6] Kuutti, K., Keinonen, T., Norros, L. and Kaasinen, E. 2007. Älykäs ympäristö suunnittelun haasteena [Intelligent Environment as a Design Challenge] In: Kaasinen, E. and Norros, L. (Eds.). Älykkäiden ympäristöjen suunnittelu – Kohti ekologista systeemiajattelua [Design of Intelligent Environments – Towards Ecological System Approach]. Teknoliogiatoollisuus. Pp. 32–51.
- [7] Gibson, J. J. 1979. The Ecological Approach to Visual Perception. Houghton Mifflin.
- [8] Bronfenbrenner, U. 1979. The ecology of human development. Harvard University Press.

- [9] Engeström, Y. 1999. Activity theory and individual and social transformation. In: Engeström, Y. Miettinen, R. and Punamäki, R.-L. (Eds.). *Perspectives in Activity Theory*. Cambridge University Press. Pp. 19–38.
- [10] Gibbons, M., Limoges, C. Nowotny, H., Schwartzmann, S., Scott, P. and Trow, M. 1994. The new production of knowledge. The dynamics of science and research in contemporary societies. Sage publications.
- [11] Jääskö, V. and Keinonen, T. 2004. Käyttäjätieto konseptoinnissa [User Information in Concept Design]. In: Keinonen, T. and Jääskö, V. (Eds.). *Tuotekonseptointi. Teknologiateollisuus*. Pp. 81–114.
- [12] Jääskö, V. and Keinonen, T. 2006. User Information in Concepting. In: Keinonen, T. and Takala, R. (Eds.). *Product Concept Design – A review of the Conceptual Design of Products in Industry*. Springer. Pp. 91–132.
- [13] Keinonen, T. 2009. Design Contribution Square. *Advanced Engineering Informatics*, 23, 142–148.
- [14] Keinonen, T. 2008. User Centered Design and Fundamental Need. Proceedings of Nordic CHI08, Lund, Sweden, 20–22 October. ACM Press.
- [15] Darses, F. and Wolffb, M. 2006. How do designers represent to themselves the users' needs? *Applied Ergonomics*, 37, 757–764.
- [16] Gaver, W., Dunne, T. and Pacenti, E. 1999. Cultural probes. *Interactions*, 6(1), 21–29.
- [17] Kelley, T. 2001. *Art of Innovation: Lessons in Creativity from Ideo*. Doubleday.
- [18] Lawson, B. R. 1997. *How Designers Think*. 3rd ed. Architectural Press.
- [19] Schön, D. A. 1983. *The Reflective Practitioner. How Professionals Think in Action*. Basic Books.
- [20] Rasmussen, J. and Jensen, A. 1974. Mental procedures in real-life tasks: A case study of electronic troubleshooting. *Ergonomics*, 17, 293–307.
- [21] Engeström, Y. 1993. Moniammatillisten tiimien toiminnan analysointi. In: Simoila, R., Harlamow, A., Launis, K., Engeström, Y., Saarela, O. and Kokkinen-Jussila, M. (Eds.). *Mallit, kontaktit, tiimit ja verkot: Välineitä terveyskeskustyön analysointiin*. Stakes, Raportteja 80.
- [22] Beyer, H. and Holtzblatt, K. 1998. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann.
- [23] Ylirisku, S. and Buur, J. 2007. *Designing with Video – Focusing the user-centered design process*. Springer.
- [24] Jonas, W. 2001. A scenario for design. *Design Issues*, 17(2), 64–80.
- [25] Pruitt, J. S. and Adlin, T. 2006. *The Persona Life cycle. Keeping People in Mind Throughout Product Design*. Morgan Kaufmann.
- [26] Fisher, G. 2003. *Meta-Design: Beyond User-Centered and Participatory Design*. Proceedings of HCI International, 22–27 June, 2003 Crete, Greece.
- [27] Gorb, P. and Dumas, A. 1987. Silent design. *Design Studies*, 8(3), 150–156.
- [28] Wiggins, D. 1998. What is the force of the claim that one needs something? In: Brock, G. Ed. *Necessary Goods. Our responsibilities to meet other's needs*. Rowman & Littlefield. Pp. 33–56.
- [29] Thomson, G. 1987. *Need*. Routledge & Kegan.
- [30] Thomson, G. 2005. Fundamental needs. In: Reader, S. (Ed.). *The philosophy of need. Royal institute of philosophy supplement: 57*. Cambridge University Press. Pp. 175–186.
- [31] Frankfurt, H. G. 1998. Necessity and desire. In: Brock, G. (Ed.). *Necessary Goods. Our responsibilities to meet other's needs*. Rowman & Littlefield. Pp. 19–32.
- [32] Lewis, J. R. 2006. Usability testing. In: Salvendy, G. Ed. *Handbook of Human Factors and Ergonomics*. Wiley. Pp. 1275–1316.
- [33] Nielsen, J. 1993. *Usability engineering*. Academic Press.
- [34] Rubin, J. 1994. *Handbook of usability testing: How to plan, design, and conduct effective tests*. Wiley.
- [35] ISO 9241-11. 1998. Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. International Organization for Standardization.
- [36] Eason, K. D. 1984. Towards the experimental study of usability. *Behaviour & Information Technology*, 3, 2, 357–364.
- [37] Chapanis, A. 1991. Evaluating usability. In: Shackel, B. and Richardson, S. (Eds.). *Human factors for informatics usability*. Cambridge University Press. Pp. 359–398
- [38] Helander, M. (Ed.). 1988. *Handbooks of human-computer interaction*. North-Holland.
- [39] Helander, M. and Tham, M. P. 2003. Hedonomics – affective human factors design. *Ergonomics*, 46(13–14), 1269–1272.
- [40] Helander, M. G. and Khalid, H. M. 2006. Affective and pleasurable design. In: Salvendy, G. (Ed.). *Handbook of Human Factors and Ergonomics*. Wiley. Pp. 543–572.
- [41] Norman, D. A. and Draper, S. W. (Eds.). 1986. *User centered system design: New perspectives on human-computer interaction*. Lawrence Erlbaum.
- [42] Norman, D. A. 1988. *The psychology of everyday things*. Basic Books.
- [43] Norman, D. A. 2003. *Emotional Design: Why We Love (Or Hate) Everyday Things*. Basic Books.
- [44] Desmet, P. 2002. *Designing Emotions*. Doctoral dissertations. Technische Universiteit Delft, the Netherlands.
- [45] Jordan, P. 2002. *Designing Pleasurable Products: An Introduction to the New Human Factors*. Taylor & Francis.
- [46] Khalid, H. M. 2006. Embracing diversity in user needs for affective design. *Applied Ergonomics*, 37, 409–418.
- [47] Koskinen, I., Battarbee, K. and Mattelmäki, T. (Eds.). 2003. *Empathic Design. User Experience in Product Design*. IT Press.
- [48] Brock, G. 2005. Needs and global justice. In: Reader, S. (Ed.). *The philosophy of need. Royal institute of philosophy supplement 57*. Cambridge University Press. Pp. 51–72.
- [49] Braybrooke, D. 1987. *Meeting Needs*. Princeton University Press.

Keynote Presentations

- [50] Doyal, L. 1998. A theory of human need. In: Brock, G. (Ed.). *Necessary Goods. Our responsibilities to meet other's needs*. Rowman & Littlefield. Pp. 157–172.
- [51] Penz, P. 1986. *Consumer Sovereignty and Human Interest*. Cambridge University Press.
- [52] Dreyfuss, H. 2003. *Designing for People*. Allworth Press.
- [53] Keinonen, T. 2009. Immediate and Remote Design of Complex Environments. *Design Issues*, 25(3), 62–67.

Session 1: Developing the Conceptual Basis of HSI Design

Perspectives on the Design Process: From a Focus on Artefacts to Practices

Victor Kaptelinin
Department of Informatics
Umeå University
901 87 Umeå
Sweden
vklinin@informatik.umu.se

Liam Bannon
Interaction Design Centre
Dept. Computer Science and Information
Systems University of Limerick
Ireland
liam.bannon@ul.ie

ABSTRACT

We argue that in order for the field of interaction design research to respond to the challenges associated with ubiquitous computing environments and other recent developments in interactive technologies, two distinct orientations, predominantly concerned with, respectively, technological innovation and practice development, need to be clearly articulated. Historical evidence presented in the paper suggests that while these orientations are mutually complementary and dependent upon each other, they are also associated with substantially different concerns and priorities. An outline of a model describing the relationship between technological innovation and practice development in the context of interaction design is introduced. The paper identifies *product design* and *practice development* as two different perspectives in interaction design, which need to be coordinated. Implications of the analysis for future research agenda of interaction design are discussed.

Keywords

interaction design, product innovation, practice development, participatory design, ubiquitous computing

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *user-centered design*.

INTRODUCTION

The field of “interaction design” encompasses scientific and engineering disciplines, the human and social sciences, and the design disciplines (here we use “interaction design”, or “IxD”, in a broad sense, as an umbrella term for research in human-computer interaction (HCI) and related areas, such as cognitive ergonomics). Given such a pluridisciplinary makeup (cf. [11]), it is not surprising that the key concerns of the field seem at times confused. In this paper, we provide a perspective on the research field that outlines the need for attention to be paid to both *technological innovation*, in terms of the design of new artefacts, and *practice development*, in terms of employing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

interactive technologies to support developmental transformation of the activities of people. The latter aspect has been relatively neglected in discussions to date, in terms of systematic analysis. This paper attempts to show the importance of this aspect of design, and provide an initial frame for discussing the topic.

The vast bulk of the literature in the area of interaction design has tended to focus on the design and development of usable and engaging new interactive products, and occasionally, services. This focus on product development has tended to overshadow another equally important perspective on the design and use of interactive systems, namely, a practice perspective that examines artefacts in use and the various ways in which people assemble, adopt, adapt, modify and make “habitable” interactive systems in a variety of local contexts. While it is the case that in certain areas of Informatics, notably in certain participative and process-oriented approaches, there has been some investigation of such aspects as end-user programming, and tailorable and extensible systems development, the relative importance of this perspective has generally been quite low within the professional community.

We believe that current work in the area of ubiquitous computing and sensor developments requires a more analytic investigation of how to support people in assembling technical hardware and software “components” for particular tasks and activities. Recent developments in user-driven innovative practices, the development of exciting forms of mash-ups of software sub-systems, innovative bricolage practices, etc. require that we investigate in a more thorough fashion the practice side of the equation.

An attempt to bring practice development approaches and cognitive ergonomics closer together was made at the ECCE 11 conference [2], where the opening plenary session featured practice development methodologies, including the “clinic of activity” [10] and “change laboratory” (see, e.g., [16, 23]). In this paper we continue this line of exploration, with the aim of better integrating practice development concerns into the conceptual framework of interaction design.

This paper is organized as follows. In the next section we argue that recent developments in ubiquitous computing call for expanding the scope of interaction design from technological innovation to practice development. After that, in Section: “Participatory Design as a Product Design Methodology or a Means for Local Change?” a short historical digression, we discuss some twists and turns of the historical developments of the participatory design approach, which indicate that maintaining a consistent research agenda when simultaneously aiming at technological innovation and practice development, can be somewhat problematic, which suggests that these two aims need to be clearly differentiated. In Section: “Technological

Innovation and Practice Development: a Tentative Model” we introduce a model, which represents the relationship between technological innovation and practice development as a network of human activity systems. Finally, in Section: “Conclusions and Directions for Future Research”, we conclude with discussing the implications of the conclusions made in the paper for the research agenda of future interaction design.

PRACTICE DEVELOPMENT – BEYOND TECHNOLOGICAL INNOVATION

The first thing we need to establish in our analysis is that technological innovation and practice development are indeed two different objectives of interaction design. The difference is not obvious. For instance, one can argue that technological innovation – that is, the creation of a novel interactive product by a professional designer, a product which is then implemented and deployed in a variety of real-life settings – is not only inseparable from practice development but can even be considered *the* activity through which interaction design contributes to practice development. In this section we argue against this view. We claim that practice development through re-mediation of human activities by different forms of interactive technologies *cannot be limited* to technological innovation, that is, a sequence of activities, which begins with identifying a need for a new technology and ends with a successful appropriation of the newly designed technology.

Of course, we do not suggest that adopting technological innovation as the key objective of interaction design means that technology is considered independently of human practices. Directly or indirectly, through related notions of “context” or, more recently, “ecology”, human practices are increasingly often taken into account in all areas of interaction design (e.g., [3, 7, 27, 35, 40]).

For instance, one of the main principles of user-centred design, according to Gulliksen and Göransson [21] is as follows: “The work practices of the users control the development” (see also [22]). “Context” is the key notion in the “contextual design” framework, which is defined Beyer and Holtzblatt [6] as “a customer-centered approach to systems design”. Forlizzi [19] proposes the “product ecology” framework, which aims at helping designers create more successful products, in particular, by supporting the selection of appropriate design research methods. These and other related design approaches (e.g., [31]) do pay close attention to human practice.

However, these approaches take practice into account not for the sake of practice *per se*, but rather with the aim of creating successful innovative technologies. From this perspective human practice is a source of new product ideas, a testbed for evaluation of prototypes, or potential market for new products. Figure 1 shows a simplified model, which represents the variety of ways, in which practice is involved in technological innovation throughout the lifecycle of user-centred product design.

The design lifecycle is initiated by identifying a problem with existing user practice. The problem is addressed by the designer by carrying out a number of design iterations, each including creating and evaluating a prototype of a novel artefact. The designed artefact is eventually implemented, deployed, and appropriated by the user. Appropriation of the artefact includes both developing a repertoire of new strategies, procedures, and tasks (that is, modifying user practices) and modifying the artefact itself (e.g., [34, 38]). In recent years the need for designers to anticipate and support various user activities related to artefact appropriation, including end-user

development, has been emphasized by interaction design researchers (e.g., [18, 26, 41]), and we will mention this aspect later in the paper.

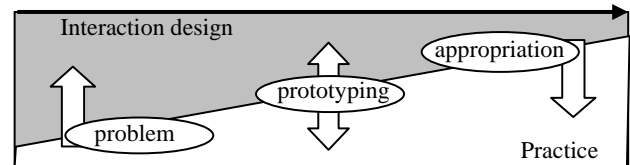


Figure 1. A simplified model of the relationship between interaction design and human practice throughout the lifecycle of user-centred design activities (“problem”: problem identification; “prototyping”: development and evaluation of prototypes; “appropriation”: implementation, deployment, and appropriation).

Undoubtedly, technological innovation can (and, arguably, often does) result in practice development. If a novel artefact addresses a real problem, if it is successfully designed, implemented, deployed, and appropriated, then the new technology may indeed transform existing practices. This best case scenario, however, does not necessarily come true. The designer can only create pre-conditions for practice development but a positive outcome cannot be guaranteed, as testified by a high ratio of unsuccessful products.

On the other hand, a successful practice development may result from an appropriation of a well-known, rather than novel, technology. When an individual or organization experiences a need for a new technology, a reasonable strategy, which is often successful, is to first look at already available tools, rather than immediately starting to design a new one.

The prevalence of technological innovation as the main orientation in interaction design (cf. [12]) has been challenged in recent years by the advent of ubiquitous computing. Ubiquitous computing does aim at developing a variety of novel artefacts. However, new ubicomp artefacts can be just “construction kits” including well known components, such as sensors, actuators, RFID tags, connectors, processors, controls, and so forth. The real design problem, solving which is the main responsibility of end-users rather than professional designers (or, perhaps, new kinds of designers, who are much more closely connected with the end-user situation), is to configure the components so that the technology brings about positive changes in human practices. Helping people solve these design problems and thus contribute to practice development, – not only technological innovation, – should be given a higher priority in HCI and interaction design (e.g., [24, 39]). There are some indications that researchers have started to acknowledge the importance of understanding and supporting users as interaction designers. For instance, a recent article by Oulasvirta [36] in the “interactions” magazine (see also [37]) is evocatively entitled “When users “do” the ubicomp”.

In sum, while technological innovation and practice development are closely related to one another, they are in fact two different phenomena, relatively independent of one another. On the one hand, for a variety of reasons, technological innovation may fail to make an impact on human practice. On the other hand, practice development may be a result of an appropriation of an existing, rather than novel, technology.

Therefore, the discussion in this section indicates that technological innovation is too limited an objective to guide all research in the field of interaction design. There is a need to explore a wider variety of the ways in which interaction design can

contribute to practice development. Such an exploration, as argued in the sections below, means deviating somewhat from a traditional view of interaction design as primarily comprised of activities located somewhere between identifying a need for a new technology and a successful appropriation of the technology.

There have been a number of attempts in interaction design research to take a broader view on the relationship between technological innovation and practice development. One of them is participatory design, or PD.

PARTICIPATORY DESIGN AS A PRODUCT DESIGN METHODOLOGY OR A MEANS FOR LOCAL CHANGE? A SHORT HISTORICAL DIGRESSION

As mentioned above, a range of frameworks, from end-user programming (e.g., [34]) to participatory design (e.g., [9, 14, 29]), have not been limiting their concerns to technological innovation and have been aimed at helping people transform and develop their everyday practices. In this section we specifically focus on Scandinavian participatory design, or cooperative design¹, as a key example of a design approach dealing with practice development. It should be noted that it is not our aim to provide a complete overview of all pertinent research. Some other approaches and directions of research, such as certain studies in the field of computer supported cooperative work, are of relevance but are not discussed here because of space limitations. Moreover, the short history of participatory design, presented below is by no means comprehensive. The discussion presents a broad picture of the development of technological innovation and practice development concerns in the history of participatory design; it is admittedly selective and incomplete.

The issue of practice development was high on participatory design's research agenda from the very outset; the interest in the issue can be found in the very bloodline of the approach. While a uniquely Scandinavian phenomenon, the approach was influenced by some other previous approaches, most notably sociotechnical systems design developed at the Tavistock Institute of Human Relations [32]. The sociotechnical systems design approach should be credited with raising the interest of the international research community in the relationship between technology and social organization, participative work design, the danger of de-skilling, and so forth. The approach had an explicit aim of transforming work practices to achieve increased productivity, learning, and workers' participation in work design. A number of conceptual tools have been developed within the approach to support the design of work. They included, for instance, a set of steps and processes to follow, comprised of analysis and evaluation of the workflow at an organization, the social system, workers' perception of their roles, and so forth, and ended up with suggestions for change. Another tool developed by the Tavistock group was a set of principles for good design. It should be noted that all these design tools, steps and principles, were all dealing with the *design of work, not specific technologies*. Creating novel technologies, while important, was not an inherent part of work design. The principles of the approach suggest that a decision regarding whether or not a *novel* technology needs to be created should be subordinated to overall aims of designing sociotechnical systems.

Of course, Scandinavian participatory design was not a direct continuation of sociotechnical systems design. The former, in fact, partly developed as an alternative to the latter and, as noted by Kensing et al. [30] sociotechnical systems design was "... heavily critiqued by Scandinavian researchers involved in trade union projects in mid- to late 1970s." The approaches differed in a number of significant ways, most notably in their political orientations, at the outset. Another of the differences, especially relevant in the context of this discussion, is that, as opposed to sociotechnical systems design, many of the later participatory design projects, at least since mid-1980s, specifically focussed on designing *technologies*.

In the early days of PD, during the 1970s and early 1980s, the attitude toward technology was, according to Ehn [13] *reactive* rather than *proactive*. Researchers, who joined forces with workers and trade unions and acted as resource persons, analyzed the effects of the introduction of IT at the workplace and attempted to build technical and organizational competence that would allow workers to have a voice in discussions and increase their bargaining (and, ideally, decision making) power regarding the transformation of their workplaces and work practices. In these studies the focus on technology virtually coincided with a focus on practice development. The aim was to conduct research and act in local contexts in order to recognize and counter potential dangers associated with work practice changes, caused by the introduction of information technologies.

Since the early/mid 1980s participatory design projects have been increasingly focusing on worker participation as a way to *design* IT-applications that "really support work" [28], with the UTOPIA project [8] as a prime example of this new orientation of research and development. Moving into this new phase of PD, and transforming the approach into a proper *design* approach, had several important consequences. A number of novel concepts and methods have been developed to support technology design activities carried out by designers together with workers. Participatory design became hugely influential internationally and made a significant impact on research and practice related to a wide range of information technologies. At the same time, the concern about practice development, as we argue below, had moved down a notch on the priorities list of the approach.

First, if a technology design perspective as adopted, then whether or not the design of a new technology is on the agenda can be a critically important argument when considering a case as suitable (or otherwise) for initiating a participatory design project. Therefore, the focus on design implies that certain types of practice development – namely, those associated with creating new technologies – are more likely to be chosen for examination than other types, which can be supported with existing technologies.

Second, adopting technology innovation as a key objective of the approach implied a dual view of human practice: as a *source* and *target* of design. Understanding an existing practice provides an input for designing a new technology. At a later point in time, when the technology is designed, implemented, and deployed, it can make an impact on human practice – this practice then becoming a target for deploying the newly designed technology. Transforming a source practice into a target practice in a local context within a reasonable amount of time can be considered a case of continuous practice development. If, however, the source practice and target practice are significantly different in place and/or time, they do not represent two phases of the development of the same practice and the practice development process is discontinuous.

¹ In this paper we use "PD", "participatory design", and "cooperative design" interchangeably.

The dual view of practice, associated with adopting technological innovation as the objective of participatory design, may cause a duality in the role of participant users. If there is a significant gap between the source and target practices, whether in place or time, the participants are not contributing to the transformation of *their own* practices. Instead, they are doing additional work, – and take time from their immediate tasks and responsibilities – the benefits of which will be enjoyed by somebody else (cf. [20]), which raises questions about the motivation and commitment of the participants.

The participatory design literature is somewhat unclear when it comes to the specific benefits of PD projects for participating users. For instance, according to Kensing and Blomberg [29], if the primary recipient group for a PD project are workers and worker organizations, then:

Workers benefit by having (1) systems that better fit the ways in which they (want to) work, (2) more participatory design practices and methods, and (3) strategies for successful integration of new technology designs. ([29], p. 178)

However, with a few notable exceptions, PD projects rarely reach implementation and deployment stages, at which workers could actually benefit from having new systems and successful strategies of integrating these systems into their work practices. And even if a project does reach these stages, the time span of a PD project can be months or even years (see [8, 9]), so that the workers who enjoy the benefits of a new system and integration strategies could differ from the workers who contributed to the design of the system.

More recent developments within the participatory design approach appear to take into account the issues discussed above, and deal with them by making some strategic decisions regarding the scope and outcomes of the approach. For instance, the MUST method for participatory design, developed by Kensing, Simonsen, and Bødker [30], explicitly distances itself from developing generic products for large markets and focuses on particular types of projects, namely, in-house/custom development or competitive bid/contract development of information technologies for complex organizational contexts. Other distinctive features of the method include involving managers and only focusing on early phases of the development process.

MUST manages to maintain a double focus on technological innovation (product design) and practice development (local change) by using a strategy, which combines (a) specifically limiting itself to a niche set of products, that is, unique, “one of a kind” technological products custom-made for complex organizations, and (b) supporting practice development in relatively large local contexts, defined in organizational terms, which helps ensure the continuity of local change. At the same time, leaving design activities taking place during implementation and appropriation of the designed technology out of the scope of the approach, indicates that the method does not cover the whole lifecycle of the local change.

The short historical excursion above indicates that when trying to accomplish the double objective of practice development and technological innovation, PD had been facing a dilemma. Adopting a *technology design* perspective, especially when “design” refers to creating *products* for a variety of future contexts, inevitably means de-emphasizing the focus on understanding and supporting *local change*. In addition, a design perspective may downplay the importance of exploring opportunities and challenges related to appropriation of *existing technologies*. Finally, a technology design perspective

can be associated with a dual view of practice, that is, differentiating between source practice and target practice.

We do not suggest, of course, that adopting a technology design perspective somehow undermines the value of PD, which has been a most insightful and influential design approach. Instead, the discussion above is intended to illustrate the following points:

- product innovation is not a necessary condition for practice development through technological re-mediation;
- practice development through technological re-mediation implies a dual view of practice, as a “source” practice and “target” practice;
- product design and practice development are inherently different perspectives, which are difficult to balance.

These points will be further developed below.

TECHNOLOGICAL INNOVATION AND PRACTICE DEVELOPMENT: A TENTATIVE MODEL

In previous sections we argued that to understand the relationship between technological innovation and practice development one needs to take into account a complex network of interdependencies between work practices, technological tools, and designers’ activities. In this section we present a tentative model describing the relationship. The development of the model can be described as a sequence of steps toward a progressively more complex representation of the situation.

First of all, the model should represent a relationship between interactive technologies and human practices. At a high level of abstraction, a representation of the development of human practices from the point of view of interaction design should include at least two components, human practice (e.g., work practice) and design of new interactive technologies. For the purposes of the current analysis we have chosen to model technology-mediated work practice and technology design as two human activity systems, using the conceptual framework of Developmental Work Research [15, 17]. The activity systems of “work” and “design” constitute a network. The “design” activity system supplies mediating artefacts for the “work” activity system. In turn, secondary contradictions between the mediating artefact, on the one hand, and subject, object, and community, on the other hand, provide a link to the “design” activity system. The aim of designing new artefacts is to resolve contradictions in the “work” activity system. The first step of the modelling is shown in Figure 2.

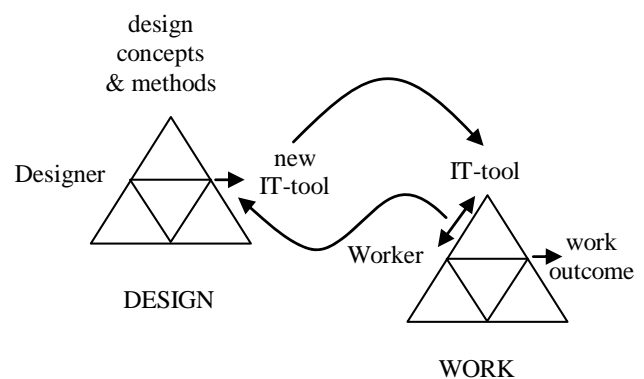


Figure 2. Step A: Activity systems of “design” and “work”.

Figure 2 includes the two most central components of the model, activity systems of work (“WORK”) and design (“PROFESSIONAL IxD”) but is incomplete in a number of important aspects. More components are added to the model in the steps that follow.

The outcomes of design are not directly supplied to workers. New technologies are distributed to users through activities of providers, such as vendors, who create a stock of novel products and then make these products available to potential users. Even if a new product is custom-made for a particular user, the user typically obtains the product from technical support people in an organization rather than directly from the designer. Accordingly, the model includes a “DISTRIBUTION” activity system, through which designers provide the outcomes of their work to users.

As discussed above, product innovation created through user-centred design implies a dual view on work practice and assumes that two types of practice are involved in analysis: the “source” practice, on which a new design is based, and the “target” practice, resulting from an appropriation of the new

product. If the “target” practice is a direct progression of the “source” practice, both of them can be represented by one and the same activity system. However, if these two practices are substantially different, they should be represented in the model as two different activity systems. Accordingly, an extended model includes two work activity systems: WORK A and WORK B.

The model also needs to take into account that in certain cases *users can act as interaction designers*, for instance, when they customize, perform end-user programming and development, integrate various digital and non-digital tools, produce “ephemeral innovations” and so forth ([18, 34, 42]). When users experience problems and need more advanced technological support, they may, instead of turning to the designer, create their own solutions that would solve their problems. Accordingly, “END-USER IxD” has been added as a separate activity system represented in the model.

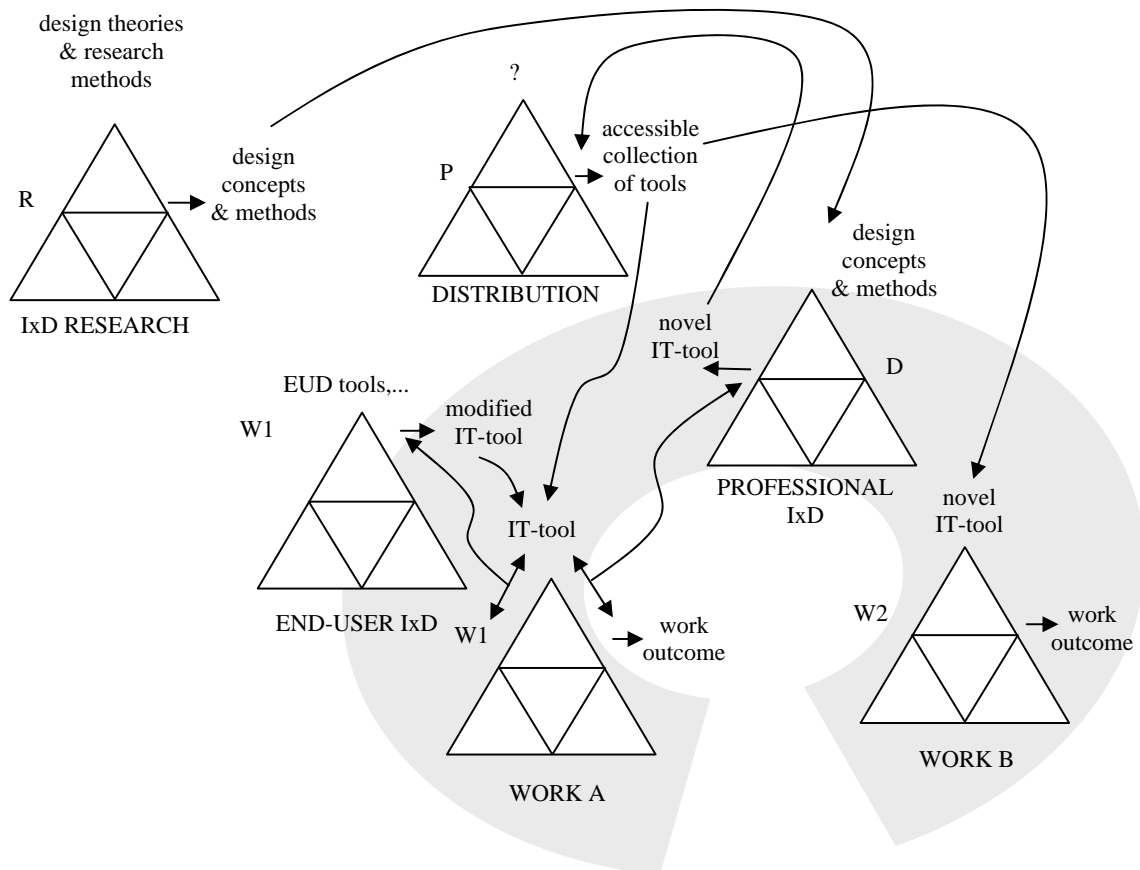


Figure 3. Step E: Activity system network of “work A” (input/outcome), “work B” (outcome), “professional design”, “end-user design”, “research”, and “distribution” (W1,W2: workers; D: designer, P: provider; R: researcher).

4.5. Finally, the model also includes interaction design *research* as a separate activity system. The “IxD RESEARCH” activity system receives inputs from other activity systems (which inputs are not shown to avoid cluttering the diagram) and provides concepts and methods for professional interaction designers².

The final model, resulting from going from step 4.1 through step 4.5, is shown in Figure 3. The model comprises a network of six activity systems: two work activity systems, two interaction design activity systems (professional designers’ and end-users’), and two additional activity systems of: (a) tool distribution and (b) interaction design research.

A typical product design lifecycle is represented in the gray arch-shaped area on the right. The lifecycle includes a source work practice WORK A, which provides an input to interaction design activity system PROFESSIONAL IxD. To address problems detected in WORK A, a novel IT-tool is designed within PROFESSIONAL IxD and eventually deployed and appropriated in a target work practice WORK B.

The model indicates that product design covers only a part of the whole picture. A number of important aspects of the complex relationship between work practices and design of interactive technologies do not fall into the category of “product design”. The model shows, for instance, that product design corresponds to just one particular case of practice development, that is, distributed development, where source practice and target practice are two different practices. And even in this case product design does not cover all aspects of practice development, since it does not include key activities involved in local changes of either WORK A or WORK B.

The model also shows that *local changes* to work practices may not involve any product design. Users may develop an innovative solution themselves (which cannot be categorized as *product design*), or obtain, through the “Distribution” activity system, an existing technology, which would meet their needs. In fact, whether or not a technology is novel does not play a significant role from the point of view of local practices; what really matters is whether the technology used is appropriate for the resulting practices.

In sum, the model introduced in this section indicates that technological innovation represents only one aspect of the range of phenomena, activities, issues, and problems, which are associated with practice development stimulated by design, appropriation, and use of interactive technologies. The conclusion about substantial differences between product design and practice development/local change, made in the previous section, can be supported with additional arguments. Product design and local change are two distinct perspectives in interaction design; they highlight different patterns in the relation between interaction design and human practice. At the same time, these two perspectives are complementary rather than mutually exclusive: each of the perspectives represents a part of the whole picture.

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

It is not an overstatement to say that research in the field of interaction design has important implications for the development of a wide range of human practices. In recent decades interactive technologies have dramatically changed the way millions of people work, learn, socialize, and play. It is often assumed, explicitly or implicitly, that the contribution of interaction design to the development of human practice is, basically, limited to creating (or supporting the creation of) novel interactive products. The discussion above challenges this assumption. While critically important, product design does not cover the whole range of possible ways, in which interaction design can make an impact on human practice.

In this paper we argue that product design represents just one perspective on interaction design. Another, substantially different but equally important, perspective is supporting local change through practice development. This, rather abstract, conceptual distinction has a concrete implication for research and practice in the field of interaction design. It points to the need to pay close attention not only to how to create novel technologies, but also to how to support people in exploiting the full potential of existing interactive technologies.

Adopting practice development as a legitimate perspective in interaction design research foregrounds a set of issues and research questions, which so far have not received sufficient attention. The discussion in the paper specifically highlights two such issues: (a) developing interaction design competence in end-users, and (b) supporting people in locating and appropriating existing technologies.

Developing interaction design competence in end-users. While, as mentioned above, end-users increasingly often act as interaction designers and are faced with a range of tasks related to extending, customizing, and integrating interactive technologies [36, 39], there is a difference between end-users and professional interaction designers. The latter employ a variety of concepts, methods, approaches, and tools, which they learn as a part of their education or professional self-development. Most end-users do not have a background in interaction design, which may mean that they lack competence necessary to carry out their tasks, and that they would probably benefit from learning about interaction design concepts and methods. Identifying the specific aims, content and appropriate forms of such learning for different categories of end-users appears to be a promising direction for future research.

Supporting people in locating and appropriating existing technologies. A couple of decades ago helping people improve their work practices through integrating a digital technology often meant developing the technology from scratch. The situation is quite different today. Thousands and thousands of potentially useful tools are not only potentially available, but are also often free or very inexpensive. Therefore, if an interactive technology used by an individual or a group does not provide sufficient support to their activities, it is likely that there is no need to develop a new technological artefact to address the problem. Increasingly often, what people need from technology experts is not developing a new technology. Because of the enormous selection of already developed IT-tools, users rather require technology experts’ help in finding and customizing the tools they need. In particular, technology experts may help the users in realizing that there is indeed a problem with existing technology, articulating their specific needs and requirements, and developing a clear idea of what type of technology would better fit the activity in question,

² In fact, the model should also include the activity system of “Development”, since designs of new products should be implemented to become usable products, but we omit this component for the sake of simplicity.

locating and comparing available technologies, and selecting the best alternative. Currently, these activities are often considered as being outside the scope of interaction design. Arguably, however, they cannot be properly addressed without the expertise developed in the field of interaction design. The discussion in this paper is intended as a step toward formulating a concrete agenda for future interaction design research, which will include helping people develop their own local practices as a key objective of the field.

The turn of the interaction design field towards local contexts and development of local practices, suggested in this paper, will require a revision of the conceptual foundations of the field. In recent years, there has been a growth of alternative approaches to investigating such issues. The open-source movement, hacker groups, self-help community fora, etc. are all seedbeds of new concepts and ideas as to how to “open-up” the design process and provide local competence and knowledge to be deployed in using new technologies for socially-relevant activities (see for example, the Bricolabs network activities, www.bricolabs.net). Another approach, outlined by Barab et al. [5], introduces the notion of “critical design ethnography”, which is defined as “...an ethnographic process involving participatory design work aimed at transforming a local context while producing an instructional design that can be used in multiple contexts”. Critical design ethnography is an example of a design approach, which also pays attention to transforming local practices. “Design” in this case means “instructional design” rather than “interaction design”, but the same strategy can, arguably, be employed in interaction design, as well.

A number of conceptual approaches, specifically oriented toward practice development, can be considered promising theoretical frameworks for understanding the effect of interactive technologies on human practice (e.g., [1, 4, 17, 33]). However, with some probable exceptions, earlier attempts to integrate such approaches with interaction design frameworks have not been particularly successful. The discussion in this paper suggests that the lack of success can be, at least partly, attributed to the fact that the dominant perspective in interaction design has been product design. Achieving a more balanced orientation of the field by complementing product design concerns with concerns about local change, will hopefully provide a precondition for more fruitful contribution of practice development frameworks to research and practice of interaction design.

REFERENCES

- [1] Argyrys, C. and Schön, D. 1996. *Organizational Learning II: Theory, Method and Practice*. Reading, MA: Addison-Wesley.
- [2] Bannon, L. and Bødker, S. 1991. Beyond the interface: Encountering artifacts in use. In: J. Carroll (Ed.). *Designing Interaction: Psychology at the Human-Computer Interface*. Cambridge: Cambridge University Press.
- [3] Bagnara, S., Pozzi, S., Rizzo, A. and Wright, P. (Eds.). 2002. *Cognition and design*, Proceedings of the 11th European Conference on Cognitive Ergonomics. Catania, Italy, September 2002.
- [4] Barab, S. and Squire, K. 2004. Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1–14.
- [5] Barab, S. A., Thomas, M. K., Dodge, T., Squire, K. and Newell, M. 2004. Critical design ethnography: Designing for change. *Anthropology and Education Quarterly*, 35(2), 254–268.
- [6] Beyer, H. and Holtzblatt, K. 1998. *Contextual design: A customer-centered approach to systems design*. San Francisco, CA: Morgan Kaufmann.
- [7] Bødker, S. and Buur, J. 2002. The design collaboratorium: a place for usability design. *ACM TOCHI*, 9(2), 152–169. DOI = <http://doi.acm.org/10.1145/513665.513670>.
- [8] Bødker, S., Ehn, P., Kammersgaard, J., Kyng, M. and Sundblad, Y. 1987. A Utopian experience. In: G. Bjerknes, P. Ehn and M. Kyng (Eds.). *Computers and democracy – a Scandinavian challenge*, Aldershot, UK: Avebury.
- [9] Bødker, S., Grønbaek, K. and Kyng, M. 1993. Cooperative design: Techniques and experiences from the Scandinavian scene. In: D. Schuler and A. Namioka (Eds.). *Participatory Design: Principles and Practices*. Hillsdale, NJ: Lawrence Erlbaum.
- [10] Clot, Y., Fernandes, G. and Carles, L. 2002. Crossed self-confrontation in the “clinic of activity”. Proceedings of the 11th European Conference on Cognitive Ergonomics (Catania, Italy, September 2002). Pp. 13–18.
- [11] Dittrich, Y., Floyd, C. and Klischewski, R. (Eds.). 2002. *Social Thinking – Software Practice*. MIT Press. Cambridge, Mass.
- [12] Dourish, P. 2006. Implications for design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Montréal, Québec, Canada, April 22–27, 2006. CHI 2006. DOI = <http://doi.acm.org/10.1145/1124772.1124855>.
- [13] Ehn, P. 1988. *Work-Oriented Artifact Design*. Stockholm: Arbetslivscentrum.
- [14] Ehn, P. and Kyng, M. 1987. The collective resource approach to systems design. In: *Computers and Democracy*, G. Bjerknes, P. Ehn, and M. Kyng (Eds.). Avebury, Aldershot.
- [15] Engeström, Y. 1990. *Working, Learning, and Imagining: Twelve Studies in Activity Theory*. Helsinki: Orienta-Konsultit Oy.
- [16] Engeström, Y., Virkkunen, J., Helle, M., Pihlaja, J. and Poikela, R. 1996. The Change laboratory as a tool for transforming work. *Lifelong Learning in Europe*, 1(2).
- [17] Engeström, Y. 1999. Activity theory and individual and social transformation. In: Y. Engeström, R. Miettinen and R. Punamäki (Eds.). *Perspectives on Activity Theory*. Cambridge: Cambridge University Press.
- [18] Fischer, G., Giaccardi, E., Ye, Y., Sutcliffe, A. G. and Mehndjiev, N. 2004. Meta-design: A manifesto for end-user development. *Communications of the ACM*, 47(9), 33–37. DOI = <http://doi.acm.org/10.1145/1015864.1015884>.
- [19] Forlizzi, J. 2007. The Product Ecology: Understanding Social Product Use and Supporting Design Culture. *International Journal of Design*, 2(1), 11–20.
- [20] Grudin, J. 1994. Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM*, 37(1), 92–105. DOI = <http://doi.acm.org/10.1145/175222.175230>.
- [21] Guliksen, J. and Göransson, B. 2001. Reengineering the systems development process for user centred design. In: Proceedings of INTERACT 2001. Hirose, M. (Ed.). EOS Press.

- [22] Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, B. and Cajander Å. 2003. Key principles for user centred systems design. *Behaviour and Information Technology*, 22(6), 397–409.
- [23] Helle, M. 2000. Disturbances and contradictions as tools for understanding work in the newsroom. *Scandinavian Journal of Information Systems*, 12, 81–114.
- [24] Henderson, A. 1991. A development perspective on interface, design and theory. In: Carroll, J. (Ed.). *Designing Interaction: Psychology at the Human-Computer Interface*. Cambridge: CUP.
- [25] Henderson, A. and Kyng, M. 1992. There's no place like home: continuing design in use. *Design at work: cooperative design of computer systems*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.
- [26] Höök, K. 2006. Designing familiar open surfaces. *Proceedings of NordiCHI 2006*. Pp. 242–251.
- [27] Kaptelinin, V. and Nardi, B. 2006. *Acting with Technology: Activity Theory and Interaction Design*. MIT Press, Cambridge, Mass.
- [28] Kensing, F. 2003. *Methods and Practices in Participatory Design*. Copenhagen: ITU Press.
- [29] Kensing, F. and Blomberg, J. 1998. Participatory design: Issues and concerns. *CSCW Journal*, 7(3–4), 167–185.
- [30] Kensing, F., Simonsen, K. and Bødker, K. 1998. MUST – a method for participatory design. *Human-Computer Interaction*, 13(2), 167–198.
- [31] Krippendorff, K. 2006. *The semantic turn: A new foundation for design*. Boca Raton, FL: Taylor and Francis.
- [32] Mumford, E. 1987. Sociotechnical systems design: Evolving theory and practice. In: Computers and Democracy. Bjerknes, G., Ehn, P. and Kyng, M. (Eds.). Avebury, Aldershot.
- [33] Mwanza, D. 2002. *Towards an Activity-Oriented Design Method for HCI Research and Practice*. PhD Thesis. The Open University, United Kingdom.
- [34] Nardi, B. 1993. *A Small Matter of Programming*. MIT Press, Cambridge, Mass.
- [35] Nardi, B. and O'Day, V. 1999. *Information Ecologies: Using Technology with Heart*. Cambridge: MIT Press.
- [36] Oulasvirta, A. and Sumari, L. 2007. Mobile kits and laptop trays: managing multiple devices in mobile information work. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose, California, April 28 – May 03, 2007. CHI 2007. DOI = <http://doi.acm.org/10.1145/1240624.1240795>.
- [37] Oulasvirta, A. 2008. When users “do” the UbiComp. *interactions*, 15(2), 6–9. DOI = <http://doi.acm.org/10.1145/1340961.1340963>.
- [38] Rabardel, P. and Bourmaud, G. 2003. From computer to instrument system: A developmental perspective. *Interacting with Computers*, 15, 665–691.
- [39] Rodden, T., Crabtree, A., Hemmings, T., Koleva, B., Humble, J., Åkersson, K.-P. and Hansson, P. 2004. Between the Dazzle of a New Building and its Eventual Corpse: Assembling the Ubiquitous Home. *Proceedings of the 5th Conference on Designing interactive systems: processes, practices, methods, and techniques*. Cambridge, MA, August 01–04, 2004. DIS 04. DOI = <http://doi.acm.org/10.1145/1013115.1013127>.
- [40] Sachs, P. 1995. Transforming work: Collaboration, learning, and design. *Communications of the ACM* 38: 36–44. DOI = <http://doi.acm.org/10.1145/223248.223258>.
- [41] Sengers, P. and Gaver, W. 2006. Staying Open to Interpretation: Engaging Multiple Meanings in Design and Evaluation. *Proceedings of the 6th Conference on Designing interactive systems: processes, practices, methods, and techniques* (University Park, PA, June 26–28, 2006) DIS 06. DOI = <http://doi.acm.org/10.1145/1142405.1142422>.
- [42] Spinuzzi, C. 2003. *Tracing Genres through Organizations: A Sociocultural Approach to Information Design*. MIT Press: Cambridge, Mass.

Challenges to Cognitive Systems Engineering: Understanding Qualitative Aspects of Control Actions

Morten Lind
Department of Electrical Engineering
Technical University of Denmark
DK-2800 Kongens Lyngby
mli@elektro.dtu.dk

ABSTRACT

The paper discusses the future role of Cognitive Systems Engineering (CSE) in contributing to integrated design of process, automation and human machine systems. Existing concepts and methods of Cognitive Systems Engineering do not integrate well with control theory and industrial automation tools. It is argued that better integration may be obtained by deeper understanding of the purposes of control actions. Examples show how process functions and control purposes are integrated in Multilevel Flow Modeling. The paper concludes that these results should be considered in future developments of CSE.

Keywords

cognitive system engineering, process control, means-end analysis, multilevel flow modeling

INTRODUCTION

Current development in industrial production and large technical infrastructures show interest in increasing efficiency and safety of operations. These goals can be reached partly by optimizing the individual technologies involved i.e. the production process, the automation and the human-machine systems. But there is an increasing awareness in industry that an integrated approach is also required so that the interactions between process, automation and human machine system design can be handled properly and more efficiently. Understanding and managing these interactions are key issues both in improving efficiency and safety of industrial installations beyond existing practices and also to reduce the costs and risk of engineering such complex systems.

Cognitive System Engineering (CSE) has the potential of playing a role in this integration, but considering more than two decades of research and development in the field it has had only limited impact on the design of human machine systems in industrial production and large technical infrastructures. Control room problems such as information overload, lack of overview and alarm cascades are still unsolved problems for

industry, even though these problems were key issues on the early CSE research agenda.

The lack of industrial impact of CSE may be seen as a purely practical issue whose solution is not a concern of research but should be found by industry through innovation of research results for application in their practices. Some improvements may be done in this direction by industry, but this will not eliminate the core problem which is a lack of theoretical understanding of the relations between process, automation and human machine design problems in industrial domains. It will be shown in the paper that these relations can be understood by a formalized analysis of the purposes of process control actions performed by either by human or machine. Analysis of control purposes deal with semantics and is therefore dealing with qualitative aspects of control actions.

It is here assumed that the reader is familiar with the main principles and methods of the CSE framework [1, 2, 3, 4]. Here we will only consider topics related to the abstraction hierarchy and the decision ladder. The abstraction hierarchy is a framework for means-end analysis of work domains and the decision ladder is used for analysis of information processing and decision task involved in control and supervision. The present author has found that the abstraction hierarchy and the decision ladder of CSE do not pay sufficient attention to analysis of the purposes of control actions. In addition they do not integrate well with both theoretical and practical concepts and methods from control engineering. These problems may explain the limited impact of CSE in process industry.

MFM [9] was originally conceived as an integrated part of CSE research, but has later also developed into methodology for means-end analysis in its own right, partly as a response to limitations of "orthodox" CSE concepts such as the abstraction hierarchy and the decision ladder when applied to control problems in process industry [7, 8, 27]. In addition the high level of formalization of MFM was developed to be able to support engineering of intelligent control systems and decision support systems for human operators. The ability to represent purposes of control actions was developed to address the needs for integrating process, automation and HMI design.

It will be argued in the paper that a theoretical understanding of purposes of control actions is not only a concern for CSE but also for control engineering. The paper will discuss these challenges based on the present author's experiences with application of means-end analysis to process control problems and in particular with Multilevel Flow Modeling (MFM) which is a methodology for functional or means-end modeling of industrial plant.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

CSE and Automation Engineering

Present tools for automation engineering in industry do not support the analytical frameworks proposed by CSE (including MFM). The engineering tools provided by automation system vendors are strongly focused on implementation aspects and therefore not concerned with higher level “non tangible” functional aspects as those captured by concepts of means-end analysis. Process schemes, piping and instrumentation diagrams and control logic diagrams comprise the main tools for design of the control systems and the human machine interface. The engineering tools deliver therefore only marginal support in design of the human machine interaction. Application of CSE comprises partly of this reason a significant additional cost for process industry seeking improved solutions to the HMI design problem. The incompatibilities of CSE with control theoretical concepts mentioned below contribute also to the lack of industrial impact. The traditional educational background of control engineers in industry is control theory combined with process technology and computer science.

CSE and Control Theory

Modeling of process dynamics is an integrated aspect of a control theoretical approach to system design and is the direct equivalent to the work domain analysis (WDA) of CSE. But the means-end concepts of WDA are qualitative and do not play an immediate or obvious role in modeling process dynamics which is predominantly quantitative. It is therefore not easy to combine results from work domain analysis and modeling of process dynamics even though they seem to address the same problem – to represent the constraints of the object to be controlled.

The decision ladder suffers not from these problems of compatibility with control theory concepts. On the contrary, the distinction of the decision ladder between different stages of information processing in a control task is largely a reflection of the fundamental elements of a feedback loop. But, due to this compatibility it can be difficult to see the added value of the decision ladder in analysis of control tasks unless you consider tasks which require considerable more intelligence than what can be realized by a traditional feedback loop (such as e.g. the shortcuts).

MULTILEVEL FLOW MODELING

Multilevel Flow Modeling (MFM) is an approach to modeling goals and functions of complex industrial processes involving interactions between flows of mass, energy and information [5, 6, 7, 8, 9, 10]. MFM has been developed to support functional modeling [11] of complex dynamic processes and combines means-end analysis with whole-part decompositions to model system functions at different levels of abstraction. System functions are represented by elementary flow functions interconnected to form flow structures representing a particular goal oriented view of the system (Figure 1). Flow structures are interconnected in a multilevel representation through means-end relations, causal roles and control functions and structures. MFM is founded on fundamental concepts of action and each of the elementary flow and control functions can be seen as instances of more generic action types [12]. The views represented by the flow structures, functions, objectives and their interrelations comprise together a comprehensive model of the functional organization of the system represented as a hyper graph. It should be noted that MFM provides a formalized conceptual model of the system which supports qualitative reasoning about control situations [13, 14].

MFM has been used to represent a variety of complex dynamic processes including fossil and nuclear power generation [15, 16, 17], oil refineries [18], chemical engineering [19] and biochemical processes [20].

Application of MFM includes model based situation assessment and decision support for control room operators [21, 22], hazop analysis [23], alarm design [24] and alarm filtering [25] and planning of control actions [15, 26]. MFM is supported by knowledge based tools for model building and reasoning [10].

The MFM concepts shown in Figure 1 will be demonstrated below with a simple modeling example.

MFM and CSE

As mentioned above MFM was initially an integrated part of CSE but has since had its individual path of development. The relation between present concepts and methods of CSE and MFM has not yet been completely clarified. It is the hope of the author that the present paper will be a contribution in this direction. Some initial clarification may be obtained by considering differences in strategy for MFM and CSE research.

The overall aim of MFM has been to provide common modeling framework for design of intelligent control systems and decision support for human process supervisors. In addition the main focus of MFM research has been on modeling problems and applications within energy and chemical engineering systems. However, recent research of the present author develops also an action theoretical foundation for MFM. This research was initiated in order to address the challenges in using means-end concepts for modeling control systems in MFM but has also a potential for facilitating a systematic application of the principles of MFM to other domains [9, 11].

The insights gained by this strategy and the subsequent development of an action theoretical basis for MFM to solve the control problem offer an approach to generalization which is theoretically founded on fundamental concepts of action. The action theoretical approach is currently under development and therefore not completely assimilated in MFM (see also later).

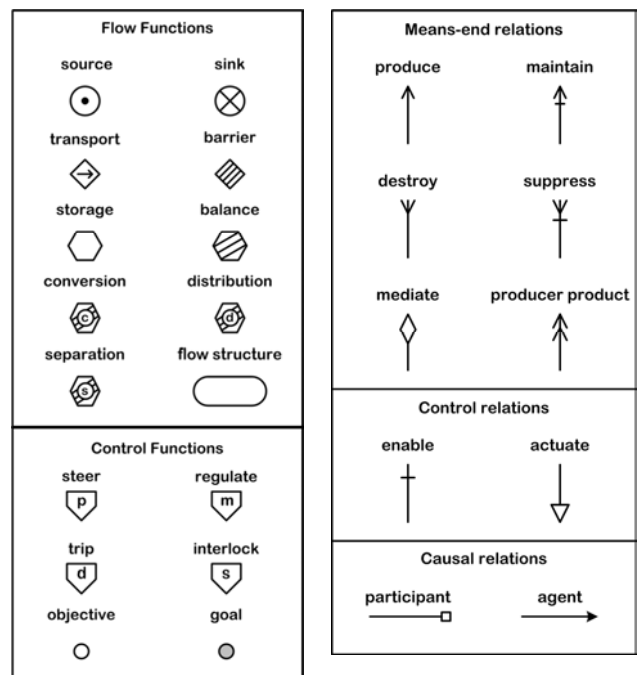


Figure 1. MFM concepts.

Somewhat in contrast with the strategy adopted in MFM research CSE has been applied to a diverse field of application domains [1] even though the original concepts and methods were developed for process control [28]. CSE has for example been developed and applied to domains such as Libraries which do not require the same deep consideration of control semantics which is of particular interest in process control and therefore for MFM.

The Importance of Control Purposes

It is not difficult to present a number of reasons why the semantics of control actions is important when applying means-end concepts in modeling work domains/control objects like industrial processes. Consider the reasons given below.

Work Domains Contain Control Agents

Most realistic work domains/control objects contains control agents which could be either human operators or automated systems. Control agents have a significant influence on the behavior and dynamics of the process and means-end analyses which ignore this fact are defective. Control principles involving several interacting control agents are especially difficult to understand without knowing the purpose each control system serve in achieving the overall control objective.

Understanding Control Behavior

The behavior of control systems is goal oriented. Knowing the goals or purpose of a control system can therefore often be sufficient for using it or for predicting its behavior. Its behavior is predictable because it is goal oriented.

Intelligent Response to Control Failures

Responding properly to control system failures require knowledge of the purpose of the failed control action. This knowledge is important for derivation of proper counteractions. Note that the purpose of the actions of a control system cannot be read from the blueprint or the control loop structure [9]. Consider an example: a simple regulator controlling the water level in a tank can have two purposes 1) it can be to maintain the water contained at a desired level or 2) it could also be to prevent the water from going over the rim. When the operator must intervene in the event of a controller failure it is important for him to know the control purpose because this knowledge is required to choose between the available means of appropriate counteraction.

Categorization of Control Means

Means of control are often categorized according to the purpose they serve. For example a valve can be used for purposes of configuration control e.g. closing or opening a pipe during process start up or shut down. In other situation the same valve may be used for control of flow rate. The purpose of the control valve depends accordingly on the situation or plant operating mode and should be known to the operator and therefore reflected in a means-end analysis of the work domain.

Summary

Even though MFM has followed its own path of development independent of other CSE research it still shares its interest in using means-end abstraction for modeling control problems and for using these models in the design of supervisory control systems. The experiences in industrial application of MFM reported here are therefore considered valid for understanding the current problems with the industrial impact of CSE.

Researchers in MFM and CSE are in the same boat so to speak due to a shared belief in the value of means-end analysis.

Some of the key results within MFM research on modeling control systems will be presented below. The results will be illustrated by a simple example and show that a significant understanding of the conceptual structure or deep semantics of control actions is required in order to see the principal relevance and power of means-end concepts for modeling control systems. The results also confirm the findings of earlier work by the author indicating that control systems play a fundamental and subtle role in the formation of functional levels in the abstraction hierarchy used by CSE for work domain analysis [27].

AN EXAMPLE

Application of the MFM concepts (Figure 1) is illustrated in the following for the simple example shown in Figure 2 below. The example is a heat transfer system with a water circulation loop and associated support system for lubrication of the circulation pump. It should be noted that the example has been selected in order to serve the specific needs of the present paper. Thus we will only consider the functions involved in circulation of lube oil and the water and ignore the functions associated with the transfer of heat through the heat exchangers. By including the means-end relations between the mass flow and energy flow functions in the heat transfer system the models would have been more complex and representative for MFM models in general. Another aspect of MFM which of the same reason is not illustrated strongly by the example is the principal differences between physical and functional topology. The interested reader can find more complex and "interesting" examples elsewhere [18, 19, 20, 23].

The water circulation loop and the lube oil system are equipped with flow measurements FM1 and FM2 and associated controllers CON1 and CON2 dealing with lube oil and water flow regulation. The purpose of the example is to demonstrate how control and process functions are integrated in the MFM models.

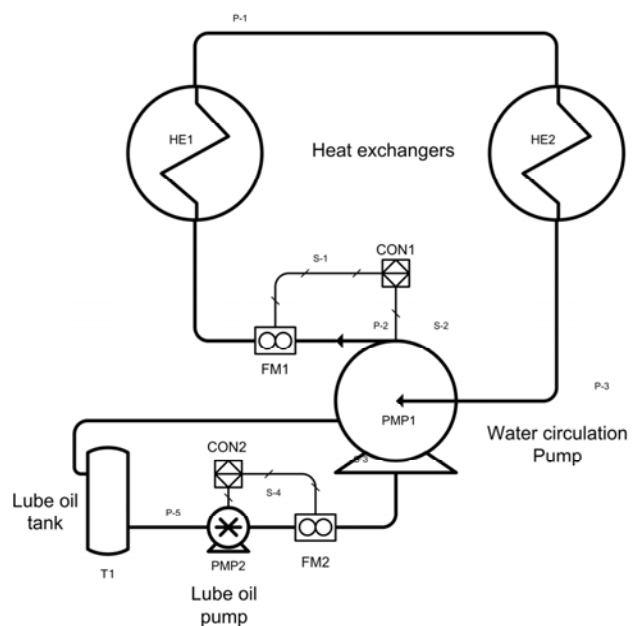


Figure 2. A heat transfer system with water circulation.

The MFM Models

We will present three models of the example system. The first model excludes the functions of the control systems. The second model show how the model is modified when the control system in the lube oil system is taken into account. The third model includes also the water flow control.

No Control Systems

The model in Figure 3 represents the objectives and functions of a water circulation loop in a heat transfer system as they are represented in MFM. The example illustrates how the MFM model provides a comprehensive understanding of the purpose and functions of the circulation loop and its subsystems. On an overall level the model can be seen as composed of three sub-models representing different views on the water circulation system.

The first view (starting from the top) represents systems aspects related to water circulation and comprises the flow structure labeled MFS1, a maintain relation and the objective O1. This part of the models represents the overall objective of the water circulation, which is to maintain a flow of water. The flow structure contains the functions provided to circulate the water. In this simplified model the transport function T1 is the means used for water circulation.

The second view is partially overlapping with the first view because what is seen here as a means (the transport T1) is in the second view seen as an end. Transport T1 is related to the means of transport which is the pumping represented by the energy flow structure EFS1. T1 and EFS1 are related by a type of means-end relation called a producer-product relation in MFM. The flow structure EFS1 is decomposed into the flow functions representing the services provided by components of the pump system (including the energy supply) in order to achieve the end, the transportation of water represented by T1.

The third view is related with the second view through the energy transport T2, an enable relation and an associated objective O2 which is the end to be maintained by the functions contained in the flow structure MFS2. The flow structure MFS2 represents the functions involved in the lubrication of the pump and the objective O2 represents the condition that should be fulfilled in order to ensure that the pump is properly lubricated. A condition which should be satisfied in order to enable the pump to provide its functions. The flow functions inside MFS2 accordingly represent the functions of the pump lubrication system.

Even though the simple example does not utilize all the concepts of MFM, it demonstrates the power of MFM to represent in a clear and logical way relations between the goals and functions of a system. The MFM modeling language has a strong syntax which defines rules for combining the different entities and relations of the language into a consistent model.

The model in Figure 3 describes the functions of the components and subsystem which contributed to the overall objective of the system (deliver water flow). No consideration was accordingly given to the purpose and function of control systems in meeting this objective. As is well known control systems are important for ensuring that process objectives are met in spite of uncertainty and disturbances in the process. This is actually one of the basic reasons for using control systems. MFM has a set of functions which can be used to represent control system functions. We will now show how these concepts are used.

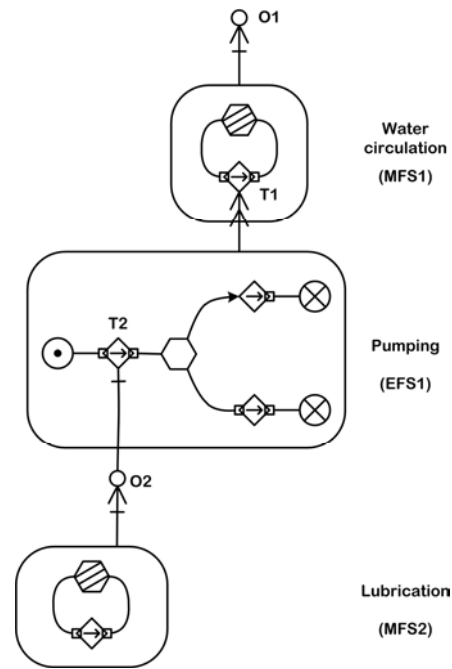


Figure 3. MFM model of the water circulation loop without control systems.

Regulation of Lubrication Flow

Assume that we need to keep the lubrication flow in the pump within specified limits in order to avoid pump problems. An engineering solution to this problem could be to use a regulator measuring the oil flow and controlling the speed of the oil pump (FM2 and CON2 in Figure 2). The function of the regulator is to maintain oil flow within limits. This function can be modeled in MFM as shown in Figure 4. The regulator function is represented by C₁.

Note that we have introduced a new objective O3 in addition to the original objective O2. It is very important to emphasize the fundamental difference between these two objectives. O2 is “process” objective specifying the value range within the lubrication flow should be kept. In contrast O3 in a “control” objective specifying the performance required of the regulated process. The control objective could specify stability margins etc. and other control attributes specifying the desired performance of the regulator [7].

It should be stressed that the “loop” formed by the maintain and the actuate relations connecting the mass flow and the control flow structures are conceptual relations and is therefore not a representation of the function or structure of a feedback loop. The concept of feedback is connected with signal or information flow but the control functions shown here do not describe information flow but the intended effect or purpose of the control action on the process (maintaining O₂).

Note also that control functions and flow functions in the flow structures representing the process under control are interdependent. This means that the inclusion of control functions in a system will influence the description made of process functions.

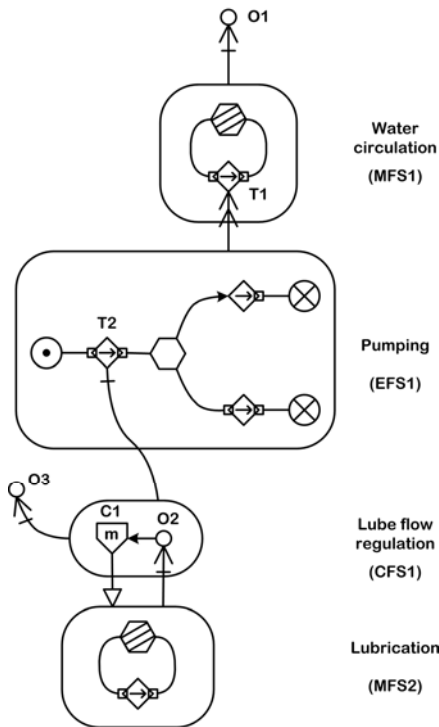


Figure 4. MFM model with lube oil regulation function.

Regulation of Water Flow

Assume that we also must to keep the water flow in the circulation loop within specified limits in order to support the heat exchange process. The solution is here to use a regulator measuring the water flow and controlling the speed of the circulation pump (FM1 and CON1 in Figure 2). The function of the regulator is to maintain water flow within specified limits. The MFM model shown in Figure 5 show how this control function can be represented by an extension of the model shown in Figure 4. The function of the water flow regulator is represented by C_2 . The actuation relation is pointing towards T_2 representing the means of control used (transport of energy to the pump rotor whose function is represented as an energy storage S_2). The objective for C_2 is represented by O_4 .

General Observations Regarding MFM

We will use the model examples in Figure 3, 4 and 5 to make some general observations regarding the nature of MFM models. This will also include relations to the abstraction hierarchy and the decision ladder of CSE and to models of dynamics used within control theory.

MFM Models are Conceptual

An MFM model is qualitative. Note that qualitative here means conceptual and not imprecise or vague because the models are formalized, have a well defined semantics and support causal reasoning. They represent a conceptualization of the process seen as an object under control. In CSE terminology they represent *conceptual* (and not parametric) constraints and structures of the “work domain”. These conceptual structures may be used for organizing the information *content* of process displays for a human supervisor (not the presentation). The obvious advantage of the organization provided is that it support reasoning about control situations in a very direct manner and make the function of control agents in the system transparent to the operator.

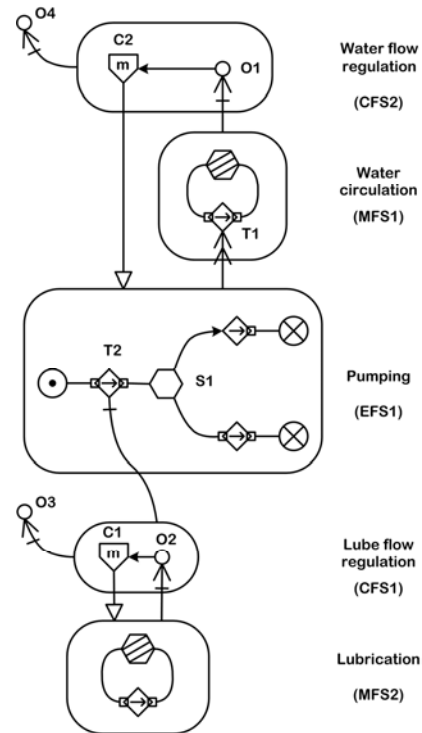


Figure 5. MFM model with lube oil and water regulation functions.

The relation between MFM models and quantitative models of dynamics is more subtle. An MFM model can here be seen as a representation of the conceptualizations which are required to define the level of granularity and abstraction in a quantitative dynamic model.

The type of conceptualization of control problems which are represented by MFM through means-end concepts are not considered explicitly in mainstream control theory. Here such qualitative information is treated informally as assumptions for the quantitative dynamic models. But MFM may prove powerful for realization of intelligent control systems which can cope with situations which require reasoning about modeling assumptions in the quantitative models.

MFM Combines Process and Control Functions

MFM models provide an integrated representation of the process and its control functions. In fact the two aspects are not separated in MFM because control functions are defined relative to the process and the process is enabled by control functions. Control functions play therefore a significant role in the constitution of levels of process functions. CSE separate the two types of knowledge in the work domain analysis (the abstraction hierarchy) and in the control task analysis (the decision ladder) and cannot therefore deal systematically with the relations between control and process functions.

MFM Models Represent Control Purposes

Purposes of control actions are explicitly represented in MFM by the effect they have on the process functions. On the other hand, MFM do not provide a representation of the information processes and decision making required for a control task. CSE use the decision ladder for this purpose. The decision ladder does not provide information about control purposes.

MFM Support Integrated Process, Control and HMI Design

The abilities mentioned above to integrate process and control knowledge and to represent the conceptual constraints of a work domain make MFM suitable as a common knowledge representation for integration of process, control and HMI design.

CONTROL SEMANTICS

It has been shown above by example that the purposes and functions of control play an important role in means-end analysis of industrial processes. For the further advancement of MFM it is therefore important to achieve a comprehensive understanding of the semantics of control actions. In the following we will briefly introduce the action theoretical foundations of MFM mentioned above and other related aspects currently under consideration by the author. These aspects seem important for understanding the semantics of control actions and are included in future developments of MFM.

Action Types

The control functions used above for representing the purpose of control systems in MFM have been developed from the logic of action proposed by VonWright[29], [9]. Von Wright derive four types of so-called elementary action types

- produce p
- maintain p
- destroy p
- suppress p.

where p is a proposition representing a state of affairs concerning the object of action.

These four types of action have been used to define the control function types in MFM shown in Figure 1. The function steer correspond to produce p, regulate to maintain p, trip to destroy p and interlock to suppress. MFM can in this way be supported by a firm theoretical foundation. With proper interpretation of the proposition p the control functions allow applications on more complex situation types such as modes during plant start up or shut down.

Counter Agency

An important aspect of Von Wrights theory of action types is that actions are defined by reference to three situations 1) the initial situation 2) the resulting situation and 3) the (counterfactual) situation which would have obtained if the action was not done. The reference to the counterfactual situation implies the notion of counter agency which is fundamental for understanding causality and the purpose of control actions (which is to suppress the effect of counter agents). Proper treatment of control actions in means-end analysis should therefore include counter agents. Handling the effect of counter agents play a significant role in more advanced control strategies such as control cascades.

Action Phases

Another significant aspect of actions relevant for means-end analysis and control is the temporal unfolding of actions in phases. The phase structure of action has been discussed by researchers interested in narratives [3]. The basic idea of action phases is to make distinctions between temporal segments where:

- there is a potentiality for action
- there is an opportunity for action
- the action is executed
- the action is completed.

The phases represent steps in the unfolding of an action. This means that each step is dependent on the realization of the previous phase. Each of these phases (further distinctions can be made) may involve aspects of control so that the unfolding of an action may involve intervention by agents which control the unfolding.

In fact, the example discussed above provides a perfect illustration of the significance of action phases for MFM and their relations with control actions. Consider T2 in Figure 5 which represents a function in the pump – the transportation of energy from the energy supply to the pump impeller. The execution of this transportation action (phase 3 above) is dependent on the availability of water in the system and that the impeller is able to rotate (note that only the last condition is represented in Figure 5). The availability of water in the system ensure that there is an *opportunity* for action (there is something to transport) i.e. phase 2. The ability of the impeller to rotate ensure that there is a *potentiality* of action (transportation is a can be done) i.e. phase 1. The purpose of the lube oil system regulator is to maintain the latter condition. This demonstrates clearly that the phase model is fundamental for understanding the formation of functional levels in MFM and that the control functions here play a crucial role.

DISCUSSIONS AND CONCLUSION

One of the concerns of MFM research has been to understand how means-end modeling complies with concepts of control theory. This question is obviously relevant and of several reasons. First of all, control engineering is the discipline which has developed the most sophisticated notions of control. A means-end modeling methodology claiming to say something new and relevant about control problems must therefore relate to that body of theory in one way or the other. Not only on at a superficial level but in its very foundations. Another more practical reason why the relation between control theory and means-end analysis is relevant is the fact that control theory is the traditional background of control engineers in industry engaged with the development, installation and operation of process control systems. If these people do not see the value of means-end analysis in solving their automation problems it will not be adopted by industry.

Unfortunately, it is not obvious how means-end analysis offer new valuable insights into control problems. This may be one of the explanations for the lack of industrial impact of CSE. However, as shown in the paper this understanding can be developed by more research in how means-end and action concepts can be used in the schematization or framing of control problems. This development should be a concern of researchers in human machine systems but also for control researchers and will also address problems in integrating process and control system design.

The Challenges

One immediate source of incompatibility between control theory and means end analysis is that the latter is predominantly qualitative and therefore for the naked eye seems to be not relevant and even inferior to control theory which has an array of quantitative methods at its disposal. But, as is concluded in the paper the problem is not the qualitative nature of means-end analysis but the lack of consideration of the qualitative aspects of control problems within control theory. On the other hand, CSE application of means-end analysis seems not to capture aspects of control which are fundamental to control theory such as the significance of counter agency.

Challenges to CSE research

The incompatibility of means-end analysis and control theory has roots in the way means-end analysis is conducted within CSE. The CSE research seems not to pay attention to the primary purpose of control actions which is either to achieve or to maintain states in the controlled process under conditions of counter agency i.e. disturbances and other uncertainties. Understanding control in this way within the framework of means end analysis would accordingly require representations of disturbances and other forms of counter agency as participants in control situations. However, CSE analyze control tasks by the decision ladder which is a qualitative representation of control as an information processing task and has therefore no direct representation of the purpose or the effect of the control action or counteragents. However, this information is essential for understanding the links between control and process design problems. Linkages between control and process design can be modeled in MFM but it is not obvious how that should be done in the present work domain and control task analyses of CSE. Work domain analysis describes the resources for control but does not give insights in the reasons why the control action is required. This explanation is given within the context of process design and would require representation of control as a means of coping with uncertainties in the operating conditions of the process and for managing its operations.

Challenges to Control Theory

Control theory is suffering from the same lack of explicit consideration of the purpose of control actions i.e. their semantics, and as a consequence hereof also the conceptual structure of control situations. Of course control engineers would know what the purposes of their systems are if you ask them but this knowledge cannot be represented by the concepts and methods of current control theory which is quantitative and do not deal with semantics. This type of information is not considered essential for their task which is to design the automated control loops. But it is becoming so if industry needs to understand how they can increase the safety and efficiency of the processes by integration of process, control and human-machine design.

Remarks on Ecological Interfaces

With this insight it is problematic that the branch of CSE called Ecological Interfaces has adopted control theoretical concepts to represent work domain constraints. Even though EID research in this way has had more success in communicating their ideas to control engineers, they have done it at the expense of a deeper understanding of the qualitative aspects of control knowledge. This view of a control problem ignore the qualitative nature of the conceptual schemas which are required for understanding the nature of control actions and their embedding in the process operations. We suggest that the constraints of the work domain also should include explicit information about the conceptual structure of control situations since this information provides the context and therefore shape the actions of the operator. These conceptual structures seem to play a more dominant role in complex industrial processes in comparison with the domains where the control challenge is perception action coordination.

REFERENCES

- [1] Rasmussen, J., Pejtersen, A. M. and Goodstein, L. P. 1994. *Cognitive Systems Engineering*. New York: Wiley-Interscience.
- [2] Vicente, K. 1999. *Cognitive Work Analysis: Toward Safe, productive, and healthy computer-based work*. Mahwah, N. J. Lawrence Erlbaum Associates.
- [3] Naikar, N., Hopcroft and Moylan, A. 2005. *Work Domain Analysis; Theoretical Concepts and Methodology*. DSTO-TR-1665 Technical Report from Air Operations Division, Defense Science and Technology Organisation, 506 Lorimer St. Fishermans Bend, Victoria 3207 Australia.
- [4] Naikar, N., Molyan, A. and Pearce, B. 2006. *Analysing activity in complex systems with cognitive work analysis: concepts, guidelines and case study for control task analysis*. *Theoretical Issues in Ergonomics Science*, 7(4), July–August, 371–394.
- [5] Lind, M. 1979. *The use of flow models for design of plant operating procedures*. In: *Proceedings of IWG/NPPCI Specialist meeting on procedures and systems for assisting an operator in normal and anomalous nuclear plant operation situations*, Garching, Germany.
- [6] Lind, M. 1981. *The use of flow models for automated plant diagnosis*. In: *Human Detection and Diagnosis of System Failures*. Rasmussen J. and Rouse, W. B. (Eds.). Plenum Press, New York.
- [7] Lind, M. 1994. *Modeling Goals and Functions of complex industrial plant*. *Applied Artificial Intelligence*, 8(2), 259–283.
- [8] Lind, M. 1999. *Plant modeling for human supervisory control*. *Transactions of the Institute of Measurement and Control*, 21(4–5), 171–180.
- [9] Lind, M. 2005. *Modeling goals and functions of control and safety systems – theoretical foundations and extensions of MFM*. Ørsted DTU, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark, NKS-R-07 Project report.
- [10] Lind, M. 2008. *Perspectives in Multilevel Flow Modeling*. In: *Proceedings of 4th International Symposium on Cognitive Systems Engineering Approach to Power Plant Control (CSEPC2008)*, Harbin Heilongjiang China, September 8–10.
- [11] Lind, M. 2007. *The what, why and how of functional modeling*. In: *Proceedings of International Symposium on Symbiotic Nuclear Power Systems for the 21st Century (ISSNP'07)*, Tsuruga Japan, July 9–11. Pp. 174–179.
- [12] Lind, M. 2005. *Modeling goals and functions of control and safety systems in MFM*. In: *Proceedings of International Workshop on Functional Modeling of Engineering Systems*, Kyoto, Japan, January 25. Pp. 1–7.
- [13] Lind, M. 1996. *Status and challenges of intelligent plant control*. *Annual Review of Control*, 20, 23–41.
- [14] Saleem, A. Us, T. and Lind, M. 2007. *Means-end based functional modeling for intelligent control: Modeling and experiments with an industrial heat pump*. In: *Proceedings of IASTED Conference on Intelligent Control Systems (ICS2007)*. Cambridge, Massachusetts, USA, November 21–23.
- [15] Larsen, M. N. 1993. *Deriving action sequences for start-up using multilevel flow models*. PhD dissertation, Department of Automation, Technical University of Denmark.
- [16] Ouyang, J., Yang, M., Yoshikawa, H., Zhou, Y. and Liu, J. 2004. *Alarm Analysis and Supervisory Control of PWR*

- Plant. In: Proceedings of Cognitive Systems Engineering in Process Control (CSEPC 2004). Sendai, Japan, November 4–5. Pp. 61–68.
- [17] Liu, J., Yoshikawa, H. and Zhou, Y. Application of Multilevel Flow Modeling to describe complex processes in a nuclear fuel cycle. In: Proceedings of Cognitive Systems Engineering in Process Control (CSEPC2004), Sendai Japan, November 4–5. Pp. 114–120.
- [18] Gofuku, A. and Tanaka, Y. 1999. Application of Derivation Technique of Possible Counteraction to an Oil Refinery Plant. In: Proceedings of 4th IJCAI Workshop on Engineering Problems for Qualitative Reasoning, Stockholm, Sweden. Pp. 77–83.
- [19] Petersen, L. W. 2005. Multilevel Flow Model of Heat Integrated Distillation Plant. MSc. Thesis, Ørsted DTU Automation, Technical University of Denmark, DK-2800 Kongens Lyngby Denmark.
- [20] Gemaey, K. V., Lind, M. and Jørgensen, S. B. 2004. Towards understanding the role and function of regulatory networks in microorganisms. In: Computer Aided Process & Product Engineering. Puigjaner, L. and Heyden, G. (Eds.). Weinheim, Germany: Wiley-VCH.
- [21] Petersen, J. 2001. Situation Assessment of Complex dynamic systems using MFM. In: Proceedings of 8th IFAC/IFIP/IFPRS/IEA Symposium on Analysis, Design and Evaluation of Human-Machine Systems, Kassel Germany, September 18–20. Pp. 645–650.
- [22] Petersen, J. 2000. Causal Reasoning based on MFM. In: Proceedings of Cognitive Systems Engineering in Process Control (CSEPC 2000). Taejon, Korea, November 22–25. Pp. 36–43.
- [23] Rossing, N. L., Lind, M., Jensen, N. and Jørgensen, S. B. 2008. A goal based methodology for hazop analysis. In: Proceedings of 4th International Symposium on Cognitive Systems Engineering Approach to Power Plant Control), Harbin Heilongjiang China, September 8–10.
- [24] Us, T., Jensen, N., Lind, M. and Jørgensen, S. B. Fundamental principles of alarm design. In: Proceedings of 4th International Symposium on Cognitive Systems Engineering Approach to Power Plant Control (CSEPC2008), Harbin Heilongjiang China, September 8–10.
- [25] Larsson, J. E. 1996. Diagnosis based on explicit means-end models. *Artificial Intelligence*, 80(1), 29–93.
- [26] Gofuku, A. and Tanaka, Y. 1997. Development of an Operator Advisory System: Finding Possible Counter Actions in Anomalous Situations. In: Proceedings 5th International Workshop on Functional Modeling of Complex Technical Systems, Paris, France, July 1–3. Pp. 87–97.
- [27] Lind, M. 2003. Making sense of the abstraction hierarchy in the power plant domain. *Cognition Technology and Work*, 5(2), 67–81.
- [28] Rasmussen, J. 1986. *Information Processing and Human Machine Interaction*. North Holland.
- [29] Von Wright, G. H. 1968. An essay in deontic logic and the general theory of action. *Acta Philosophica Fennica*, Vol. 21.
- [30] Bremond, C. 1980. The Logic of Narrative Possibilities. *New Literary History*, 11, 387–411.

Mental Contents in Analyzing Presence in Human Technology Interaction: the Approach and Preliminary Results

Pertti Saariluoma

University of Jyväskylä, Finland

Department of Computer Science and Information Systems

Fin-40014, Finland

ps@jyu.f

José J. Cañas

University of Granada, Spain

Faculty of Psychology, Campus de Cartuja

1807 Granada, Spain

delagado@ugr.es

ABSTRACT

By content-based psychology we mean psychological research, which explains behavioral phenomena on the ground of mental contents. Here, we are interested in investigating how people represent presence, which has become an important concept in HTI. Our research showed that such mental abilities factors such as spatial orientation, introversion, ability to construct mental models and empathy can predict variation on presence factors such as direct instruction, group cohesion, information integration and resolution (or problem solving) during online e-learning. The results suggest that deep level programs operating in the latter phenomena such as problem solving strategies or ability to construct social network images are connected to feeling of presence in online e-learning. Nevertheless, our study implies that it is important to develop direct measures for mental contents in presence research.

Keywords

presence, virtual learning, content-based analysis

ACM Classification Keywords

H.5.2 User Interfaces – Evaluation/methodology; H.5.3 Group and Organization Interfaces – *Web-based interaction*.

INTRODUCTION

People represent world. They constantly encode many of its aspects and they adapt their actions on the ground of the knowledge they have encoded.

This means that the contents of mental representations explain many types of psychologically relevant phenomena in investigating human technology interaction. An approach in psychology, which works to explain human behavior on the ground of the information contents in mental representations or on the ground of mental contents, can be called the psychology of mental contents or content-based psychology as the properties of mental contents form the explanatory basis of its

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

arguments (Saariluoma 1990, 2003).

The analysis of mental contents opens important new explanatory possibilities in designing human technology interaction, because it improves our understanding of the functions of mental representations (Saariluoma 2003). Content-based psychological thinking provides a perspective for us to consider how people represent the interaction situations, technologies and their own goals when using technical devices. In this way, it allows us to elaborate the analysis of many traditional problems.

Mental contents may entail many types of content elements such as unorganized sensory patterns, concepts, propositions, diagrams or stories. A good example of mental content is the notion of presence. It is an important and a much researched phenomenon and concept when ambient technologies are developed, but it is essential in all interaction with technologies (Thornson, Goldiez and Le, 2009). Presence means how people experience what they are momentarily doing, what are the relevant aspects of their experiences connecting them to their own going actions and the current phases of their performances. Presence is thus an important aspect of human experience in using technical devices for reaching their action goals.

From content-based point of view, presence is a content-element in human mental representations when they use technical devices. People represent in their minds what is the present situation, what are its major attributes and what is their position in the situation and the presence unifies such information contents in their mental representations. This is why understanding presence is possible only, when we fully understand the mental contents associated with presence. This is why it is logical

As presence has called substantial interest among the researchers of HTI, it provides a good case problem to develop content-based analysis of human technology interaction (Gilkey and Weisenberger, 1995, Thornson, Goldiez and Le, 2009). It is widely applied and intuitively presence is something that is incorporated in human mental representations. therefore, it is very sense making problem area, to develop the basic ideas of content-based research in the context of human technology interaction.

As the first step to analyze the presence in mental contents, it is essential to solve a number of problems. Firstly, it is essential to solve the foundational issues. For doing that, it is necessary to outline and to test previous approaches for analyzing presence in the context of the HTI. Then, we can explore how we can get data from mental contents, how we can analyze them and what is the logic of the empirical data analysis. And finally, we must consider how content-based way of thinking can be place among the many important alternative approaches. In this paper, we intend to solve the first part of these problems

and to show how we can study the problems in case of the problems of presence.

Some researchers have been exploring the possibility of explaining Presence by taking it as the “ability to feel it”. They assume that a person would feel presence if she has had “some training on some mental capacities”. Therefore, they take the topic of Presence as whether or not you would feel presence, but not “what are your mental contents of your feeling of presence”. The methods they had used have been based mainly on subjective questionnaires.

Most of the subjective measures of presence are pencil and paper questionnaires with items with a limited number of responses (e.g. Thomson, Goldiez and Le, 2009). These methods assume that it is possible to use qualitative and ethnographic methods or those based on subjective measures of presence (see Gilkey and Weisenberger, 1995). Participants in the assessment of presence can be extensively interviewed, either individually or in a focus group, as to the nature of their responses. Although, these measurement methods may have a high level of validity, they do not measure the contents of the feeling of presence. For example, one inspection of the methods listed by ISPR (http://www.temple.edu/ispr/frame_measure_t.htm) shows that any of those methods measure contents of presence. From the content-based approach to presence, we want to know when and how people feel presence when interacting with technology.

However, in our approach we propose that as long as we focus only on the ability to experience presence, we do not necessarily understand what the contents of presence might be. Therefore, we like to set the basic question also from the content-based point of view. This means that we are not only interested in the fact that people are presence, but we also want to know “what” they feel, when they are present and how we should understand that presence is incorporated in mental contents.

GENERAL LOGIC FOUR RESEARCH

We decided to take very conservative position and begin with the existent empirical work. The question we wanted to address in this first part of our research is whether a questionnaire developed for measuring mental ability would predict presence in a virtual environment. We select e-learning as our virtual environment since there is much research done on the role of presence already.

We took two existing questionnaires developed for analyzing presence from recent studies that have given admirably clear empirical results concerning results and therefore they seemed to give us a solid connection to the existent investigation into presence. (Arbaugh, Cleveland-Innes, Diaz, Garrison, Ice, Richardson and Swan 2008, Thomson, Goldiez and Le, 2009)..

One of these questionnaires was the one developed by Thomson, Goldiez and Le (2009) who found that presence could be in factor analytic research divided into six important factors: *Cognitive Involvement, Spatial Orientation, Introversion, Passive Cog. Involvement/Absorption, Ability to Construct Mental Models and Empathy*. Some of the factors are here cognitive, part emotional or socio-emotional (see Appendix 1).

The second one was the questionnaire developed by Arbaugh, Cleveland-Innes, Diaz, Garrison, Ice, Richardson and Swan (2008) to measure presence in the context of e-learning (see Appendix 2).

We decided to use these two questionnaires as they have their strengths. Our interest was to continue the theme of the study

of Thomson, Goldiez and Le (2009, p. 75) in which they have developed a set of factors to investigate and predict the strength of presence but left it for further investigation. Thus we presented to a number of students both questionnaires and studied to what degree we can predict presence in e-learning on the ground of on the variation of the abilities to feel presence.

PRESENCE IN VIRTUAL LEARNING

Online teaching and learning is well established issue in today educational systems and online learning environments continue to evolve. Thus, researchers working on e-learning already know that presence would be essential for designing course material, planning the teaching, and predicting learning outcomes. Their effort has already produced important theoretical models and sound empirical results.

Model of Presence in Virtual Learning

Researchers involved in the so called “Community of Inquiry Framework” (<http://communitiesofinquiry.com/>) have developed the Model of a Community of Inquiry. The model proposes that the participants in virtual learning experience three types of presence (see Figure 1):



Figure 1. Dimensions of Presence according to the Community of Inquiry Framework.

1. Social presence is the ability of learners to project their personal characteristics into the community of inquiry, thereby presenting themselves as ‘real people.’
2. Cognitive presence is the extent to which the participants in any particular configuration of a community of inquiry are able to construct meaning through sustained communication.
3. Teaching presence is defined as the design, facilitation, and direction of cognitive and social processes for the purpose of realizing personally meaningful and educational worthwhile learning outcomes.

The questionnaire developed by Arbaugh, Cleveland-Innes, Diaz, Garrison, Ice, Richardson and Swan (2008) tried to measure those factors.

A very logical continuation is to ask how well the factors of the first type explain the phenomena of the latter type. This question can be analyzed by presenting the two questionnaires

to the same group of subjects. In this kind of setting it is possible to investigate how much the six basic factors of Thornson, Goldiez and Le (2009) can explain of the variation in the study of Arbaugh, Cleveland-Innes, Diaz, Garrison, Ice, Richardson and Swan (2008). In other words we ask, whether general properties of presence can predict the three types of presences in online e-learning.

A critical word is necessary here. From our point of view the concept validity of the studies is not optimal, as they are not directly focused on the mental representation but rather presence in action. The questions concentrate on how subjects have taken their action but not how they have mentally represented their action. The questions have thus little to say directly about mental contents. Questions like “The instructor clearly communicated the important course topics”, “The instructor provided feedback in a timely fashion”, “I felt comfortable in converging through online medium” or “ I can describe ways to test and apply the knowledge created in this course” do not directly analyze involved mental contents. Of course, the present way of designing the question is not in any way problematic for studying presence in context of action, but the answers provide us only with relatively external picture of the experience. Nevertheless, this means that we have to infer issues relevant for mental contents on the ground of actions.

To investigate the possibility to predict the experience of presence, it is natural to construct a regression analytical setting. This means that factors of the first kind explain the variation of the factors of the second type. This kind of setting shows association between the items. In general, when we analyze mental contents involved in some action regression setting is logical way of investigating connections.

METHOD

Participants

Nine out one-hundreds and sixty students that took part in a course on Cognitive Ergonomics at the University of Granada, Spain volunteered to participate in the study. They take the course in the first semester of their third year in their studies for getting a B.A. in Psychology. They all used a Moodle platform during the course. This platform, called AGORA, is used by the Faculty of Psychology at the University of Granada for e-learning courses at both undergraduate and graduate levels. The main activities in AGORA are communication between teachers and students, among students themselves, exchanging of documents related to course contents, and open discussion of those contents. The platform also allows students to exchange information about course organization and planning, and social and cultural events.

Materials

We used Thornson, Goldiez and Le (2009) questionnaire to evaluate six mental capacity factors found by the authors. The questionnaire developed by Arbaugh, Cleveland-Innes, Diaz, Garrison, Ice, Richardson and Swan (2008) was used to measure presence in e-learning environment

Procedure

Participants received the questionnaires through AGORA and were asked to send them back to the second author personal e-mail address. They were instructed to fill in the questionnaire thinking in their overall experience using e-learning environment during their three years at the University of Granada.

RESULTS

We run 10 regression analyses on the data, one each dimension of presence. The predictors were those proposed by Thornson, et al. (2009). The results are presented in Table 1.

Table 1. Summary of the regression analyses.

	P1	P2	P3	P4	P5	P6
DP1						
DP2						
DP3		.76 ³	-		.99 ²	-1.2 ¹
DP4			.58 ⁴			
DP5						
DP6		.92 ¹				
DP7						
DP8						
DP9		1.1 ²	-1 ¹			
DP10		.89 ¹	-			-.39 ³
			.99 ²			

Arbaugh et al. (2008) Dimensions of Presence are in the columns. The codes for those dimensions are: Teaching Presence (DP1 = Design and Organization; DP2 = Facilitation; DP3 = Direct Instruction); Social Presence (DP4 = Affective Expression; DP5 = Open communication; DP6 = Group Cohesion); Cognitive Presence (DP7 = Triggering event; DP8 = Exploration; DP9 = Integration; DP10 = Resolution). Thornson, et al. (2009) predictors are in files. The codes for the predictors are: P1 = Cognitive Involvement; P2 = Spatial Orientation; P3 = Introversion; P4 = Passive Cog. Involvement/Absorption; P5 = Ability to Construct Mental Models; P6 = Empathy. Betas for each significant predictor are in cells. The super index in each beta means the order in which they enter in the equation. The empty cells mean that the betas were not significant.

Teaching Presence. Only the component “Direct Instruction” was predictable by some of the Thornson et al. (2009) variables. The important predictor was “Empathy”. The most emphatic participants seemed to be the ones that perceived and experienced most negatively direct instruction. Contrary, people with less evidently social orientations experienced instruction better. Presumably, emphatic and social people suffer from instruction made by computing devices.

Secondly, ability to construct mental models made subjects experience positively instruction. On the ground of concrete questions it is logical to argue that people appreciate direct instruction, because it helps them in constructing mental models. Information, which is given by a teacher, aids students in their attempts to comprehend the problems.

Thirdly, the less introvert students, i.e., extrovert, appreciated more direct instruction. This finding also makes sense. Students, who have social abilities and skills benefit more from

the community and teaching instructions. Socially directed subjects benefit from community type work.

Fourthly, subjects who are good at creating spatial representations find the instructions better. It is well known that computing presupposes mental images and therefore imagery ability is important for them (Saariluoma and Sajaniemi 1994).

Taken these results we can say that direct instruction is positively related with mental ability to create mental representation of the learning situation, whereas it is negatively related with personality characteristics such as empathy and introversion.

Social Presence. Group Cohesion was also predictable by Spatial Orientation. Participants with higher spatial orientation and ability were able to create a mental representation of group cohesion.

This is again logical when we think that computer use is often based on images (Saariluoma and Sajaniemi 1994). This means that the subjects with most vivid images have best possibilities to make a mental representation of the group. It is a spatial network of people and being able to represent on one's mental images.

Cognitive Presence. Integration was predicted by Spatial Orientation ability and negative Introversion or simply extroversion. The most able in Spatial Orientation ability were the one that integrated the course material better. However, Introversion had a negative relation with this dimension. The most introverted students were less able to integrate the material. Presumably, communities are good for extroverted people.

Finally, resolution was predicted by Empathy, Spatial Orientation and extroversion. Resolution measure how much the students could apply the knowledge acquired in the course to the solution of real problems related with the course. Spatial Orientation was the most important predictor. Participants with higher Spatial Orientation ability were more able to apply their knowledge. However, the less introvert, i.e., extrovert, and the less empathic were less able to apply their knowledge to solve real problems. This means that social orientation is a positive predictor for being able to participate to community courses.

DISCUSSION

The results do not directly open us involved mental contents as the questionnaires were not focused on these issues. However, we can find indirectly many important phenomena relevant to thinking mental contents, or the information contents of involved mental representations.

Firstly, there are systematic differences between the participants in the way they experience teaching. Of course, they get the same instruction by the same instructor. This means that on perceptual level their experience is the same, but still there are big differences between the participants in the ways they experience the interaction in the course.

The origin of the differences cannot be in the way people retinally represent the teaching, because there are no differences on this level, which would be connected to such personality features as extroversions or introversion. Instead, the differences in experiences must have their origins in the way students conceptually encode their mental representations. This process can be called apperception (Saariluoma 1990, 2003).

By apperception is meant integration of mental representations (Saariluoma 1990). The notion itself has a long history in psychology and philosophy (Kant 1781, Wundt 1920). This process often presupposes the combination of existent memory schemata with perceived information. Of course, it may be the case that no perceptual stimulus is needed in apperceiving. For

example, when we cope with such abstract mental contents as electrons, possible, future events or infinite. However, in this case the experience is really combination of cognitive and socio-emotional factors, which explain how students experience the teaching.

Our results show some preliminary content types, which are relevant in analyzing e-learning courses. Firstly, there are a number of cognitive contents. Firstly, spatial knowledge is beneficial when we think the positivity of subjects' experiences. Secondly, instructions, which help student in forming mental models, are important.

It is also interesting that socio-emotional contents play a role in forming the experience of presence. There is relatively little investigation so far on emotional and socio-emotional factors involved in e-learning (Juutinen and Saariluoma 2007). Nevertheless, their role is obvious.

One may ask here, how emotions can be mental contents or contents of mental representations. This question makes very much sense and it is important to answer to it. As a matter of fact, emotions have many important dimensions of thinking mental contents and mental states. Firstly, emotions have valence. This means that they are either positive or negative. Secondly, emotions have their themes so that fear differs from courage or joy (Lazarus and Lazarus 1994, Power and Dalgleish 1997). These types of action relevant emotional contents are important when we analyze human mental representations.

If we think this study to be concrete, it is evident that the attitudes towards teaching instructions have their emotional dimensions. Students assess whether teaching was good or bad and this type of assessment is necessarily emotional.

If we think the notion of presence, it is defined as feeling of being there. Obviously, emotional contents make sense in this type of mental state and the respective mental representations, which underlie the experience. Good and bad feelings, for example, must be seen in the prevailing action context.

It seems thus that our preliminary results enable us to build a connection between mental representations of certain content types and the feeling of presence in e-learning. This means that the results encourage constructing more directly mental contents related analysis of presence in the continuation. At least, it is possible and logical to raise these questions from the existing presence research.

ACKNOWLEDGMENTS

Our thanks to the administrators of the AGORA platform for their support in conducting our research.

REFERENCES

- [1] Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C. and Swan, K. P. 2008. Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample Internet and Higher Education, 11, 133–136.
- [2] Garrison, D. R., Anderson, T and Archer, W. 2000. Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2–3), 87–105.
- [3] Gilkey, R. H. and Weisenberger, J. M. 1995 The sense of presence for the suddenly deafened adult – implications for

virtual environments. Presence-teleoperators and virtual environments, 4, 357–363.

- [4] Juutinen, S. and Saariluoma, P. 2007. Usability and emotional obstacles in e-learning. In: (Eds.). *Managing world wide operations and communications* New York: IGI Publishing.
- [5] Kant I. 1787/1985. *Kritik der Reinen Vernunft*. Philip Reclam, Stuttgart
- [6] Lazarus, R. S. and Lazarus, B. N. 1994. *Passion & reason: Making sense of our emotions*. Oxford: Oxford University Press.
- [7] Power, M. and Dalglish, T. 1997. *Cognition and emotion. From order to disorder*. Hove: Psychology Press.
- [8] Saariluoma, P. 1990. Apperception and restructuring in chess players' problem solving. In: Gilhooly, K. J., Keane, M. T. G., Logie, R. H. and Erdos, G. (Eds.). *Lines of thought: Reflections on the Psychology of Thinking*. Wiley: London. Pp. 41–57.
- [9] Saariluoma, P. 2003. Apperception, content-based psychology and design. In: Lindeman, U. (Eds.). *Human behaviour in design* Berlin: Springer. Pp.72–78.
- [10] Saariluoma, P. and Sajaniemi, J. 1994. Transforming verbal descriptions into mathematical formulas in spreadsheet calculations. *International Journal of Human-Computer Studies* 421, 915–948.
- [11] Thomson, C. A., Goldiez, B. F. and Le, H. 2009. Predicting presence: Constructing the Tendency toward Presence Inventory. *International Journal of Human-Computer Studies*, 67, 62–78.
- [12] Wundt, W. 1913. *Grundriss der Psychologie*: Stuttgart: Kröner.

Appendix 1

A.1. FACTOR I – Cognitive Involvement (Active – Games):

1. When I'm involved with the characters in a videogame, I momentarily "forget" that the characters aren't actually real people.
2. I sometimes catch myself wondering about what will happen to the characters in a videogame.
3. Even though I'm in one location, while playing a video game it can feel as though my mind has
4. When involved with the fictional characters in a videogame, I'm able to feel what they're feeling (anger, sadness, grief, etc.)
5. When I play a videogame, it's easy for me to imagine myself as an integral part of the action.
6. When involved in a videogame, it seems as if the real world around me disappears.
7. I often find myself physically reacting to something that occurs inside the game as if it were real.
8. When I'm playing a videogame, I don't seem to realize how quickly time is going by.

A.2. FACTOR II – Spatial Orientation

1. I know I have a good sense of direction.

2. I rarely get lost.
3. When driving to a new location, even if given explicit and accurate directions, I usually make at least one mistake.
4. When I'm inside a building, I can point outside to another location I know – and be absolutely accurate.
5. After having driven somewhere once, I can find it again pretty easily.
6. Sometimes, I just seem to instinctively know which direction is north.
7. I need to consult printed directions and/or a map several times before going somewhere for the first time.
8. When someone gives me directions to a location, I can picture the map showing the route of how to get there in my mind.
9. When someone shows a new technique to me, I find I need little to no practice before I can do it on my own.
10. I lack good hand-eye coordination.

A.3. FACTOR III – Introversion

1. People often describe me as being "quiet."
2. I'm often quiet when I'm around strangers.
3. People describe me as "the life of the party."
4. I usually talk to a variety of people at parties rather than sticking to a few people I know well.
5. I'm uncomfortable meeting new people.
6. People describe me as calm or reserved.
7. I don't mind being the center of attention.
8. I prefer to keep to myself mostly, away from the scrutiny of others.

A.4. FACTOR IV – Passive Cog. Involvement/Absorption

1. When involved with fictional characters in a TV show, movie, or book, I'm able to feel what they are feeling (anger, sadness, grief, etc.)
2. When involved in a TV show, movie, or good book, it seems as if the world around me disappears.
3. When I'm watching something I enjoy, or reading a good book, I don't seem to realize how quickly time is going by.
4. When choosing a book to read (other than a textbook), I will choose fiction (science fiction, fantasy, mystery, etc.) over non-fiction (history, biographies, etc.)
5. After I'm finished watching a TV show or movie, or have read a good book, I might think about the characters and wonder what's going to happen to them now.
6. I do not enjoy spending time imagining possibilities.
7. I often played make-believe or role-playing games (house, war, etc.) as a child.

A.5. FACTOR V – Ability to Construct Mental Models

1. As a child, I loved to pull things apart to see if I could reconstruct them.
2. When I was little, I spent hours building sophisticated designs with construction toys (blocks, Lego sets, etc.) or other materials.

3. I enjoy one or more hobbies related to making things (e.g., carpentry, arts, crafts, etc.)
4. I can usually draw a pretty accurate (to scale) representation of the rooms of a house or building that I know well.
5. If I'm trying to locate an office in an unfamiliar area of town, I prefer that someone draws me a map.

A.6. FACTOR VI – Empathy

1. Most people use the words warm, compassionate, and sympathetic to describe me.
2. Taking other people's points of view into account is a top priority for me.
3. Making sure that everyone gets along in my circle of friends is one of my priorities.
4. I think I'm too tenderhearted and quick to forgive insults.

Appendix 2

Community of Inquiry Survey Instrument (draft v14)

Teaching Presence

Design & Organization

1. The instructor clearly communicated important course topics.
2. The instructor clearly communicated important course goals.
3. The instructor provided clear instructions on how to participate in course learning activities.
4. The instructor clearly communicated important due dates/time frames for learning activities.

Facilitation

5. The instructor was helpful in identifying areas of agreement and disagreement on course topics that helped me to learn.
6. The instructor was helpful in guiding the class towards understanding course topics in a way that helped me clarify my thinking.
7. The instructor helped to keep course participants engaged and participating in productive dialogue.
8. The instructor helped keep the course participants on task in a way that helped me to learn.
9. The instructor encouraged course participants to explore new concepts in this course.
10. Instructor actions reinforced the development of a sense of community among course participants.

Direct Instruction

11. The instructor helped to focus discussion on relevant issues in a way that helped me to learn.
12. The instructor provided feedback that helped me understand my strengths and weaknesses.
13. The instructor provided feedback in a timely fashion.

Social Presence

Affective Expression

14. Getting to know other course participants gave me a sense of belonging in the course.
15. I was able to form distinct impressions of some course participants.
16. Online or web-based communication is an excellent medium for social interaction.

Open Communication

17. I felt comfortable conversing through the online medium.
18. I felt comfortable participating in the course discussions.
19. I felt comfortable interacting with other course participants.

Group Cohesion

20. I felt comfortable disagreeing with other course participants while still maintaining a sense of trust.
21. I felt that my point of view was acknowledged by other course participants.
22. Online discussions help me to develop a sense of collaboration.

Cognitive Presence

Triggering Event

23. Problems posed increased my interest in course issues.
24. Course activities piqued my curiosity.
25. I felt motivated to explore content related questions.

Exploration

26. I utilized a variety of information sources to explore problems posed in this course.
27. Brainstorming and finding relevant information helped me resolve content related questions.
28. Online discussions were valuable in helping me appreciate different perspectives.

Integration

29. Combining new information helped me answer questions raised in course activities.
30. Learning activities helped me construct explanations/solutions.
31. Reflection on course content and discussions helped me understand fundamental concepts in this class.

Resolution

32. I can describe ways to test and apply the knowledge created in this course.
33. I have developed solutions to course problems that can be applied in practice.
34. I can apply the knowledge created in this course to my work or other non-class related activities.

Session 2: User Experience

Do Web Pages Have Personalities?

Phil Turner

p.turner@napier.ac.uk

Susan Turner

Centre for Interaction Design, School of Computing
Edinburgh Napier University, Edinburgh, UK
s.turner@napier.ac.uk

Leanne Wilson

l.wilson@napier.ac.uk

ABSTRACT

We are fundamentally social animals: we are geared to understanding each other; to gauging each other's moods and states of mind; and we are very adept at judging each others personalities. This ability to judge personality can also be generalized to a range of interactive technology including web sites. We present evidence that judgments of personality of different genres of website are not only internally consistent but are also correlated with perceptions of the sites' usability and aesthetics. It is proposed that this approach may be helpful in designing websites which are not only usable and attractive, but are also more predictable and better tailored to their target users. The vocabulary of personality traits should also support clearer communication between designers and clients

Classification Keywords

H.5.1.2 [User/Machine Systems].

INTRODUCTION

It is now commonplace to look beyond usability in the quest for enhanced user satisfaction. User satisfaction is usually associated with quality in use, yet defining the quality of an interface is not a simple matter. Not only may quality be perceived differently by different people, but as observed by [1] 'quality of use' as a broad concept, includes "...aspects of pleasure, fun and emotion", aspects which fall squarely in the realm of aesthetics.

This paper concerns the aesthetics of web design. We approach this complex domain at somewhat of a tangent, drawing on research in the attribution of personality to people and consumer products. There is evidence that we can judge someone's personality from a photograph of their face, that we can do that very quickly, and that this ability has value as a functional adaptation. Moreover, it has been established by design and product researchers that consumer products also have discernible personalities, that perceptions of product personalities may be described reliably and that specific personalities can be created by designers. There is also evidence that people may prefer products with personalities they judge to be similar to their own. We report the results of asking people the following question: if this webpage were a person, what kind of personality would it have?

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

THE AESTHETICS OF INTERACTIVE TECHNOLOGIES

Although intrinsically complex, the term 'aesthetics' is widely used to describe features of our everyday life; architecture, interior design as well as people, most frequently with reference to physical appearance. Aesthetics can imply anything pleasurable to the senses, but most commonly it refers to our visual senses in how we experience and see things, and their impact on our emotions. Lindgaard [2] offers a comprehensive account including a host of definitions from various authors "...beauty in appearance" [3], visual appeal [4], an experience [5], an attitude [6], a property of objects [7], a response or a judgment [8], and a process [9]. The common thread is some idea of a feeling of pleasure towards something or someone. Further, aesthetics can be regarded as possessing a dual nature "On the one hand, it is being viewed as properties of objects, and on the other, it is regarded as human responses to such properties" [10].

There has been comparatively little reported research concerning the aesthetics of interactive technologies. The reason for this is unclear: however it may be related to reluctance by those in the field of computer science to compromise matters of substance and usefulness in favour of artistic merit. While technical and analytical aspects are typical of this domain, the less tangible aspects such as 'look and feel' are sometimes disregarded [11]. Literature in this area is renowned for its complexities, perhaps another reason why many researchers tend to take a wide berth. The definition of aesthetics, as noted above, is contested. Nonetheless there have been a number of attempts to devise theories of aesthetics in this context.

Lavie and Tractinsky [3] established an empirically based classification of visual aesthetics for the web which adopts two dimensions, namely classical and expressive. Classical aesthetics are described as those steeped in history, which have reigned through changing trends. Properties include order, logical layout, and clear-cut design. Reference to classical aesthetics in interactive systems concerns properties such as colour, layout, and menu design, while expressive aesthetics on the other hand represents the judgments and feelings of the subject when interacting with the system; the user experience. The expressive dimension is associated with creativity and originality and shows little regard for design conventions. We can see that there are some contradictions between the two: indeed the authors themselves observe that "there is an intricate interplay between the cognitive and the emotional aspects of aesthetics".

In contrast, Lindgaard [11] takes a more cognitive approach, building on Berlyne's work [12, 13], based on the concept that beyond a certain level of complexity, the arousal level of the user will drop, suggesting that an experience will become less favourable as complexity is increased – in short, simplicity is preferred over complexity. Thus extra design elements can create unnecessary design problems [14]. Apple's ability to design simple products has established a reputation for straightforward,

(aesthetically pleasing) good design: a conscious effort to keep features out has allowed Apple to differentiate itself.

A more holistic conceptualization of aesthetics is presented by Petersen [15] who propose a Pragmatist framework to distinguish between approaches to aesthetics in interactive systems. They examine how aesthetics fits within our everyday lives. This approach implies that aesthetics involves a seamless integration of context, use and instrumentality. They base this on the premise that any artefact can possess aesthetic potential, but it will only be released through attachment to context and use. In this view aesthetics is not closely related to artistic properties. Rather, it is related to our experience of quality and value. The focus in the design of interactive systems shifts from promoting visual pleasure to fostering “everyday experiential qualities”. Croon Fors and Stolterman [16] also emphasise experiential aspects, arguing that “The sublime and beautiful dimension provides notions and concepts that can be used in order to frame this technology as an expressive form with an organic character...” They reject traditional boundaries of inside and outside such as those between object and subject. When describing the relationship between information technology and people, they see us as equally important participants of the design fabric as the technology itself. Technology is a mix of practical experience that occurs as one part of the composition of our life world.

Zettle [17] offers yet a further slant, with specific relevance to the web: “Aesthetics is not an abstract concept but a process by which people examine a number of visual elements, and their reactions to them”. Krauss [18] adds further detail: “The aim of visual aesthetics is to induce the user to unknowingly, unconsciously, and unsuspectingly choose to become involved in the message and the website of concern. Here aesthetics is a communication mechanism.

Empirically based studies of judgments of web aesthetics illustrate further the complexities of operationalising the concept. Park [19] report studies with both users and designers, intended to investigate “aesthetic fidelity” – the degree to which designers’ aesthetic intentions are conveyed to users – concluding *inter alia* that user perceptions of aesthetics are subject to strong individual differences. Interestingly for the personality-based approach we explore in this paper, the 13 aesthetic dimensions identified by authors include tense, strong, popular, adorable and hopeful – characteristics that are strongly redolent of human personality traits. Lack of consensus in aesthetic judgment is also noted by [20]. This study used as stimuli web pages that had been selected as exemplifying high or low levels of visual appeal, although expert designers were unable to reach a reliable consensus as to the graphical properties underlying such appeal. While participants exposed to the stimuli were able to assess visual appeal after only about 50ms (microsecond) exposure, the authors conclude that understanding the specific design factors contributing to such judgments was “too hard to do”. Speed of judgment was also recorded by Robins and Holmes [21] who found that sites with a “higher aesthetic treatment” were judged as more credible than the same sites stripped of visual enhancements. Such assessments were made in an average time of 3.42 seconds.

Finally perhaps the most comprehensive treatment of web and mobile service aesthetics thus far is the work reported by Hartmann [22]. The authors propose a framework for user judgments of design quality based on Adaptive Decision making theory, arguing strongly for the task- and context-bound nature of such judgments. Using measures of aesthetics developed by Lavie and Tractinsky [3] and the attractiveness heuristics from Sutcliffe [23], participants evaluated websites

with strongly contrasting aesthetics and the customization features of mobile services. Perceived aesthetic quality dominated overall judgment of quality, even where substantial usability problems were also apparent, but this effect was reversed in a context of serious information seeking. Substantial individual differences were however noted: the authors speculate that “there may be subgroups of aesthetically sensitive and non-aesthetically sensitive participants”.

To summarise, the definition and theorization of aesthetics is contested, both in general and in the context of interactive technologies, while the investigation of the role of aesthetic factors in the experience of such technologies is further confounded by issues of context and individual difference. However, as we shall see below, there is evidence about the way we can make rapid judgments of human personality traits which suggest it may be fruitful to treat the personality of a website as a convenient proxy for its aesthetic.

JUDGING HUMAN PERSONALITY TRAITS

The identification and description of stable, reliable personality traits have long been a concern for psychologists. Milestones in include the work of Sheldon, for example, who linked personality to body shape [24], but the first fully systematic model was that of Cattell who developed a 16 factor model, operationalised as the once widely-used 16PF Questionnaire [25]. In work roughly contemporary to that of Cattell, Eysenck developed a theory of personality comprising two dimensions, extraversion-introversion and neuroticism-stability, each associated with component traits [26]. Later researchers have failed to replicate Cattell’s 16 factors, and the consensus model now has five factors – the so-called ‘Big Five’ – extraversion and neuroticism being completed by agreeableness, conscientiousness and openness, each again comprising a number of more specific traits [27, 28]. The Big Five are also viewed by many as causal dispositions, and have been shown to be cross-culturally valid and reliable [29]. A brief summary of the five factors is shown in Table 1 below.

Table 1. The ‘Big Five’ personality traits, after Costa and McCrae (1985).

Personality traits	Characteristics
Extraversion	Sociable vs. retiring Fun-loving vs. sober
Agreeableness	Affectionate vs. reserved Soft-hearted vs. ruthless Trusting vs. suspicious
Conscientiousness	Helpful vs. uncooperative Well organised vs. disorganised Careful vs. careless
Neuroticism	Self-disciplined vs. weak-willed Worried vs. calm Insecure vs. secure
Openness	Self-pitying vs. self-satisfied Imaginative vs. down-to-earth Prefers variety vs. prefers routine Independent vs. conforming

Despite such consensus other factors continue to be investigated, as we shall see in the next section.

JUDGING PERSONALITY FROM FACES

Among the stimuli for the current work have been empirical results suggesting that judgments of personality are made with exceeding rapidity from facial appearance alone, paralleling the rapid judgments of website appeal reported by [20]. Willis and Todorov [30] show that people can make a judgment about the traits of attractiveness, likeability, trustworthiness, competence and aggressiveness based on unfamiliar faces depicted in a monochrome photograph in less than 100ms. Similarly, Penton-Voak *et al.* [31] have demonstrated that people are able to make judgments of personality in terms of the Big Five traits from pictures of faces. More recently, accurate perception of extraversion has been elicited after 50ms exposure to photographs [32]. These inferences can be characterized, in the terms of dual-process theory [33] as “fast, intuitive and unreflective”. It has been further argued that such assessments are based on quickly extracted, rudimentary information which allows us to form rapid impressions by a process of analogy with people we already know [35, 36].

Work investigating the perception of specific traits has shown that judgments of trustworthiness reflect the overall positive or negative valence of face evaluations [37] and this quality may be the subject of an automatic categorization process in the amygdala [38] which may in turn support the allocation of attention and the adoption of appropriate approach/avoidance behaviours. Of the Big Five personality dimensions, agreeableness (followed at some distance by extraversion) is the most prevalently attributed trait in open-ended judgments of people in photos, videos and face-to-face encounters [34]. The authors suggest this effect is grounded in peoples’ concern to anticipate the behaviour of others towards them and the associated need to structure relations accordingly. However, agreeableness is assessed quite inaccurately when compared to ratings made by the target subjects themselves and their friends and colleagues and to ratings of other Big Five traits.

ATTRIBUTING PERSONALITIES TO PRODUCTS

Just as in the case of human personality, it has been amply demonstrated that people can readily attribute personality traits to consumer products based on their appearance alone. Three of the Big Five personality dimensions – extraversion, agreeableness and conscientiousness – were found by Govers [39] to be salient to products, while Jordan [40] showed that consumers were able to rate photographs of vacuum cleaners, alarm clocks, kettles and toasters according to the Myers-Briggs Personality Indicator [41] dimensions: extrovert/introvert; sensible/intuitive; thinking/feeling and judgmental/perceptive. In a later small-scale study designed to elicit and apply personality descriptors which were more meaningful to the layperson, 17 dimensions – including, *inter alia*, authoritarian/liberal, bright/dim and conformist/rebel – were used to assign personalities to irons, shavers, shaver bags, epilators, air-cleaners, hair-dryers and coffee-makers, again depicted in photographs [42]. Similarly, Govers *et al.* [43] suggests that respondents could attribute happiness, cuteness and toughness to drawings of domestic consumer products, in this case, irons. The attributions made matched the personality traits intended by the product designers.

Mugge [44] developed this approach further in the derivation of a product personality scale. Their work combined items

from human personality scales with existing instruments from design and marketing studies designed to capture personality associations. These items were complemented by data from qualitative studies in which consumers were asked to describe a range of household products “as if they were a person”. The final scale items (aloof, boring, cheerful, childish, cute, dominant, easy-going, honest, idiosyncratic, interesting, lively, modest, obtrusive, open, pretty, provocative, relaxed, serious, silly and untidy) were found to be reliable in the attribution of personality to pictures of cars and vacuum cleaners.

In the domain of interactive technologies, there is also substantial evidence that people often think of and treat interactive technology as though it was their friend, a pet or another person [e.g. 45, 46] and ascribe a broad range of human attributes including personality to interactive technology (e.g. [47, 48, 49, 50, 51]).

Designers also appear to be able to design products with specific personalities, although there are fewer reports here. As noted above, Govers *et al.* ([39] report that domestic irons designed by students to embody a range of personality traits were accurately recognised by respondents, while Desmet *et al.* [52] established that devices intended to have a dominant, elegant or neutral (tangible) interaction style conveyed these traits effectively.

Product Personality Preference

The balance of evidence to date falls towards product preferences that mirror consumers’ own personalities. Jordan’s 1997 study suggests such a trend, based on participants’ self-rating of their own personality. This is also evident in Govers and Mugge [39], albeit using a 3rd party method where participants made judgements about the attachment of fictional consumers described in scenarios, to ‘extrovert’ and ‘conscientious’ toasters. Participants chose between statements such as “*This toaster has no special meaning to Susan*” and “*This toaster is very dear to Susan*” (sic).

The more extensive study reported by Govers and Schoormans [53] investigates this trend in greater depth. Forty-eight participants first described the personalities of each of several variants of screwdrivers, coffee-makers, soap-dispensers and wines “as if it were a person”, then completed a questionnaire scale designed to capture the degree of perceived similarity between their own personality and that of the product and lastly a scale capturing the perceived quality, desirability and attractiveness of the product. Products which were perceived to be similar to the participant’s own personality were significantly preferred. Finally, we should note that Jordan [42] found no such relationship between participant personality and product preference, albeit using data from a workshop with only four participants.

Personality and Design Qualities

Although there is rather less extant work which links personality traits with specific design qualities, one such study is reported by Brunel [54] and Brunel and Kumar [55]. In this instance participants rated a range of products represented in black-and-white photographs – automobiles, telephones, TVs and wall-clocks – against the five brand personality dimensions identified in Aaker [56] – sincerity, excitement, competence, sophistication, and ruggedness, so far paralleling the procedure of many other studies. However, participants were also required to rate products against the aesthetic facets of recognition, simplicity, harmony, balance, unity, dynamics, timeliness/fashion, and novelty. A significant relationship was found between each of the personality dimensions and

evaluations of aesthetic facets. Excitement, for example was related to timeliness and dynamism, while competence was associated with dynamism, unity and novelty.

JUDGING WEBSITE PERSONALITIES

Thus, people are able to judge human personalities reliably and very quickly from visual appearance alone; such judgments can be credibly applied to consumer products; and people frequently treat interactive devices as if human.

It is therefore hypothesized that people can make judgments about the ‘personality’ of websites from their appearance alone. Based on previous work, the widely-accepted Big Five account of personality is the most obvious candidate for this study. So, to amplify the opening question, “If this webpage was a person ...”, our investigation will ask people to judge this ‘personality’ in term of the traits of extroversion, agreeableness, conscientiousness, neuroticism, and open-mindedness.

Method and Participants

A booklet was prepared comprising five questions reflecting the five personality traits. These were of the form:

“If this web page were a person, I would judge its personality to be hard working and dependable (tick a box).”

Very conscientious					Not very conscientious					
+5	+4	+3	+2	+1	0	-1	-2	-3	-4	-5

In addition to these questions, people were also asked to judge the attractiveness and usability of the page using the same scale.

The participants in these studies were drawn from the second year undergraduate computing students at Edinburgh Napier University.

The number of participants was 55 (38 male and 17 female) and this yielded 48 completed, usable questionnaires.

The data collection was conducted by a small group of postgraduate students at Napier as part of their practical work on the module ‘user experience’. These students chose both the website ‘genres’ and the instance of each – three clothing retail sites, namely, (www.gap.co.uk, www.firetrap.com and www.republic.co.uk); three news sites (www.bbc.co.uk, www.yahoo.com and www.msn.com) and three social networking sites (www.facebook.com, www.myspace.com and www.youtube.com).

The home page of each website was displayed in turn to the participants using a standard lecture theatre projector for 5 minutes each.

Results

This section reports the results for each website genre. For each group of sites, we first provide the mean scores for each of the Big Five personality traits, then the correlations between individual traits and ratings for usability and attractiveness. We take attractiveness to be a simple indicator of aesthetic appeal. We also report data on the relationship between these qualities and examine correlations between usability and attractiveness.

Comparing Retail Clothing Sites

Table 2 holds the mean scores for the three retail clothing home web pages for the five personality traits. All are rated moderately positively (0 is the neutral point) for all traits except neurosis, which is rated negatively indicating its a relative absence.

Table 2. Mean scores for personality traits for clothing sites.

	Gap	Firetrap	Republic
conscientiousness	1.59	0.72	1.24
agreeableness	1.97	0.60	2.04
neurosis	-0.44	-0.98	-1.24
open mindedness	2.19	1.72	2.00
extraversion	1.13	2.06	2.56

Figure 1 is a plot of these results for the three websites. From inspection it is evident that the personality profile of each is similar.

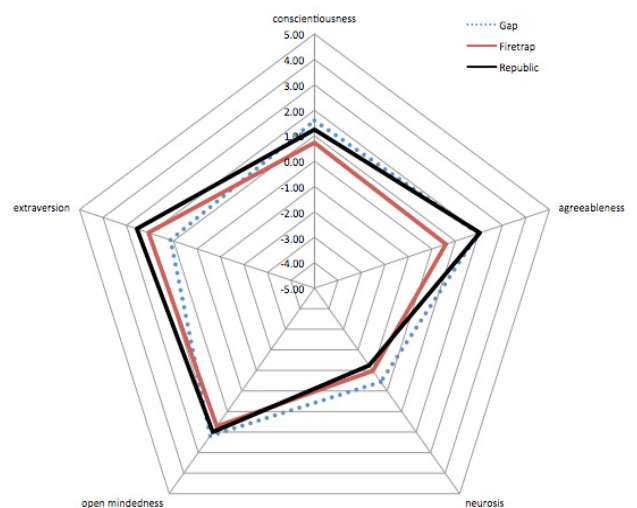


Figure 1. The ‘personalities’ of the Gap, Firetrap and Republic home web pages.

Table 3 holds details of a series of pair-wise correlations between the judgments of conscientiousness, agreeableness, neurosis, open-mindedness and extraversion and usability, attractiveness and familiarity. The figures in bold indicate significant correlations, (p < 0.05). Both agreeableness and open-mindedness are positively correlated with attractiveness in all instances. Similarly conscientiousness and (not being) neurotic are significantly correlated with attractiveness in two of the three judgments.

Usability is similarly correlated with conscientiousness for two of the three but otherwise there is no apparent pattern.

Table 3. Correlations between personality traits, usability and attractiveness for retail clothing sites.

Attractiveness	Gap	Firetrap	Republic
conscientiousness	0.23	0.55	0.76
agreeableness	0.41	0.40	0.61
neurosis	-0.40	-0.14	-0.38
open-mindedness	0.41	0.41	0.37
Extraversion	0.07	0.04	0.22
Usability			
conscientiousness	0.12	0.43	0.47
agreeableness	0.24	0.17	0.19
neurosis	-0.15	-0.17	-0.16
open-mindedness	0.18	0.36	0.17
extraversion	0.43	0.28	-0.01

Comparing News Sites

Table 4 holds the mean scores for the three news home web pages for the five personality traits. All are rated moderately positively (0 is the neutral point) for all traits except neurosis, which is rated negatively indicating its relative absence.

Table 4. Mean scores for personality traits for news sites.

	BBC	Yahoo	MSN
conscientiousness	3.06	1.81	2.76
agreeableness	0.35	1.37	0.90
neurosis	-1.23	-0.74	-0.69
open-mindedness	1.00	1.22	1.55
extraversion	0.45	1.07	0.62

Figure 2 is a plot of these three websites. Again the sites are rated positively for all traits except neurosis. From inspection it again appears that these home web pages have congruent ‘personality profiles’.

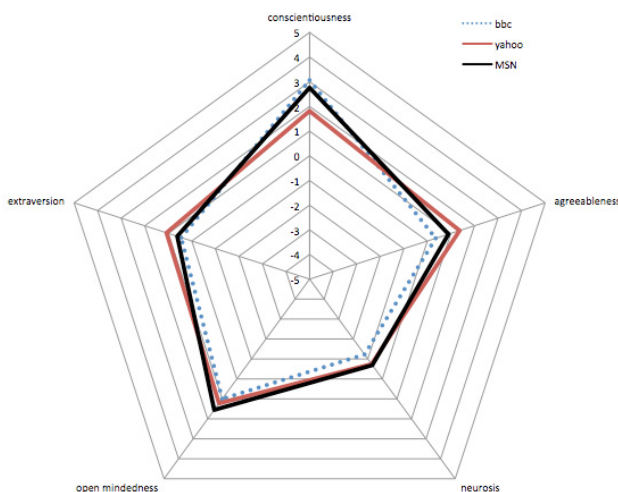


Figure 2. The ‘personalities’ of the BBC, Yahoo and MSN websites.

Table 5. Correlations between personality traits, usability and attractiveness for news sites.

Attractiveness	BBC	Yahoo	MSN
conscientiousness	0.16	0.43	0.06
agreeableness	0.00	0.59	0.56
neurosis	-0.35	0.13	-0.03
open-mindedness	0.44	0.54	0.71
extraversion	0.02	0.48	0.54
Usability			
conscientiousness	0.04	0.33	0.04
agreeableness	0.06	0.52	0.45
neurosis	-0.20	-0.23	-0.11
open-mindedness	0.56	0.50	0.85
extraversion	0.12	0.17	0.55

Table 5 holds details of a series of pair-wise correlations between the judgments of conscientiousness, agreeableness, neurosis, open-mindedness and extraversion and usability, attractiveness and familiarity. The figures in **bold** indicate significant correlations, ($p < 0.05$).

Again agreeableness is positively correlated ($p < 0.05$) with attractiveness for two sites, as is extraversion, while for all three sites open-mindedness is associated with attractiveness.

In the news context, usability and open-mindedness appear to be related for all three sites.

Comparing Social Networking Sites

Table 4 holds the mean scores for the three news home web pages for the five personality traits. All are rated moderately positively with only one negatively rated web page.

Table 6. Mean scores for personality traits for social networking sites.

	FaceBook	YouTube	MySpace
conscientiousness	1.40	0.75	0.69
agreeableness	1.31	1.19	1.44
neurosis	0.32	-0.25	0.44
open-mindedness	1.79	2.81	1.47
extraversion	0.68	3.28	2.66

Figure 3 is a plot of these three websites. Inspection suggests that while the profiles of the sites are similar for conscientiousness, agreeableness and neurosis, YouTube and Facebook in particular differ on the dimension of extraversion – YouTube being the more extraverted, while YouTube is also perceived as being more open-minded than comparison sites.

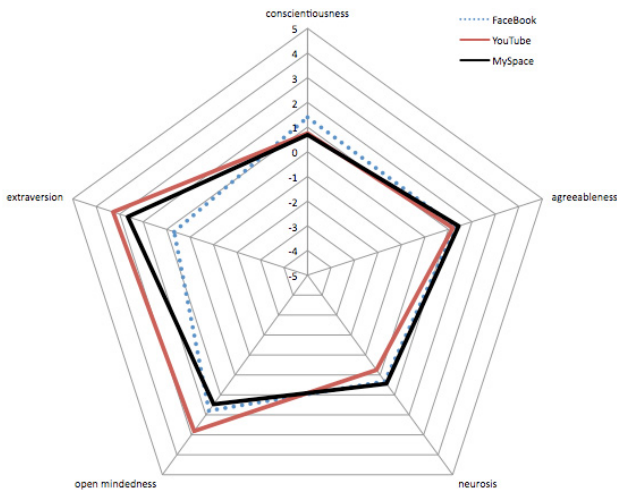


Figure 3. The ‘personalities’ of the Facebook, YouTube and MySpace websites.

Table 7 holds details of a series of pair-wise correlations between the judgments of conscientiousness, agreeableness, neurosis, open-mindedness and extraversion and usability, attractiveness and familiarity.

Open-mindedness is positively correlated ($p < 0.05$) with both attractiveness and usability. For two of the sites (Facebook and YouTube), usability is related to agreeableness, open-mindedness and extraversion.

Table 7. Correlations between personality traits, usability and attractiveness for social networking sites.

Attractiveness	Facebook	YouTube	MySpace
conscientiousness	0.22	0.22	0.24
agreeableness	0.19	0.68	0.29
neurosis	0.07	-0.31	0.11
open-mindedness	0.38	0.51	0.43
extraversion	0.48	0.08	0.17
Usability			
conscientiousness	0.04	0.16	0.25
agreeableness	0.35	0.47	0.26
neurosis	0.02	-0.28	-0.15
open-mindedness	0.51	0.65	0.62
extraversion	0.31	0.38	-0.07

DISCUSSION

The most striking finding we report is that personality does seem to be a meaningful concept in this context. We have identified four immediate uses for our findings:

Firstly, it may be possible for designers to create websites with specific personality traits which would render them more predictable, and as such, acceptable for their potential users.

Secondly, it may be possible to tailoring design for “people like us”. There is evidence that consumers prefer product personalities that accord with their own [52]. It is reasonable to suppose websites could be created to match the personality

traits of their intended users. Further, if we can design for specific personality traits then we can design for personas, thus forging a direct link between website design, website aesthetics and persona-based design

Thirdly, the use of this ‘personality profiling’ for websites may be a tool which could be used to differentiate between two designs of similar usability.

Finally, designers of websites are often faced with the challenge of talking to their client about aesthetics without having a clear, common language – hence the extensive use of iterative prototyping and “enlightened trial and error”. However, in adopting this ‘personality’ based approach there is a ready-made and completely comprehensible language which can empower designer and client alike.

ACKNOWLEDGMENTS

We would like to acknowledge the hard work of the 2009 postgraduate *User Experience* class at the School of Computing, Edinburgh Napier University, who collected the data reported here.

REFERENCES

- [1] Frøkjær, E., Hertzum, H. and Hornbæk, K. 2000. Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (The Hague, The Netherlands, April 01–06, 2000). CHI ‘00. ACM, New York, NY. Pp. 345–352.
- [2] Lindgaard, G. 2007. Aesthetics, Visual Appeal, Usability and User Satisfaction: What Do the User’s Eyes Tell the User’s Brain? Australian Journal of Emerging Technologies and Society, 5(1), 1–14.
- [3] Lavie, T. and Tractinsky, N. 2004. Assessing dimensions of perceived visual aesthetics of web sites. International Journal of Human-Computer Studies, 60(3), 269–298.
- [4] Lindgaard, G. and Dudek, C. 2003. What is this evasive beast we call user satisfaction? Interacting with Computers, 15(3), 429–452.
- [5] Ramachandran, V. S. and Blakeslee, S. 1998. Phantoms in the Brain. William Morrow & Co.
- [6] (New York.) Cupchik, G. C. 1993. Component and relational processing in aesthetics, Poetics, 22, 171–183.
- [7] Porteous, J. D. 1996. Environmental Aesthetics: Ideas, Politics and Planning. Routledge, London.
- [8] Hassenzahl, M. 2004. The interplay of beauty, goodness and usability in interactive products. Human-Computer Interaction, 19(4), 319–349.
- [9] Langer, S. 1967. Mind: An Essay on Human Feeling Vol. I. The Johns Hopkins Press, Baltimore.
- [10] Tractinsky, N. 2006. Noam Tractinsky on Aesthetics and HCI by Dr Gitte Lindgaard. Hot Topics [Online]. Publication of the Human Oriented Technology Lab, Carleton University. Available from: <http://www.carleton.ca/hotlab/hottopics/Articles/June2006-NoamTractinskyon.html>. [Accessed 13/07/09].
- [11] Lindgaard, G. and Chatratchart, J. 2007. Usability testing: what have we overlooked? Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems 2007. Pp. 1415–1424.

- [12] Berlyne, D. E. 1971. *Aesthetics and Psychobiology* Appleton-Century-Crofts: New York.
- [13] Berlyne, D. E. 1972. 'Experimental aesthetics'. In: Dodwell, P. C. (Ed.). *New Horizons in Psychology* Penguin, Harmondsworth, UK.
- [14] Lidwell, W., Holden, K. and Butler, J. 2003. *Universal Principles of Design*. Rockport Publishers: Beverly, MA.
- [15] Petersen, M. G., Iversen, O. S., Krogh, P. G. and Ludvigsen, M. 2004. Aesthetic interaction: a pragmatist's aesthetics of interactive systems. In: *Proceedings of the 5th Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques*, Cambridge, MA, USA, August 01–04, 2004. DIS '04. ACM, New York, NY. Pp. 269–276.
- [16] Croon Fors, A. and Stolterman, E. 2003. *Information Technology and Everyday Aesthetics*. The 5th European Academy of Design Conference (EAD), τέχνη – Design Wisdom, Barcelona, April, 2003.
- [17] Zettl, H. 1999. *Sight, Sound, Motion: Applied Media Aesthetics*. Wadsworth Publishing Company, USA.
- [18] Park S., Choi, D. and Kim, J. 2004. Critical factors for the aesthetic fidelity of web pages: Empirical studies with professional web designers and users. *Interacting with Computers*, 16, 2, 351–376.
- [19] Lindgaard G., Fernandes G., Dudek C. and Brown, J. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, 25(2), 115–126.
- [20] Robins, D. and Holmes, J. 2008. Aesthetics and credibility in web site design, *Information Processing and Management*, 44, 386–399.
- [21] Hartmann, J., Sutcliffe, A. and De Angeli, A. 2008. Towards a Theory of User Judgment of Aesthetics and User Interface Quality. *ACM Transactions on Computer-Human Interaction*, 15(4). (No page numbers).
- [22] Sutcliffe, A. G. 2002. Assessing the reliability of heuristic evaluation for website attractiveness and usability. *Proceedings of the Hawaii International Conference on System Sciences (HICSS 35)*. IEEE Computer Society Press, Los Alamitos, CA. Pp. 1838–1847.
- [23] Sheldon, W. 1954. *Atlas of Men: A Guide for Somatotyping the Adult Male of All Ages*. New York, Harper.
- [24] Cattell, R. B., Eber, H. W. and Tatsuoka, M. M. 1977. *Handbook for the Sixteen factor Personality Factor Questionnaire*, Champaign, (Illinois): IPAT.
- [25] Eysenck, H. J. 1975. *The Inequality of Man*, San Diego, California: Edits Publishers.
- [26] Digman, J. M. 1990. Personality structure: Emergence of the five-factor model, *Annual Review of Psychology*, 41, 417–440.
- [27] Goldberg, L. R. 1993. The structure of phenotypic personality traits, *American Psychologist*, 48, 26–34.
- [28] Schmitt, D. P. and 126 others 2006. The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations. *Journal of Cross-Cultural Psychology*, 38(2), 173–212.
- [29] Willis, J. and Todorov, A. 2006. First Impressions Making Up Your Mind after a 100-Ms Exposure to a Face. *Psychological Science*, 17(7), 592–598.
- [30] Penton-Voak, I. S., Pound N., Little A. C. and Perrett D. I. 2006. Accuracy in personality attributions made on the basis of facial characteristics. *Social Cognition*, 24, 607–640.
- [31] Borkenau, P., Brecke, S., Möttig, C. and Paelecke, M. 2009. Extraversion is accurately perceived after a 50-ms exposure to a face, *Journal of Research in Personality*, 43, 703–706.
- [32] Kahneman, D. 2003. A perspective on judgment and choice. *American Psychologist*, 58, 697–720.
- [33] Ames, D. R. and Bianchi, E. C. 2008. The Agreeableness Asymmetry in First Impressions: Perceivers' Impulse to (Mis)judge Agreeableness and How It Is Moderated by Power, *Personality and Social Psychology Bulletin*, 34, 1719–1736.
- [34] Bar, M., Neta, M. and Linz, H. 2006. Very first impressions. *Emotion* 6, 269–278.
- [35] Bar, M. 2007. The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Science*, 11(7), 280–289.
- [36] Todorov, A., Baron, S. G. and Oosterhof, N. N. 2008. Evaluating face trustworthiness: a model based approach, *SCAN*, 3, 119–127.
- [37] Engell, A. D., Haxby, J. V. and Todorov, A. 2007. Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala, *Journal of Cognitive Neuroscience*, 19(9), 1508–1519.
- [38] Govers, P. C. M. and Mugge, R. 2004. 'I love my jeep, because its tough like me': the effect of product personality congruence on product attachment. In: Kurtgözü, A. (Ed.). *Proceedings of the Fourth International Conference on Design and Emotion*, Ankara, Turkey.
- [39] Jordan, P. W. 1997. Products as personalities. In: Robertson, S. A. (Ed.). *Contemporary ergonomics*, London: Taylor & Francis. Pp. 73–78.
- [40] Myers, I. B. and McCaulley, M. H. 1985. *Manual; A Guide to the Development and Use of the Myers-Briggs Type Indicator*, 5th ed., Palo Alto, CA: Consulting Psychologists Press.
- [41] Jordan, P. W. 2002. The personalities of products. In: Green, W. S. and Jordan, P. W. (Eds.). *Pleasure with products: beyond usability*, London: Taylor & Francis. Pp. 19–48.
- [42] Govers, P. C. M., Hekkert, P. and Schoormans, J. P. L. 2004. Happy, cute and tough: can designers create a product personality that consumers understand? In: MacDonagh, D., Hekkert, P., Van Erp, J. and Gyí, D. (Eds.). *Design and emotion, the experience of everyday things*. Taylor & Francis, London. Pp. 345–349.
- [43] Mugge, R., Govers, P. C. M. and Schoormans, J. P. L. (to appear). The development and testing of a product personality scale, *Design Studies*.
- [44] Reeves, B. and Nass, C. 1996. *The Media Equation: How People Treat Computers, Television And New Media Like Real People And Places*. Cambridge, UK: Cambridge University Press.

- [45] Nass, C. and Moon, Y. 2000 Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1), 81–103.
- [46] Fogg, B. J. and Nass, C. 1997 Silicon sycophants: The effects of computers that flatter. *Int. Journal of Human-Computer Studies*, 46, 551–561.
- [47] Brave, S., Nass, C. I. and Hutchinson, K. 2005. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int. Journal of Human-Computer Studies*, 62(2), 161–178.
- [48] Johnson, J. D., Marakas, G. M. and Palmer, J. W. 2006. Differential social attributions toward computing technology: An empirical examination. *International Journal of Human-Computer Studies*, 64, 446–460.
- [49] Johnson, J. D., Marakas, G. M. and Palmer, J. W. 2008. Beliefs about the Social Roles and Capabilities of Computing Technology: Development of the Computing Technology Continuum of Perspective. *Behaviour and Information Technology*, 27(2), 169–181.
- [50] Turner, P. and Sobolewska, E. 2009. Mental Models, Magical Thinking and Individual Differences. *Human Technology*, 5(1), 90–113.
- [51] Desmet, P. M. A., Nicolás, J. C. O. and Schoormans, J. P. 2008. Product Personality in Physical Interaction. *Design Studies*, 29, 458–477.
- [52] Govers, P. C. M. and Schoormans, J. P. L. 2005. Product personality and its influence on consumer preference, *Journal of Consumer Marketing*, 22(4), 189–197.
- [53] Brunel, F. F. 2006. Design and the Big Five: linking visual product aesthetics to product personality, Boston University School of Management Working Paper.
- [54] Brunel, F. F. and Kumar, R. 2007. Design and the Big Five: linking visual product aesthetics to product personality, *Advances in consumer research*, Association for Consumer Research, 34, 238–239.
- [55] Aaker, J. 1997. Dimensions of brand personality, *Journal of Marketing Research*, 34(3), 347–356.

How Cognitive Ergonomics Can Deal with the Problem of Persuasive Interfaces: Is a Criteria-based Approach Possible?

Alexandra Nemery
University Paul Verlaine
User Experience Labs
57006 Metz (France)

alexandra.nemery@univ-metz.fr

Eric Brangier
University Paul Verlaine
User Experience Labs
57006 Metz (France)

brangier@univ-metz.fr

Steve Kopp
SAP-BusinessObjects Division
157-159, rue Anatole France
92309 Levallois-Perret (France)

steve.kopp@sap.com

ABSTRACT

Today the user experience covers areas such as usability, satisfaction and accessibility which are known as critical factors for success of user interfaces. However, studies about persuasion, relying on credibility of the product for instance, are less recognized. Our goal is to promote the introduction of persuasive methods in software elaboration through psychosocial theory, especially in the Business Intelligence area. Currently, we are proposing a criteria grid measuring persuasive dimensions in interfaces; this grid is being validated.

Keywords

user experience, persuasion, credibility, criteria grid

INTRODUCTION

Have you ever received spam in your email box that tried to sell a product or a service that you do not want? Maybe you have subscribed to an RSS feed and then regretted your decision after? Wouldn't you want to have data that would help you convince your colleagues to take the decision you find the best? As you can see, persuasion is beginning to invade our technical systems, particularly in the area of social networks. Evidently, new problems are appearing in the field of computer ergonomics. It takes into account social dimensions and put user at the center of the man-machine relationship. Some authors have developed the concept of persuasive technology to emphasize HCI influence on social conduct.

The aim of our communication is to discuss the concept of technological persuasion, to note how useful it could be during the software evaluation or conception phases and to insist on the importance of ethics in this field.

As a first step, we will define the contours of the theoretical persuasion technology. We will then propose a framework for the analysis of persuasive evidence as a tool for the design of interfaces, while discussing the ethical problems. Finally, we will conclude by referring to the prospects for validation of our proposed grid.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

THEORETICAL POINT OF VIEW

The first work using the persuasive power of technology took place in the 1970s and 1980s. It was about promoting behaviors related to health or to the improvement of the employees' productivity. But the evolution of persuasive methods really started in the 1990s with the emergence of the Internet. At the moment, web sites are the favorite media for trying to change the attitude and behavior by its rich interactive possibilities. The main contributor in this discipline is Fogg [1] who proposes to create a science named Captology. This word is based on the acronym "Computer As Persuasive Technology". The notion of captology has existed for several years (the 4th International Conference on Persuasive Technology will take place in 2009) and states a focus on behavioral change resulting from human computer interaction.

Definitions

We will define the persuasive technology and then present the framework for the analysis of persuasive elements. According to us, persuasive technology can therefore be seen as a vehicle to influence and persuade people through HCI.

Fogg believes that persuasion technology works as both (a) a tool since the technology can help individuals achieve their objective, (b) as a media interaction which creates an experience between the user and technology and (c) as a social actor. The social characteristic deals with properties to use strategies of social influence.

Our definition presents persuasion technology as an action to influence and persuade people through HCI. The impact of persuasion technology affects the fields of social work, psychology and ethics and obviously the social organization. Indeed, the technology becomes persuasive when people give it qualities and properties that may increase its legitimacy, reliability and perceived credibility. Persuasion technology is characterized by the fact that the intention of changing the attitude and behavior is subtle, hidden and sometimes pernicious. Persuasion technology is at the crossroads between ergonomics, social psychology, organizational management and of course the design of GUI.

Forms of Persuasion

Fogg [1] distinguishes between the macro and micro-persuasion. Macro-persuasion represents products whose main intent is to persuade. E-commerce websites clearly belong to this category since their main objective is to change the purchasing behavior of visitors [2]. The preventive health programs that seek to

modify dietary behavior or sexual behavior also reside under this level of persuasion. Their goal is to improve the lifestyle of the people using them. The micro-level concerns all products whose primary purpose is not to persuade but which are using methods of persuasion in order to satisfy a different objective. For instance, edutainment software is designed to teach educational information to children. To do so, elements of micro-persuasion are used such as the reward system or messages of encouragement to increase the child’s motivation to continue to play and learn.

Fields and Areas

Persuasion technology affects many areas. Technology development also initiates a diversification of applications. The rise of e-commerce websites in recent years is propitious to the use of persuasive methods, both in the field of design and ergonomics, trying to change purchasing behavior. It explains why marketing is a beacon of persuasion technology. eBay is a good example of persuasive technology. Indeed, stars to assess the seller reliability are a confidence index and tend to alter the intentions and therefore the purchasing behavior of visitors. A major new field of research concerns the field of health, both in the prevention of risk, monitoring of disease and the promotion of sport. For example in New Zealand, a video game group has been created [3] as an aid to stop smoking. The game is aimed at the Maori population and is based on elements of the collectivist culture to change or prevent smoking behavior of young Maori.

Finally, whether in education, health, consumption, entertainment and especially work, all areas are affected by persuasion technology. It is therefore important to develop ergonomic practices.

PROPOSED CRITERIA

The ergonomic computer has often produced grids used in the measurement of the ergonomic quality of goods and services [4, 5]. In this perspective, we seek to establish a grid to focus on the persuasive dimensions of interfaces and their effects; a grid that is robust, reliable, useful, relevant and easy to use for ergonomists.

Organizational Principles of the Criteria

Our proposal is based on a bibliographic analysis and draws up a grid that distinguishes forms and processes of social influence, respectively the static and dynamic aspects of the interface.

Table 1. General Articulation of the Criteria.

Static Aspects of the Interface	<ol style="list-style-type: none"> 1. Credibility of the interaction 2. Legitimacy of the system act 3. Guarantee of privacy 4. Suggestibility 5. Responsiveness 6. Social conformity 7. Displaying format that may reinforce behaviors
Dynamic Aspects of the Interface	<ol style="list-style-type: none"> 8. Invitation 9. Priming, Initiation of the users 10. Commitment 11. Freely accepted compliance 12. Ascendency and possibility of addiction

Static Aspects of the Interface

In interfaces, some prerequisites are necessary to promote the acceptance of an engaging process. These criteria are based on the content of technological influence.

Credibility of the Interaction

Definition: Giving enough information to the user allows him to identify the source of information to be reliable, expert and trustworthy.

Justification: Credibility affects use and is seen as a form of loyalty. Credibility is the combination of the perceived reliability and perceived expertise of the product [1]

Example: Presenting updated information and the date of the update.

Legitimacy of the System Act

Definition: Reassure the user by justifying the influence of interface elements and increasing the stability and understanding of the interaction goals.

Justification: If the user sees the interface as legitimate, it will be easier for him to accept its influence [6].

Example: Famous brand could be perceived as a moral strength.

Guarantee of Privacy

Definition: Do not persuade the user to do something that publicly exposes his private life and which he would not consent to do.

Justification: Privacy is an important point about ethics. Respect for human rights must be respected by persuasive technologies [7].

Example: Private question about religion or politics orientation must be avoided.

Suggestibility

Definition: Present indirect and non-intrusive elements with incentives.

Justification: The suggestibility is to introduce into the mind of the user some elements that could affect them later [8].

Example: Interface elements that induce incentives to perform an action.

Responsiveness

Definition: Adapt the form of persuasion to the user to make it more likely that the user will respond in a desired way.

Justification: Acting on a willingness to accept is a predictor of the internalization of the persuasive message [9]

Example: Recognize the user and welcome him by his name.

Social Conformity

Definition: Social conformity provides information about other users in order to converge the views and behavior of the user.

Justification: It reflects a need for accuracy and confirmation of individuals. People tend to act in the same way as a person to whom they look similar [10].

Example: Emphasize the social bonds of trust to strengthen the adherence of the user.

Displaying a Format that May Reinforce Behavior

Definition: Strengthening surface design and the presentation of persuasive interactive elements, while taking into account the perceptual and cognitive characteristics of the user. Prompt the user to do what is expected of him.

Justification: Enhancing surface is related to the persuasive design. Controlling the physical elements of the interface and maximizing the visual impact can cause membership and create or reinforce a behavior [11].

Example: The choice of colors as a reinforcement of the message.

Dynamic Aspects of the Interface

Regarding dynamics, there is also a means to bring the user in a process of interaction to strengthen the progressive engagement of the user to the elements of its interface.

Invitation

Definition: To identify the profile of the user information in order first to suggest a personalized approach and more likely to correspond to its needs.

Justification: Using the information given by the user allows for elements of hook and plan a sequence of engagement [12].

Example: Display a welcoming message.

Priming, Initiation of the Users

Definition: Triggering the first boot-engaging action of the user by creating a point of entry, stirring interest.

Justification: In social psychology, the notion of commitment [13] is initiated by a first act which is inexpensive and is consented to be done.

Example: A free way to subscribe to an offer.

Commitment

Definition: Continue commitment to involve the individual through a process of accession to the objectives of the GUI.

Justification: Having accepted an inexpensive first step, it will be easier to accept the following steps each time increasing the persuasive force [14].

Example: Improve the frequency of the final behavior or attitude expected.

Freely Accepted Compliance

Definition: Expanding spiral binding sequences increases their frequency and impact force. Maintain interaction and capture the user.

Justification: By segmenting the persuasive message, we follow the thinking of the user. Tunneling is how to assist the user during the process of persuasion [1].

Example: Continue to catch his attention by frequent solicitation.

Ascendency and Possibility of Addiction

Definition: Show engaging scenario completion, follow up its influence and control its evolution over time.

Justification: The last step is the culmination of the process leading to behavior and attitude initially expected. We can then speak of voluntary submission [14].

Example: The individual accepts information that he would not have accepted voluntarily.

DISCUSSION

Ethical Preoccupation

Our grid is under development. Validation results are in progress. Also, even without discussing its quality, its limitations and its interest in ergonomic practice, we want to discuss some points related to these issues in contemporary ergonomics.

Persuasion technology plans to change attitude and behavior, which naturally raises questions of individual freedom. It is a fascinating topic, but persuasion is not without mentioning acts of proselytism or propaganda in history, affecting among others the field of religion and politics. The technology itself, as a tool that can influence the masses, makes reference to some fears of manipulation embedded in the collective unconscious. Persuasion technology, like any persuasive method, is not unethical but depends in fact on its manner and use. Important work is being conducted among the young population, in terms of education or preventive health through persuasive games. But young people can be vulnerable towards attempts by third parties to disclose personal information, hence the importance of ethics in these type of interactions.

Rules

To address these problems, 7 ethical principles of persuasive design have been proposed [16]:

- The results of persuasive technology should never be considered unethical; if persuasion was undertaken without the technology or if the result would have occurred regardless of persuasion.
- The motivations behind the creation of a persuasive technology must remain ethical even if it is to lead to a more traditional persuasion (i.e. not mediated by technology)
- The creator of a persuasive technology should consider all sides and take responsibility for all results reasonably in its foreseeable use.
- The creator of a persuasive technology should ensure that it considers the privacy of users, with at least as much respect as if it were their own privacy.
- A technological persuasion that relays personal information about a user to a third party must be subject to scrutiny with regard to personal information.
- The creators of a persuasive technology should disclose their motivations, methods and expected results, unless such disclosure would seriously undermine another objective ethics.
- Persuasion technology does not deceive in order to achieve a final persuasion not avowed or undeclared.

In addition to these 7 rules, there is the golden rule of persuasion: “The creators of a persuasive technology should never seek to persuade one or more persons to do something they would not consent themselves to be persuaded to do.”

CONCLUSION

This communication allows us to explain 12 criteria divided into 2 dimensions – static and dynamic aspects of the interface – seeking to improve the performance evaluation of persuasive elements in interfaces. These criteria can also serve as guidelines or rules guiding the choice of design. This research also focuses on the social behavior of interaction with technology, the hidden dimension in nature. Moreover, classically

inspect ergonomics software is judging one's ability to be effective, efficient, error tolerant, easy to learn and satisfying for its users [17]; persuasion is generally outside the scope of the inspection. However, the intrusive aspects, ethical handling of certain domestic or professional interactions cannot remain outside the ergonomic analysis, particularly as these factors affect the attitudes of users to technology. By applying knowledge about how humans work on the psychosocial level, our inspection schedule is therefore based on a normative approach of what is and what is not a persuasive technology related to a product. It presupposes the existence of a relatively generic model of the human. However, the diversity of users and situations of use belies this narrow conception of human. For this reason, the performance of inspection techniques may be minimal. In any event, they should be supplemented by other evaluation methods. Therefore, the validation phase of this grid is to be achieved.

ACKNOWLEDGMENTS

This research has been supported by ETIC Labs – University of Metz in France and SAP-BusinessObjects Levallois-Perret in France.

REFERENCES

- [1] Fogg, B. J. 2003. *Persuasive Technology. Using computers to change what we think and do.* Amsterdam: Morgan Kaufmann.
- [2] Brangier, E. and Bastien, J. M. C. 2009, à paraître. *Ergonomie des produits informatiques: faciliter l'expérience utilisateur en s'appuyant sur les notions d'accessibilité, utilisabilité, émotionnalité et d'influçabilité.* In: Vallery, G., Zouinar, M. and Leport, M.-C.. (Eds.). *Ergonomie des produits et des services médiatisés*, Presses Universitaires de France.
- [3] Khale, R., Biddle, R., Noble, J., Barl, P. and Fischer, R. 2006. *Persuasive Interaction for Collectivist Cultures*, ACM International Conference Proceeding Series, 169, Archive Proceedings of the 7th Australasian User Interface Conference, Hobart, Australia, 50. Pp. 73–80.
- [4] Jordan, P. W. 1998. *An introduction to usability.* London: Taylor & Francis.
- [5] Nielsen, J. 1994. *Heuristic Evaluation.* In: Nielsen, J. and Mack, R. L. (Eds.). *Usability Inspection Methods*, New York: John Wiley & Sons, Inc.
- [6] Aldrich, H. E. and Fiol, C. M. 1994. *Fools rush in? The institutional context of industry creation.* *Academy of Management Review*, 19, 645–670.
- [7] Berdichevsky, D. and Neuenschwaner, E. 1999. *Toward an Ethics of Persuasive Technology.* *Communications of the ACM*, 42(5).
- [8] Madrian, B. B. and Shea, D. F. 2001. *The Power of Suggestion: Inertia in 401 (k) Participation and Savings Behavior*, *Quarterly Journal of Economics*, MIT Press.
- [9] Joule, R. V. and Beauvois, J. L. 1998. *La soumission librement consentie.* Presses universitaires de France.
- [10] Mugny, G., Souchet, L., Codaccioni, C. and Quiamzade, A. 2008. *Représentations sociales et influence sociale.* *Psychologie française*, 53, 223–237.
- [11] Chak, A. 2003. *Guiding Users with Persuasive Design: An Interview with Andrew Chak*, by Christine Perfetti, *User Interface Engineering*.
- [12] Dhamija, R., Tygar, J. D. and Hearst, M. 2006. *Why Phishing Works.* Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2006).
- [13] Joule, R. V. and Beauvois, J. L. 1998. *La soumission librement consentie.* Presses universitaires de France.
- [14] Joule, R. V. and Beauvois, J. L. 2004. *Petit traité de manipulation à l'usage des honnêtes gens.* Presses Universitaires de Grenoble, France.
- [15] Kiesler, C. A. 1971. *The psychology of commitment; experiments linking behavior to belief.* New York, Academic Press.
- [16] Berdichevsky, D. and Neuenschwaner, E. 1999. *Toward an Ethics of Persuasive Technology.* *Communications of the ACM*, 42(5).
- [17] Brangier, E. and Barcenilla, J. 2003. *Concevoir un produit facile à utiliser.* Paris: Éditions d'Organisation. 260 p.

Early Adopters' Experiences of Using Mobile Multimedia Phones in News Journalism

Tiina Koponen
Tampere University of Technology
Korkeakoulunkatu 1
33720 Tampere
tiina.koponen@tut.fi

Heli Väättäjä
Tampere University of Technology
Korkeakoulunkatu 1
33720 Tampere
heli.vaataja@tut.fi

ABSTRACT

Today's mobile multimedia phones with broadband connectivity have brought novel ways for journalists to live out their hectic profession. We interviewed six forerunners in Finnish news organizations on mobile phone usage in news production. Here, we present our results on how and why they use mobile phones in news production and provide implications for design and development. Our results show that users are delighted to use mobile phones in journalistic work if they can gain personal benefits from it, like more efficient use of time. Participants were not satisfied with the audio quality of mobile phones or mobile text editors. However, they considered mobile phones suitable for capturing short Internet videos. One reason constricting mobile production is editorial processes that are not designed for receiving material produced with a mobile phone. This causes extra work both in the newsroom and in the field.

Keywords

mobile phone, journalism, experience, work

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *user-centered design*.

General Terms

Design, Human Factors.

INTRODUCTION

The work of news journalists and photographers is facing new challenges due to the requirement for multi-skilled professionals and developments in digital network technologies [2]. Furthermore, there is a need for increased speed in publishing in online and mobile media. The development of mobile phones with advanced multimedia capabilities and network connectivity has brought novel ways for journalists to live out their hectic profession. News stories, photos and videos can be produced or even published straight from the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

field, making the news production process faster and more efficient. Mobile phones also provide a potential means to retrieve up-to-date or contextual information through the mobile Internet or context-aware services. In addition, calendar, navigation and the obvious communication functionalities are examples of their usefulness to journalists. However, the adoption of these converged mobile devices by professional users for news reporting is still at the early phases in media organizations.

In the field of human-computer interaction (HCI) Bellotti et al. [1] are one of the first to suggest that using handheld mobile devices, such as PDAs, could be useful for reporters due to the mobility of their work. Fagrell et al. [3] developed and tested a knowledge management system for PDAs to support the early phases of news-making in journalistic work in the mobile context. Furthermore, Hickey et al. [4] report on nine journalists' experiences of using a mobile phone with a 1-megapixel camera to capture photos and videos for a web magazine. They conclude that the key to adoption is the higher quality of multimedia items than was achievable with the used device. In addition, the work flow with multiple devices and software when transferring material from the mobile phone to be published in a web magazine was found to be complicated.

Väättäjä et al. [7, 8] and Jokela [5] report a field study on the user experience of mobile journalists. A mobile phone with a 5-megapixel camera and a mobile journalism application was used by ten students of journalism and nine students of visual journalism for producing complete news stories with multimedia items and publishing them in a web publication directly from the field. In this study, the quality of the multimedia items was not found to be the only critical factor. For writing the text editing functionalities were found to need improvement, whereas for capturing photos and video clips more functionalities and adjustments were wished for. Among the most important things for sending were, for example, the reliability and performance of the network connections. Issues related to network connectivity and limitations in battery power have been discussed for example by Zhang et al. [9] and Sørensen et al. [6]. Sørensen et al. [6] also identified infrastructural issues related to mobile device interoperability, mobile access and file sharing as important for users.

In this paper we present results and implications for design from interviews with six professionals on mobile phone usage for news reporting. We first describe the study and continue by presenting the findings. We then present the design implications and conclude with a discussion on the results.

PROCEDURE

We conducted six interviews to explore how and why mobile multimedia phones and applications are used in Finnish

newspapers and what kind of experiences users have; for example what are the problems users have faced and what development needs they have. Our goal was to find implications to support the design and development of a novel mobile multimedia service for mobile news journalism.

We interviewed six forerunners of mobile news journalism from four large to medium sized Finnish newspapers. Two photographers, two journalists, one video producer, and one editor-in-chief were interviewed. Their work experience varied from 3.5 to over 10 years. We wanted to select participants who had used mobile phones for producing journalistic material and found these six by contacting managing editors and online desks of newspapers who suggested suitable interviewees. Despite the fact that the interviewees had experience in mobile journalism it was novel and uncommon in their organizations. The participants were one of the few or the only one in their organization who used mobile phones for reporting news in the form of text, photos or videos.

The interviews were semi-structured, lasting from 60 to 90 minutes. The interview themes covered for example the usage of mobile multimedia phones and services in a work context, experiences of usage, the editorial process in news journalism and general opinions on mobile journalism. All interviews were recorded. Five of the interviews were conducted in summer 2008 and one in fall 2008. The data from the interviews was analyzed by content analysis. We first transcribed the recorded data and then grouped it under themes emerging from the data using inductive reasoning.

RESULTS

Here we describe the results of the study. The results are presented based on different factors affecting usage experience (photo and video capture, text, submission, process) and users' motivations.

Usage of Mobile Devices and Experiences of Usage

All of the interviewees had used multimedia phones for taking photos or video and for sending this material from the mobile phone to the newsroom. Two of them had also edited videos with a mobile phone. Two of the interviewees had used a mobile phone connected to an external Bluetooth QWERTY keyboard for writing articles, one regularly and another experimentally. Different solutions were used in sending the material, varying from email and MMS (multimedia messages) to special submission software like *Pro Reporter*³.

Photo and Video Capture

In newspapers, mobile phones' multimedia capabilities are mainly used for video capturing. Camera phone quality was not considered good enough for still photographing but adequate for short video clips to be released on the Internet. This was the general opinion of all interviewees concerning photo and video capturing with mobile phones. One photographer (working for the biggest newspaper in Finland) was especially worried that the quality of videos captured with mobile phones does not correspond to the image of the newspaper he was working for: *"It (taking photos and recording video with a camera phone) feels like we are cheating our readers. Our readers are used to high-quality journalism and expect a certain level of quality from us. It does not matter if the publication channel is print or web."*

Instead, camera phones are justified for authentic news journalism, such as accidents, when material is needed quickly but no other camera is available. In these situations quality plays only a minor role. One photographer describes catastrophe journalism: *"It is always better to get a grainy picture of the actual explosion, instead of photographing the smoky ruins with a professional camera."*

The users criticized the audio quality of video clips they had captured with camera phones. Camera phones were susceptible to background noise and had to be placed very close to the object to capture also the speech, almost *"into the mouth of the interviewee"* as one of the photographers expressed it. He said he would like to use different angles when capturing an interview video, instead of filming only a talking head, but that was not possible with the multimedia phone he had. If he moved any further from the object the camera phone could not capture the speech anymore. Other deficiencies in camera phones were a lack of manual settings and size. Small size and lightness are assets when carrying equipment, but make photo and video capturing unstable.

A big problem concerning especially processing video files was the battery life-time of the mobile phone. Editing and submitting video clips ate a lot of battery power. Users said they had to recharge the battery every few hours when using a mobile phone for video editing or submitting videos.

Text

Two users had written articles, news items and columns using a mobile phone connected to an external Bluetooth QWERTY keyboard (Nokia SU-8W), which they found useful and functional. One did this regularly every week and another experimentally. The users described the Bluetooth keyboard almost as comfortable as a normal sized desktop keyboard. Without it, using only the T9 text entry method, they would write at maximum 200 character stories. They felt that the QWERTY keyboard was a must for any longer articles.

The limiting factor in writing text with a mobile phone was the display size. Users said they could write shortish articles of up to 2500–3500 characters using a mobile phone and external keyboard, but not any longer than that because getting an overview of the structure of the text would become too difficult. It was also difficult to notice spelling mistakes and other typing errors on the small display. A spell checker would be beneficial to help users correct misspellings and create text that is ready for publishing without proofreading in the newsroom.

Mobile devices lack text editors appropriate for journalistic work. One user had tried out several text editors designed for mobile phones, both free-ware and commercial, finding only one editor including the most important feature in the journalistic context: character counter. Mobile text editors that participants had used did not support text formatting, for example using headings or bolding.

Submission

Material submission caused the users problems because of deficiencies in submission systems and inadequate capacity in the cellular networks that submission systems used.

The cellular networks were simply too slow for transmitting one- to two-minute video files, especially if there was no 3G network available. Users said they often had to wait several minutes to accomplish submission. The problem became emphasized when submitting material from abroad, which took up to 30 to 60 minutes. A notable detail affecting uplink times is that HSUPA (High-Speed Uplink Packet Access) was not in

³ <http://www.fromdistance.com/en/products/mobile-journalism/>.

use in Finland when the study was carried out. Besides being slow, submission was also unreliable. Videos had disappeared on their way from the mobile phone to the newsroom, or they had become damaged and unusable.

The submission systems participants had used did not offer sufficient information concerning the system state, i.e. if the submission was proceeding, finished or had got stuck. Therefore, users had to make a call to the newsroom to ensure material they had sent had arrived. Five of the six interviewees made a check call every time they had submitted anything from a mobile device.

Process

Using mobile multimedia phones in producing news stories often caused extra work because of undeveloped editorial processes. Editorial systems in newspapers were not designed for receiving material produced with a mobile phone. Material provided by a mobile phone had to be manually transmitted to the editorial system (independent whether the material was submitted via e-mail, MMS or a special mobile submitting application). Besides needing to be copy-pasted to the editorial system, material often needed editing or other manual processing like text formatting, converting videos into a different format, adding metadata etc. In addition to causing extra work in the newsroom, using mobile tools also caused extra work for the reporter in the field, who had to arrange a contact person in the office to receive incoming mobile material. Otherwise no one in the newsroom would necessarily notice that there is incoming material taken with a mobile phone waiting to be published.

It is essential to attach descriptive data to the picture or video, such as the names of the people in the video or picture, the date and location, and the photographer's name. The submission systems that the participants used did not offer functional ways to attach this metadata, which was therefore often sent separately. For example, video was sent using a submission application and descriptive text as an SMS message. This also complicated the process in the newsroom because one story could consist of several elements sent separately, which needed to be put together in the editorial system.

Motivations of Usage

The reasons for using mobile systems varied from an "order from the employer" and more effective use of time to eagerness to try out new technology (see Table 1). Photo or video capture was used in urgent situations to capture a unique newsworthy event when no better camera was available (for example accidents) or when material needed to get published quickly (for example local elections).

However, there was no aspect relating to quick publication in the background when mobile devices were used in writing text. In these cases motivations were more personal and brought personal benefits to the users, such as more efficient use of time and the possibility to write whenever and wherever inspiration hits. For example, the interviewee who regularly used a mobile phone for writing articles did this to make his work more efficient by making use of idle waiting moments for example when commuting, not because of a need for instant publication of the text. Also an interest towards new technologies and devices was strongly perceivable among those who used a mobile phone for writing articles.

All the interviewees were strongly aware that as "mobile journalists" they represented a small minority among all photographers and writing reporters. When asked why all

journalists are not inspired by the possibilities of mobile technology the interviewees assumed the main reasons to be low technical skills and low interest towards technology. One interviewee reckoned that their colleagues can not see mobile phones as versatile as they truly are.

Table 1. Motivations to use mobile phones in journalism.

Photo and video capturing	Writing news stories and columns
<ul style="list-style-type: none"> - Fast publishing. - Makeshift for unexpected situations when no other camera is available. - Order from the employer. 	<ul style="list-style-type: none"> - Interest and eagerness towards new technology. - Possibility to write in idle moments (e.g. between meetings or when commuting), which makes workdays shorter. - Lightness of tools, no need to carry a computer. - Possibility to write whenever and wherever inspiration hits. - Following news when out of the office. - Enjoying people see the user as a technological forerunner.

DESIGN IMPLICATIONS

In this section we have listed the design implications found based on the study to provide guidance in designing mobile phones and services for journalistic use.

1. Provide **feedback concerning the submission state**. Give user information on the state of submission; is it for example proceeding, waiting, or finished.
2. Support a **fluent, straightforward editorial process**. Material submitted from the mobile device should be automatically transferred to the right place in the editorial system/to the web/to the archive without any need for manual copy-pasting or converting file formats.
3. Provide a possibility to attach **metadata** such as the **reporter's name, the photographer's name, caption, time, date, and GPS coordinates** to multimedia and text elements.
4. Provide a text editor for the mobile phone including the following features:
 - a. **Character counter**. The length of journalistic articles is always predetermined.
 - b. **A spell-checker** would be beneficial to help the user notice misspellings and reduce the number of errors.
 - c. **Compatible text formatting with the editorial system** or publication platform makes it possible to finalize the text in the mobile device. The most important text formatting features for journalists are: **heading, subheading, head-note, and bolding**.
5. Using an **external microphone** would improve the audio quality of video clips and give more mobility in video capturing when the camera does not have to be placed "in the mouth of the interviewee".

DISCUSSION AND CONCLUSIONS

We conducted an interview study for six journalists working in Finnish newspapers in 2008 concerning how and why they use mobile multimedia phones in news production. Based on the findings we presented a list of design implications for further development of mobile phones and services for journalistic work.

News production with mobile multimedia phones is at an experimental stage at the moment. There are ongoing trials in many news organizations. For example, Reuters and BBC have already discovered the potential of these devices and experimented with them in their news production. Today's technology enables mobile production and even direct publishing from the field, but the underdeveloped integration of the mobile production into the editorial systems restricts mobile production from expanding. Extra work caused by mobile device usage does not encourage using these devices. Organizations that are planning to start using mobile multimedia devices in news production should not only create an effective and fast mobile work process but one coherent and flexible editorial process that takes into account that material can be produced using different devices (systems camera, video camera, mobile phone, computer), from different locations (newsroom, the field), and for different media (paper magazine, the web). This is also an important point for any other industry when increasingly moving towards mobile ways of working.

The next step in mobile production could be extending editorial systems to mobile phones. Then users could not only submit material from a mobile phone to the editorial system but also see the real layout i.e. how their text, photos, or video will appear in the publication. That would give more control to the journalist in the field over material (s)he has created, as final editing, like text formatting and layout positioning, could also be done in the field with the mobile phone.

In news journalism, mobile phones are used mainly for video capturing, but also for writing articles and photographing, especially if no better camera is available. Using mobile phones gives benefits such as faster publication, as material can be sent straight from the mobile device to the newsroom. A mobile phone and external keyboard can be easily carried in one's pocket at all times, which gives a reporter the possibility to write an article or column whenever and wherever (s)he gets inspiration. The limitations still present when using mobile phones related to battery power and the reliability and performance of network connectivity when submitting as well as embedding mobile phones into the editorial processes and work flows seem critical for the wider adoption and success of mobile phones in professional news reporting. However, the balance between the utility, convenience, suitability for the

context and usage situation may overcome the potential limitations when technology is seen as a means to an end for completing tasks or pragmatic goals.

ACKNOWLEDGEMENTS

This work has been supported by the PALTU project (2007–2009) funded by the Ministry of Education, Finland.

REFERENCES

- [1] Bellotti, V. and Rogers, Y. 1997. From Web press to Web pressure: multimedia representations and multimedia publishing. CHI '97. ACM, New York, NY. Pp. 279–286.
- [2] Deuze, M. 2005. What is journalism?: Professional identity and ideology of journalists reconsidered. *Journalism*, 6(4), 442–464.
- [3] Fagrell, H., Forsberg, K. and Sanneblad, J. 2000. FieldWise: a mobile knowledge management architecture. CSCW '00. ACM, New York, NY. Pp. 211–220.
- [4] Hickey, S., Rosa, C. and Isomursu, M. 2007. Evaluating the use of a audio-video mobile phone for web magazine reporters. *Mobility '07*. ACM, New York, NY. Pp. 503–506.
- [5] Jokela, T., Väättäjä, H. and Koponen, T. Mobile Journalist Toolkit: A Field Study on Producing News Articles with a Mobile Device. In: *Proc. of MindTrek 2009*, 30th Sept. – 2nd Oct. 2009, Tampere, Finland. In press.
- [6] Sørensen, C. and Gibson, D. 2004. Ubiquitous Visions and Opaques Realities: Professionals Talking About Mobile Technologies. *INFO: The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media*, 6(3), 188–196.
- [7] Väättäjä, H., Männistö, A., Vainio, T. and Jokela, T. 2009. Understanding User Experience to Support Learning for Mobile Journalist's Work. In: *The Evolution of Mobile Teaching and Learning*. Ed. Guy, R., Informing Science Press. Pp. 177–210. Available at: <http://books.google.com>.
- [8] Väättäjä, H., Roto, V. and Koponen, T. Developing practical tools for user experience evaluation – a case from mobile news journalism. In: *Proc. of the European Conference on Cognitive Ergonomics, ECCE2009*, Finland. In press.
- [9] Zhang, D. and Adipat, B. 2005. Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. *International Journal of Human-Computer Interaction*, 18(3), 293–308.

Does Cognitive Style Affect Student Performance on a Web-based Course?

Lamis Hammoud
School of Information
Systems, computing and
Mathematics
Brunel University
West London UK
lamishammoud@yahoo.com

Steve Love
School of Information
Systems, computing and
Mathematics
Brunel University
West London UK
Steve.love@brunel.ac.uk

Willem-Paul Brinkman
Delft University of Technology

The Netherlands
w.p.brinkman@tudelft.nl

ABSTRACT

A lot of research has been carried out to assess web-based courses. In many studies the concern has been the students' satisfaction and achievement in web-based courses and traditional face-to-face courses, and the comparison between the two. Other studies have focused on the development of web-based courses to meet the requirements of educational institutes. Studies about students' cognitive styles may be important for the designers of web-based courses because of the potential to enhance learning. However, the relationship between the students' cognitive styles and their satisfaction and achievement has not been researched fully and the implications are inconclusive. The aim of this study is to investigate the relationship between students' cognitive styles, their satisfaction, achievement, and their way of using a web-based course. Cognitive Styles Analysis (CSA) by Riding and Rayner [11] was selected as the instrument to determine whether students were field-dependent or field-independent. Students' attitudes toward using WebCT (Web Course Tools) were measured by a questionnaire specially designed for this intention. Students' activities on WebCT were observed through the tracking system which provided information about students' use of every tool and page on WebCT. The study does not provide data to support a relation between students' cognitive style and their use of online learning environments such as WebCT. However cognitive style seems to have an effect on student achievements.

Keywords

web-based learning, WebCT, cognitive styles, students' performance

INTRODUCTION

Most of the universities in the UK are using technology to develop courses that meet students' educational needs and goals [10]. Technology features can enhance learning outcomes by facilitating efficient delivery of instructional strategies and by supporting certain activities such as cognitive problem-

solving and decision-making processes of the learner [1]. Universities are implementing different types of technology-supported learning. This study will focus on web-enhanced courses only. Web-enhanced courses are traditional face-to-face course which include web-related materials. Web-enhanced courses usually adopt a course management system e.g. WebCT (Web Course Tools) [14].

WebCT is an important application for higher education. It has been developed by Murray Golderg, a faculty member at the University of British Columbia [2, 16]. WebCT is an integrated set of educational and management tools and an important provider of e-learning programs. It is specifically used for the design and development of teaching and learning materials. WebCT is mainly used to create sophisticated World Wide Web-based educational environments either by creating entire online courses, or simply by publishing materials that supplement existing courses. Users of WebCT do not need a lot of technical expertise as all content is accessible via a standard Web browser [16].

Technology has the possibility to enhance and transform teaching, but it can also be used incorrectly or in ways that may interfere with learning so it is important to know how we can achieve effective learning online [13]. Different ways can be used to measure the effectiveness of web-based courses. Therefore studies in distance education differ in what they use as evidence of online course effectiveness. Wells [17] studied the effect of an on-line computer-mediated communication course, prior computer experience and internet knowledge and learning styles on students' internet attitude.

Other research [12] investigated the relationship between student perceptions of others in an online class and both affective and cognitive learning outcomes. They demonstrated the significance of student-student as well as teacher-student interaction in online classes. They highlighted the importance of instructor presence and interaction among students to attitudes about the class. They believe that interaction between students is an integrated part of the class and that instructor should encourage and support the interaction. Thought, facilitating interaction, is time-consuming and often demanding.

Psychological studies have shown that personal beliefs/opinions about learning and environmental preferences affect learning behaviours. However, these learner characteristics have not been widely discussed in the web-based context [19]. "Cognitive style is seen as an individual's preferred and habitual approach to organising and representing information" [11, p. 8]. For example some people prefer learning by being told (e.g. lecture); others prefer learning by exploration (e.g. searching on Internet).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Research has been conducted to find the relationship between different cognitive style and web-based learning and design. Graff [4] investigated the interplay between cognitive learning styles and the effectiveness of online courses in delivering instructional content. Students were categorized on a range from wholistic to analytical. Wholistic learners view ideas as complete wholes and are unable to separate the ideas into discrete parts. In contrast, analytical learners are able to comprehend ideas in parts but have difficulty in seeing the complete picture. Along another axis, learning styles were arrayed from verbalizers to imagers. Verbalizers do well with text-based material, whereas imagers deal well with spatial data. The results showed that analytics performed better than the wholistics in the long-page format, which was 11 pages long with much content on each page. That is because Analytics were able to learn the content in parts, and could integrate the information. Also, imagers were superior to verbalizers on the recall test in the short-page format, which contained 23 pages of content with little on each page. The study concluded that Web-based learning environments should be matched to the cognitive style of the user.

In the same perspective Summerville [15] stated that matching cognitive style to teaching environments may be important because of the potential to enhance learning. However, at this time, the relationship between matching cognitive style and learning has not been researched fully and the implications are inconclusive, especially for hypermedia learning environments.

In another study, Jelfs and Colbourn [16] studied students' learning approaches within a group and how this affected their adoption or rejection of the electronic medium. They found weak correlations between deep, strategic and surface approaches to learning and perception of Communication and Information Technology. They said that measures of the deep, strategic and surface approaches to learning indicate potentially interesting relationships. They also suggested that to improve student interest in the use of computer-mediated communication and to motivate students then it has to be relevant to their course of study and that teaching staff have to also be active in their use of the technology. Students will quickly lose interest if they think that teaching staff are not paying attention to their students' contributions.

One of the most widely investigated cognitive styles with respect to student learning is field dependence [3]. Field dependence refers to an individual's ability to perceive a local field as discrete from its surrounding field [18].

PURPOSE OF THE STUDY

The aim of this study is to investigate the relationship between students' cognitive styles, their satisfaction, achievement, and their way of using a web-based course.

RESEARCH METHODS

The study was conducted at Brunel University, UK. All undergraduate and taught postgraduate courses delivered by the School of Information Systems, Computing and Mathematics at Brunel University are supported by WebCT.

Participant

There were 72 students taking the observed module. 51 students (23 females and 28 males) respond to both attitude questionnaire

and cognitive style analysis test CSA (Ridding). The age of respondents ranged between 18–20 years old.

Research Instruments

A questionnaire was designed to measure students' attitude toward WebCT. A five point Likert scale type of question was used in the questionnaire. The Likert scale is highly used in similar studies to assess respondents' attitude as for example: [5, 17]. The questionnaire contained 25 statements to which students can show their agreement or disagreement.

Information on students' use of WebCT throughout term time was obtained from the tracking system. The tracking system provides information on how many times each student visited each page in WebCT and how much time they spent exploring it. Moreover, the module leaders' approaches to using WebCT were explored by monitoring the web pages of their modules. These observations provided information about how they designed their modules, which tools they used, and how often they answered the students' questions.

In this study, the level of Field Dependence has been investigated along the cognitive style dimension. Field Dependence can be measured by the number of instruments that have been developed such as the Group Embedded Figures Test (GEFT) [18] and the Cognitive Styles Analysis (CSA) [11]. The CSA offers computerized administration and scoring. Therefore, the CSA was used as the measurement instrument for Field Dependence in this study.

Procedure

Students' cognitive styles were determined using the CSA test instrument during the term time in one of their lap sessions. Statistical data about students' use of WebCT was collected weekly. The statistical data was mainly in numbers giving information about how many times each student visited the web page for a module. Moreover, it provided records about how many times a student read or posted on the communication board. Also, it gave information about how many times they visited each page within a module and how much time they spent on them. In order to measure students' attitude toward WebCT a questionnaire was submitted on paper to the students at the end of the lectures on the module before the examination period began.

Students' general uses of WebCT were measured by the number of times each student accessed WebCT pages or used the discussion board for the observed modules. Students' achievement was measured by their grades (coursework and exam). Students' attitudes towards WebCT were measured by using the designed questionnaire. The questionnaire was submitted to the students during the term time in one of their lap sessions after they continued the CSA test.

RESULTS

31.4% of the students found to be field dependent, 39.2% intermediate, 29.4% field independent.

Differences were found between these three groups in terms of attitude and the way they used WebCT. Table 1 shows these differences.

Table 1. The mean of students' attitude, use of WebCT in relation to their FD/FI cognitive style.

FD_FI	Attitude	Time	Read	Post	Assessment	Content folder	Files	Total	CW	Exam
Field dependent	3.43	22.24	121.37	.19	9.00	117.19	184.38	56.00	63.19	49.06
Intermediate	3.49	21.89	178.00	.50	7.85	120.05	189.95	56.80	60.80	52.60
Field independent	3.44	31.75	73.13	.27	7.07	122.87	207.87	44.27	49.47	38.73
Total	3.45	24.90	129.39	.33	7.98	119.98	193.47	52.86	58.22	47.41

(Time (hours): overall time spent using WebCT; Read/Post: number of messages read/posted in the communication board; Assessment: number of times students practiced using the online tests. Content folders: number of time students accessed the lecture slides folder. Files: number of times students accessed available files such as study guide, coursework slides, and seminars questions and answers. Total: overall grades. CW: coursework grades. Exam: exam grades)

In order to find out whether or not these differences are statically significant, an ANOVA test was carried out. Table 2 shows the ANOVA results for students' attitude in relation to cognitive styles. The test results (as shown in Table 2 below) indicate that cognitive style does not appear to be a significant factor in students' attitude towards WebCT.

Table 2. ANOVA of the students' attitude towards WebCT.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.037	2	.018	.114	.893
Within Groups	7.780	48	.162		
Total	7.817	50			

The next analysis to be carried out looked at students use of WebCT. Table 3 shows the ANOVA results for students' use of WebCT, number of times each students accessed WebCT and the total time they spent on WebCT. The results indicate that cognitive style does not appear to have a significant effect on students' patterns of use when using WebCT.

Table 3. ANOVA of the students' use of WebCT.

		Sum of Squares	df	Mean Square	F	Sig.
Sessions	Between Groups	2773.693	2	1386.847	.406	.669
	Within Groups	164131.48	48	3419.406		
	Total	166905.17	50			
Time	Between Groups	998.787	2	499.394	1.502	.233
	Within Groups	15961.826	48	332.538		
	Total	16960.613	50			

The next analysis to be carried out looked at student achievement. Table 4 shows the ANOVA on the final exam, course work, and overall grades obtained by the students. The dependent variable is students' cognitive style. There were significant difference found between performance in the exam, coursework and overall grades ($p < 0.05$). Field dependent students got better grades in coursework and written exam than field independent students. However, in the coursework computerized test, field independent students got better grades than field dependent students.

Table 4. ANOVA of the students' grades.

		Sum of Squares	df	Mean Square	F	Sig.
CW	Between Groups	1677.257	2	838.628	3.969	.025
	Within Groups	10143.371	48	211.320		
	Total	11820.627	50			
Exam	Between Groups	1711.682	2	855.841	3.360	.043
	Within Groups	12226.671	48	254.722		
	Total	13938.353	50			
Total	Between Groups	1575.906	2	787.953	4.163	.022
	Within Groups	9086.133	48	189.294		
	Total	10662.039	50			

DISCUSSION AND CONCLUSION

Based on the students' responses to the attitude questionnaire the overall attitude of the students toward using WebCT was positive. This result backs up much research in the area; however, this study does not provide evidence that students' cognitive style affects their attitude towards using WebCT.

Field-dependent students were found to spend less time using WebCT than field-independent students. Although, field-dependent students used the communication board more than field-independent students, this could be explained by the [9]

study which indicated that field-dependent students rely on others for information, guidance, and maintenance of attitudes. At the same time field-dependent students accessed the available files on WebCT more often than field-independent students.

The differences between the three groups (FD, FI, and Intermediate) are clear however they are not statically significant. This leads to the conclusion that cognitive style seems not to be a significant factor in students' attitude toward WebCT. Furthermore, cognitive style has not been found to be a significant factor the in students' way of using WebCT (the number of times each students visited WebCT, time spent, number of pages visited, and posted or read messages). These results back up the findings from studies such as [8]. Students' Field Dependence does not have an impact on their learning performance in WebCT [8].

The results showed a significant difference between the means of the students' grades, which suggests that students' cognitive styles did affect their achievement. The results also show that field-dependent students achieved better marks in the course exam and their coursework. We can't connect this result to the students' use of WebCT; however, it can be explained by the subject area of the observed module subject.

This study found that students' cognitive styles seem not to be a significant factor in students' attitude toward WebCT. Also, the results suggest that students have positive attitude towards using WebCT regardless of their cognitive styles. Moreover, field-independent students did not differ significantly from field-dependent students in their way of using WebCT (the number of times each student visited WebCT, time spent, number of pages visited, and posted or read messages). In other words, students with different cognitive styles are able to learn equally well on WebCT online courses.

REFERENCES

- [1] Alavi, M. and Leidner D. E. 2001. Research commentary: Technology-mediated learning. A call for greater depth and breadth of research. *Information System Research*, 12(1), 1–10.
- [2] Burgess, L. A. 2003 WebCT as an e-learning tool: a study of technology students' perception. *Journal of Technology Education*, 15(1) Fall, 6–15.
- [3] Chen, S. Y., Ghinea, G. and Macredie, R. D. 2006. A cognitive approach to user perception of multimedia quality: An empirical investigation. *International Journal of Human-Computer Studies*, 64(12), 1200–1213.
- [4] Graff, M. 2003. Individual Differences in Sense of Classroom Community in a Blended Learning Environment. *Learning, Media and Technology*, 28(2), 203.
- [5] Hisham, N., Campton, P. and FitzGerald, D. 2004. A tale of two cities: A study on the satisfaction of asynchronous e-learning system in two Australian universities. In: *Proceeding of the 21st ASCILITE Conference*, December 5–8, 2004, Perth, Western Australia. Pp. 395–402.
- [6] Jelfs, A. and Colbourn, C. 2002. Do Students' Approaches to Learning Affect Their Perceptions of Using Computing and Information Technology? *Journal of Educational Media*, 27(1), 41–53.
- [7] Jones, G. H. and Jones, B. H. 2005. A comparison of teacher and student attitudes concerning use and effectiveness of web-based course management software. *Educational Technology & Society*, 8(2), 125–135.
- [8] Lu, J., Yu, C. and Liu, C. 2003. Learning style, learning patterns, and learning performance in a WebCT-based MIS course. *Inf. Manage.*, 40(6), 497–507.
- [9] Mcleod, D. B. and Adams, V. M. 1979. The interaction of field independence with discovery learning in mathematics, *Journal of Experimental Education*, 48(1), 32–35.
- [10] O'Neill, K., Singh, G. and O'Donoghue, J. 2004. Implementing eLearning Programmes for Higher Education: A Review of the Literature. *Journal of Information Technology Education*, 3, 312–323.
- [11] Riding, R. J. and Rayner, S. 1998. *The Learning Enhancement Programme; Cognitive Style and Effective Teaching*, (2nd Ed.). Birmingham, Learning and Training Technology.
- [12] Russo, T. and Benson, S. 2005. Learning with Invisible Others: Perceptions of Online Presence and their Relationship to Cognitive and Affective Learning. *Educational Technology & Society*, 8(1), 54–62.
- [13] Salter, G. 2003. Comparing online and traditional teaching- a different approach. *Campus-Wide Information System*, 20(4), September, 137–145.
- [14] Sivo, S. A., Pan, C. C. S. and Hahs-Vaughn, D. L. 2007. Combined longitudinal effects of attitude and subjective norms on student outcomes in a web-enhanced course: A structural equation modelling approach. *British Journal of Educational Technology*, 38(5), 861–875.
- [15] Summerville, J. B. 1998. The Role of Awareness of Cognitive Style in Hypermedia. Available at: <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED423865> [Accessed May 5, 2009].
- [16] Volery, T. and Lord, D. 2000 Critical success factors in online education, the international. *Journal of educational management*, 14(5), 216–223.
- [17] Wells, J. G. 2000. Effects of an On-Line Computer-Mediated Communication Course, Prior Computer Experience and Internet Knowledge, and Learning Styles on Students' Internet Attitudes Computer-Mediated Technologies and New Educational Challenges. *Journal of industrial teacher education*, 37(3), 22–53.
- [18] Witkin, H. A., Moore, C. A., Goodenough, D. R. and Cox, P. W. 1977. Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, 47(1), 1–64.
- [19] Yang, F. and Tsai, C. 2008. Investigating university student preferences and beliefs about learning in the web-based context. *Comput. Educ.*, 50(4), 1284–1303.

Developing a Rapid Questionnaire-based Metric for Experience of Form

Ville Lehtinen¹, Riikka Hänninen², Ville Toivonen³,
Antti Oulasvirta¹, Sari Kuuva⁴, Pertti Saariluoma⁴

¹Helsinki Institute for Information Technology HIIT, Finland

²University of Art and Design Helsinki, Finland

³Helsinki University of Technology TKK, Finland

⁴University of Jyväskylä, Finland

ABSTRACT

In this paper we report on the process and results of creating a questionnaire-based metric for experience of form. The questions for the metric were chosen on the basis of an expert evaluation and short iterative pilot tests, with the emphasis on ergonomics and the usability of the metric for both the commissioner of the questionnaire and the subjects filling it in. Basic questionnaire design guidelines were applied for building a logical and understandable structure of questions and the process for the evaluation event. The metric was finally adapted for the evaluation of four everyday drinking glass designs and tested in a small field study with 20 subjects. The usability of the questionnaire was evaluated according to chosen parameters for speed, understandability, and subjective experience, while usefulness was evaluated in terms of how well the metric could highlight defining characteristics for the objects under evaluation. Overall, satisfactory values were received for all parameters defined, and the questionnaire did manage to bring out defining characteristics for the glasses in the context of a rapid implementation.

Keywords

product semantics, questionnaire design, usability, ergonomics

ACM Classification Keywords

H.4.m [Information Systems Applications] Miscellaneous.

INTRODUCTION

Human emotional system is decisive when people decide what is personally important for them [9]. The effect of various factors, such as beauty, goodness, and usability, has been studied in relation to the user experience associated with objects, considering different weights of functionality and aesthetics [1, 2, 3, 11, 12].

Different approaches, such as Kansei engineering [7] and the concept of emotional design [8], have been proposed for taking customers' emotions into account in product design. "Mapping emotions" to an object is challenging because the experience is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

rarely limited to only the form of the object. It is essentially situational and influenced by factors such as prior experience and the object's inferred functionality. Creating methods for measuring this experience of product form is one essential part for solving this problem. Researchers in product semantics have created some methodologies aimed at effectively quantifying this experience by mapping emotions to different adjectives. Questionnaires and forms have usually been adopted for collecting the data from users for further analysis.

Our Objective

The purpose of our research is to develop a useful and usable method for measuring how potential users perceive the overall form of a design in a setting corresponding to the pre-purchase evaluation of a product. By "useful" we suggest that the method should convey a clear understanding of how the design is perceived by the potential users with respect to what the designers meant to express and, more importantly, that the method can be used to qualify the differences between differing design solutions. By "usable" we suggest that the method should be easy and quick to use especially in the iterative beginning phase of product design and, even more key, that it will be both understandable and rewarding for the subject providing the evaluation. Our approach is to develop a questionnaire that corresponds to the aforementioned goals. In this paper, we report on the process of developing a questionnaire for a specific class of design – everyday drinking glasses – and on a small-scale implementation of the questionnaire using four different glass designs and 20 test subjects.



Figure 1. The four glasses used in the study. From left to right: Iittala's Kartio, Aino Aalto, Ote and Geric, later referred to as glasses #1, #2, #3 and #4 respectively.

Drinking glasses provide a very good case for studying emotional interaction as they are much used and relatively simple technical objects. Moreover, emotional appeal is a key goal in designing them. We wanted to emphasize the level at which the subject connects to the evaluated object. For this purpose, the four chosen drinking glasses provide for an interesting evaluation object: the approach allows the user to touch or even hold the item while performing the evaluation. Furthermore, the four drinking glasses used are fairly similar in both size and usage, increasing the challenge of bringing out

meaningful differences among them. The four drinking glasses used for the purpose of this study (see Figure 1) were provided by Iittala, a Finnish design company specializing in housewares.

A problem in developing a metric such as ours is that the experience of form is a complex, subjective matter involving emotions and previous experiences. It may be difficult to segment and express this experience in some measurable form in such a way that the results of several subjects can be collected and analyzed unambiguously. By carefully designing a questionnaire to include both a logical and comprehensive structure and clear and relevant questions, we potentially can reduce the problem of ambiguity. On the other hand, the questionnaire itself can have a strong effect on how the subject will form and segment the overall experience of the design. By emphasizing the presence of the object during the evaluation phase and encouraging the subjects to study and hold the object in the hands, we hope to lessen the effect of the questionnaire itself in forming the opinion in relation to the object.

Another problem in evaluating the design of a product is the effect of predisposition to the evaluated design. As discussed by Karapanos et al. [5], the effects of social significance on experience of a certain design increase as the maturity of the released product approaches, in contrast to its other aesthetic qualities. In fact, two of the four drinking glasses in our studies were fairly well known designs for the targeted subjects, so the overall evaluation process for these glasses may have been very different. While our research setting was not specifically set for studying this difference, the results did show clear differences between the two pairs of glasses and offer some suggestive evidence of the magnitude of this effect.

Previous Work

In their study, Hsu et al. [4] used a semantic differential method (SDM) to evaluate different telephone design concepts with two different groups of subjects. The evaluation form was issued for both industrial designers and everyday users. It was found that the two groups of subjects evaluated the telephones differently and that some adjectives were, in fact, difficult to interpret for the normal users in the context of the given objects. The authors conclude that designers should adopt a user-centered approach in order to better understand user needs and preferences.

Petiot and Yannou [9] for their part describe an integrated method for measuring product semantics, wherein the differences between different designs are mapped by means of adjective pairs and pairwise comparisons between the designs. The methodology is extended to cover the needs of an industrial design process from the evaluation of different existing products to the launch of a new product design process, although the suitability of the methodology for other product design and evaluation needs is implied. Although the basic needs for measuring product semantics in the aforementioned study, emphasizing practicality and usefulness, are similar to ours, the approach taken is also quite different. Essentially, our study places emphasis on the usability of the methodology and is meant as a lightweight approach for expressing the semantics, perceived usability, and the feelings associated with a design at a more integrated level. For this purpose, the evaluation event in our method concentrates on a single design at a time, reducing the possible crossover effects apparent in pairwise comparison.

Also in alignment with studies on the effects of product aesthetics on perceived usability of an object [1, 2, 3, 11, 12], and the importance of social dimensions to the perceived

product aesthetics [1, 2], we decided not to limit the evaluation space of the methodology used to purely aesthetic adjectives but to include instead all kinds of relatable measuring dimensions that correspond to the experience of a product as a whole. Of course, in our case of everyday drinking glasses, the functional dimensions of the object are bound to be less relevant than, for example, in the case of previously studied Web sites [6].

THE PROCESS

The approach in our study was to create a fast and easy-to-complete questionnaire for evaluating everyday drinking glasses and to test it by using a set of four glasses. A major challenge concerning the specific set of items evaluated was to find subtle differences among objects whose functionality, size, and context of use are relatively similar. For the purpose of speeding up the process of creation of the questionnaire, a short iterative questionnaire design process was adopted instead of the relatively frequently used semantic differential method used in the previously mentioned studies [4, 9]. Our process consisted of 1) building a master list of all possible questions, 2) reducing the list to a compact size by means of an expert evaluation, and 3) performing a short iterative pilot study. In comparison to an SDM-based method, we hoped to create a more comprehensive mapping of the experience of form by not restricting the question space to only semantic differential scales. In fact, a Likert-scale-type expression such as “The item makes me want to touch it” would be difficult to present via a single adjective, let alone to map to two opposites as would be needed for the valued semantic differential scale.

A preliminary set of usability- and ergonomics-related target criteria were set for the questionnaire. The questionnaire was to be economical in that its completion was not to take more than 10 minutes for one glass, understandable in that at least 99% of the questions should be answered, comfortable to fill in (which would be measured by two feedback-type questions after filling in of the questionnaire), error-free (meaning that the users would understand the questions as designed), and useful, in that it would bring out differences between different glasses. More specifically concerning the last objective, at least two defining characteristics should be found for each glass, even though the designs in our study were fairly similar.

Mapping of the Semantics

The first step was to collect a comprehensive set of different questions and question types that could be used for expressing the experience of form. For this purpose, the results of an earlier study done in another project were used as a preliminary list. The list was then extended by means of a brainstorming session. After elimination of duplicate and clearly infeasible questions, we ended up with 150 questions. In view of known questionnaire-related practices, we also included five distinct question types in our preliminary list: Likert scale, semantic differential, rank order, checklist, and multiple-choices. The questions were then divided roughly into six categories: *material, appearance, touching, aesthetics, and inclination to buy*. A lengthy categorized list of these questions, along with a section on the different question types, was then evaluated by eight experts in the field of design and/or questionnaires. Based on this evaluation, the most suitable and valuable questions were chosen for each category.

Iterative Testing

The questionnaire was tested in pilot tests with two students to ensure that the questions were understandable and that answering

them for the full set of glasses wouldn't take too much time. One questionnaire took about five minutes; thus, the combination of all four glasses took about 20 minutes, which was considered acceptable. From the pilot test, a few further questions were removed from the list, on the basis of a verbal protocol for addressing the understandability of the questions. More specifically, the subjects had to briefly discuss each question in groups of 3–4 questions during the second glass's evaluation.



Figure 2. A subject filling out the questionnaire while holding one of the glasses.

The Questionnaire

The final questionnaire consisted of 45 questions for each glass, divided into 35 Likert-scale statements and 10 semantic differential adjective pairs. Additionally, some demographic questions were added to the front page along with brief instructions and two feedback questions for evaluating the questionnaire itself. All six of the categories originally identified were represented in the questionnaire. The adjective pairs mainly included questions in the categories of appearance and touching. The order of questions inside a questionnaire was not randomized but instead, according to found guidelines, was designed in such a way that the evaluation would progress from general aspects into more detailed aspects without skipping too much between the specified categories.

EVALUATION

The final phase of the study was carried out in a spacious luncheon restaurant near the industrial school of design. The questionnaire on the four glasses was completed by 20 persons of ages 20–60 years, equally distributed by gender and familiar with the chosen manufacturer's products. The four glasses were placed on different tables, and participants were to move between these in the order indicated in the questionnaire (four different sequences equally distributed between the subjects). The participants' questionnaire completion times were measured. After filling in the form, participants received a small design candlestick from the same manufacturer as a reward. After the empirical study, the criteria set for the questionnaire were checked and results were gathered for further analysis of the profiles of the four glasses.

The usability of the Questionnaire

All in all, the usability objectives for the questionnaire were well achieved. The results for the ergonomic criteria are listed in Table 1. The final speed of filling out the questionnaire for one glass ended up below three and a half minutes, since still in the pilot testing a value of approximately five minutes was achieved for the same questionnaire and same glasses. Almost every one of the 20 subjects also answered all questions in the questionnaire

(only three items were left unanalyzed), which would suggest that overall understandability of the questionnaire was achieved. Finally the desired comfort level of the questionnaire was achieved, according to the two feedback items at the end of the questionnaire. A Cronbach's alpha value of 0.93 was calculated for our metric signifying good internal consistency. It should however be noted that the scale of our metric is probably not unidimensional. The dimensionality of the metric was not investigated at this point of our study due to a low number of subjects in the testing of the metric.

Table 1. Observed usability of the metric.

Criterion	Metric	Received average
Speed	Time taken per questionnaire per glass	3 minutes 20 seconds
Understandability	Percentage of questions answered	99.9%
Comfort	Statement: "I could participate in a similar study in the future" (1–5)	4.25
Stress	Statement: "Filling in the questionnaire was tiring" (1–5)	2.1
Usefulness	Number of characteristics identifiable for a glass	3.75

The usefulness of the questionnaire

The usefulness of the questionnaire depends ultimately on its ability to pinpoint the subtle differences between designs. On the basis of our testing of the questionnaire with 20 subjects and four drinking glasses, the results were analyzed in terms of mean values, kurtosis, and variances. As the results of our study reflect not only the differences of the objects but also those of the subjects and their preferences, the variance and kurtosis values could also be expected to convey valuable information. One glass could, for example, split opinions regarding a certain aspect while the views on the other glass are unanimous. A general view of the differences can still be seen from a representation of the mean values for different questions (see Figure 3).

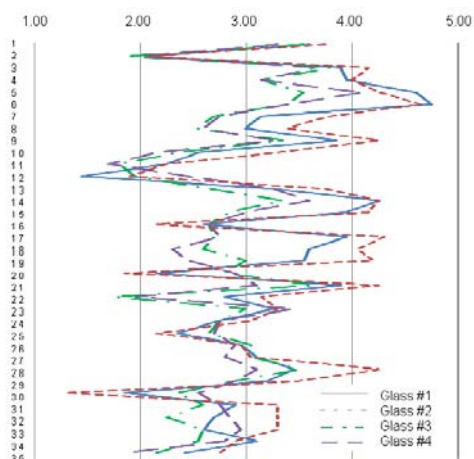


Figure 3. A representation of the differences between the four glasses based on mean values for the 35 Likert-scale statements (1 = disagree, 5 = agree).

As can be seen from the graph, the two groups of glasses can be clearly distinguished from each other. The familiar glasses,

#1 and #2, are very close to each other for the general questions, towards the beginning of the questionnaire. Some differences can be seen for the more detailed questions, towards the end of the questionnaire. The second group, glasses, #3 and #4, diverged more – and, from the variances, these glasses were also more opinion-splitting. As a summary of the questionnaire as a whole, Table 2 shows the most visible characteristics that can be highlighted on the basis of our results. The questions corresponding to the characteristics were chosen for glasses where kurtosis exceeded a threshold of 2.0. Kurtosis signifies the so-called “peakedness” of the distribution. A high kurtosis value means that the subjects answered the question at hand in the same way, leading to a peak normally close to the mean value.

The two more familiar glasses, #1 and #2, were regarded more positively in general throughout the study. According to the analysis of variance and kurtosis, various questions were shown to manifest peaks usually for one or two of the four glasses. Examples of such items were “The object fits my hand well” and “The thickness of the glass is appropriate” of the Likert-scale questions and the adjective pairs “fragile–durable” and “classical–modern.”

Table 2. Characteristics identified for glasses on the basis of kurtosis (>2.1), with mean values in parentheses (scales: 1–5 for Likert statements and 1–8 for semantic differential (SD)).

Glass #1	Glass #2	Glass #3	Glass #4
Pleasant surface (4.15)	Has familiar aspects (4.75)	Pleasant surface (3.80)	Fits well in the hand (4.10)
Fits well in the hand (4.26)	Is not practical (1.45)	Fragile vs. durable (SD: 6.10)	Worry about breaking the object (1.30)
Suitable for daily use (4.25)	Suitable for daily use (4.25)	Delicate vs. firm (SD: 6.20)	
Is disgusting (1.30)	Suitable thickness (3.90)		

CONCLUSIONS

By taking into account general questionnaire design guidelines and by conducting a small-scale expert evaluation of a master list of questions in combination with an iterative pilot study, we succeeded in quickly forming a usable and efficient metric for evaluating a set of similar product designs. The number of questions was scaled to take into account the number of objects studied and the reward given to the subjects. The 4 x 45 questions used in our study led to an average completion time of a little over 12 minutes with a high user satisfaction (4.25 out of 5 for comfort). Furthermore, we propose that, with small changes to the questions in the current questionnaire, the metric could be adopted for similar kitchenware-type products or even more disparate types of products.

Overall, the questionnaire created in the study performed fairly well for the main objectives of usability and usefulness. The

targeted usability metrics that were set for conveying the questionnaire for four glasses at a time were well met, and, above all, the subjects perceived the evaluation process in a positive way. As for the usefulness of the questionnaire, differing defining characteristics and statements could be found for each of the four drinking glasses, even though the designs of these everyday drinking glasses were initially chosen to be fairly similar in terms of size and usage.

ACKNOWLEDGEMENTS

This work was funded by THESEUS project and supported by Iittala.

REFERENCES

- [1] Hassenzahl, M. 2004. The interplay of beauty, goodness and usability in interactive products. *Human-Computer Interaction*, 19, 319–349.
- [2] Hassenzahl, M. 2004. Beautiful objects as an extension of the self: a reply. *Human-Computer Interaction*, 19, 377–386.
- [3] Hassenzahl, M. and Tractinsky, N. 2006. User experience – a research agenda. *Behaviour & Information Technology*, 25(2), 91–97.
- [4] Hsu, S. H., Chuang, M. C. and Chang, C. C. 2000. A semantic differential study of designers’ and users’ product form perception. *International Journal of Industrial Ergonomics*, 25(4), 375–391.
- [5] Karapanos, E., Hassenzahl, M. and Martens, J.-B. 2008. User experience over time. *CHI 2008 Proceedings*. Florence, Italy, April 5–10, 2008. Pp. 3561–2566.
- [6] Lavie, T. and Tractinsky, N. 2003. Assessing dimensions of perceived visual aesthetics of web sites. *Human-Computer Studies*, 60, 269–298.
- [7] Nagamachi, M. 1995. Kansei engineering: a new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics*, 15, 3–11.
- [8] Norman, D. 2004. *Emotional design: Why we love (or hate) everyday things*. Basic Books, New York.
- [9] Petiot, J.-F. and Yannou, B. 2004. Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics, *International Journal of Industrial Ergonomics*, 33(6), June 2004, 507–525.
- [10] Power, M. J. and Dagleish, T. 1997. *Cognition and emotion: From order to disorder*. Hove, UK: Psychology Press.
- [11] Tractinsky, N., Katz, A. S. and Ikar, D. 2000. What is beautiful is usable. *Interacting with Computers*, 13, 127–145.
- [12] Tractinsky, N. and Zmiri, D. 2006. Exploring attributes of skins as potential antecedents of emotion in HCI. In: Fishwick, P. (Ed.). *Aesthetic Computing*. MIT Press.

Session 3: Coping with Emergency Situation

Change Blindness and Situation Awareness in a Police C2 Environment

Gabriela Mancero

g.mancero@mdx.ac.uk

William Wong
Middlesex University
The Burroughs, London
w.wong@mdx.ac.uk

Martin Loomes
Middlesex University

m.loomes@mdx.ac.uk

ABSTRACT

We conducted a field study at the British Transport Police (BTP) Control Room in London to evaluate three hypotheses: that operators in a command and control environment miss changes; that missing changing information affects Situation Awareness (SA); and that interruptions affect operators' ability to detect changes.

Our results showed that if a controller's attention was drawn away, reading an immediately available log was sufficient for detection and understanding of relevant changes. Thorough incident logging in a well highlighted display was found to be an excellent recovery tool. However, a number of issues emerged about the operators' integration of information and spatial understanding requirements to maintain situation awareness during these incidents. A hypothesis that emerged from this study is that change blindness could occur in environments with graphical-tactical interfaces as opposed to the text-based ones used by the BTP.

This paper suggests that BTP operators' critical challenge is the integration of information and the need for spatial understanding to maintain situation awareness rather than the detection of visual changes *per se*.

Keywords

change blindness, situation awareness, attention, CDM

ACM Classification Keywords

H.5.2 [User Interfaces (D.2.2, H.1.2, I.3.6)]; H.5.3 [Group and Organization Interfaces].

INTRODUCTION

The literature suggests that Change Blindness (CB) could occur in operational environments [1, 2] especially those in which operators use several monitors and frequently may be heavily loaded with visual search, situation assessment, voice communications, and control display manipulation tasks that have to be performed simultaneously [3].

Several studies conducted in laboratories and real-world settings have shown that observers demonstrate high rates of

change blindness [4–7] especially when a change occurs during a visual disruption although such disruption is not always essential for the failure to occur [8].

Laboratory experiments have used a range of static stimuli ranging from simple shapes to more realistic photographs, dynamic simulations in aviation [9–11], combat [3] and driving [12] environments. In all cases, observers failed to notice changes that happened in their visual field. Relevance, expertise and familiarity with the scene reduce but do not eliminate CB. If the change was relevant to the task, detection rates increased. Expertise on the domain had a positive impact on detection but experts were still prone to missing changes [13]. Familiarity with the objects and scenes increased the rates of detection [14].

Theorists have argued that the first step toward good situation awareness (SA) is to notice objects and events in the surrounding environment [15]. Studies of Change Blindness (CB) and Inattentional Blindness (IB) have shown the importance of the role of attention in SA.

We hypothesised that change blindness occurs in real-world operational environments. We hoped to use our study of the nature of operational change blindness as a basis for the design of interfaces that could tackle this problem.

We needed access to a real-world operational environment that met the characteristics identified in the literature: high volumes of information, presented across several monitors with which operators need to multitask. We were permitted to conduct our study in the British Transport Police (BTP) Control Room in London, an environment which met all of these criteria.

We evaluated three main hypotheses: (1) Operators in a command and control environment miss changes, (2) missing changing information affects SA, and (3) interruptions detrimentally affect operators' ability to detect changes.

This paper examines these hypotheses, suggests that BTP operators' critical challenge is the integration of information and need for spatial understanding to maintain situation awareness and not the detection of visual changes *per se* and discusses its implications for the HCI community.

CHANGE BLINDNESS AND INATTENTIONAL BLINDNESS

There are two related but distinct phenomena that deal with visual lapses. Change Blindness refers to the failure to see changes that occur in our field of view during a visual disruption [16]. Inattentional Blindness has been described as the *looked-but-failed-to-see* effect and refers to the failure of seeing an unexpected event in the visual field when the observer's attention is diverted to a primary task [17].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

In the context of change blindness, change refers to a modification of the colour, orientation, or presence of a visual stimulus that is usually masked by a flash or a blink eliminating the signal that the change would have presented otherwise. Change differs from motion which is a variation referenced to location, while change is a variation referenced to structure [16].

In a study that used a flight simulation, pilots were asked to detect changes in the movement or altitude of weather systems or aircraft traffic. Changes in heading and airspeed were detected 40% of the time, while only 12% of changes in altitude were noticed [11]. Failures in change detection were reduced but not eliminated if the change was relevant to the task.

In a compelling real-world demonstration an experimenter asked pedestrians for directions. While the pedestrian was providing directions, two additional experimenters, carrying a door, passed between the initial experimenter and the pedestrian. During this brief interruption, a different person replaced the original experimenter. Even though the two experimenters looked different and had distinctly different voices, half the subjects failed to notice that they were talking to a different person [18].

IB focuses on the failure to detect unexpected stimuli. Evidence for inattention blindness comes mostly from relatively simple laboratory tasks, but the phenomenon has many daily analogues. For example, automobile accident reports frequently describe driver claims that they “looked but failed to see” the other vehicle. Critically, the difficulty of the primary task in an inattention blindness task increases the probability that people will miss the unexpected object [19]. In practical terms, the more people focus on aspects of their visual world other than the detection of unexpected objects, the less likely they are to detect such objects. Recent evidence suggests that talking on a cell phone dramatically increases the probability of missing an unexpected object while driving [20].

Many researchers have hypothesised that operators in environments with high volumes of information and rapid rates of changes could be susceptible to change and inattention blindness and may miss time-critical information [1–3].

Previous research has suggested change blindness as a problem that could potentially affect complex critical operational environments. Durlach, for example, stated that “as process monitoring and control systems rely on humans interacting with complex visual displays, there is a possibility that important changes in visually presented information will be missed if the changes occur coincident with a visual transient or distraction” [2]. Di Vita et al. affirmed that operators of computer-based systems are often heavily loaded with visual search, situation assessment, voice communications, and control-display manipulation tasks that must be performed concurrently. These operators usually work with several monitors, and the process of shifting their attention from one display to another creates an opportunity for changes to occur on unattended screens and time-critical information may be missed unless the interface specifically draws attention to the changed information [3]. However, despite the theories, we have not found any papers that found this phenomenon in a real-world operational environment.

BRITISH TRANSPORT POLICE

There are 43 police forces in England and Wales. Some operate within specific geographical territories, while others have responsibility for particular areas of activity. This system aims to prevent political interference in policing and avoids giving

any single organisation power over the entire police service. All regional forces are monitored by HM Inspectorate of Constabulary (HMIC).

The BTP is the specialist, national police service for Britain’s railways. It is the national police force for the railways and also polices the London Underground, Eurostar, the Channel Tunnel Rail Link, the Docklands Light Railway, the Croydon Tramlink and the Midland Metro [21].

As of March 2009, BTP operates two control rooms and one Call Handling Centre. The *First Contact Centre* is responsible for handling all routine telephone traffic. The *Force Control Room – Birmingham (FCRB)* is based in Birmingham – alongside the First Contact Centre – and is responsible for the East Midlands, West Midlands, Wales, the North West of England, the North East of England, the South West of England and Scotland. Finally, the *Force Control Room – London (FCRL)* is responsible for the Greater London area (including the London Underground and Mainline), London North and London South areas which are usually known as the Home Counties.



Figure 1. Map of England. The coloured areas represent the BTP FCRL operational areas. London North in light green, London South in dark green and the London Underground in dark blue.

The FCRL’s goal is to provide seamless communications to the three policing areas of the mass transit system; London North, London South, and the London Underground, covering more than 10,000 miles of track.

The FCRL physical layout is divided into four areas: call takers, radio dispatchers, CAD operators and the supervisors’ top desk. The monitoring and control of units is carried out in the dispatch area. This area is divided into three sectors mirroring the operational sectors: London North, London South and London Underground.

BTP operators work with three to five displays that provide them with access to the radio and telephony communication

functions, the command and control system, GIS information, CCTV for the London Underground, and the Metropolitan Police Computer Aided Dispatch (CAD) System.

When a call is received, a call taker obtains details, determines the incident classification and its grade or priority. ‘Immediate’ incidents are graded as G1 and units will use their sirens. ‘Prompt’ incidents are graded as G2 and units are required as soon as possible but they cannot use their sirens.

When it has been determined that an incident needs to be resourced, the information is sent to a radio dispatcher. The dispatcher’s task is to review the description of the incident, review available and assigned police units, and monitor ongoing incidents and outstanding jobs. The dispatcher may decide to change the priority classification. In addition, the dispatcher receives reports directly from officers in the field regarding new and ongoing incidents, which the dispatcher logs into the computer system. The dispatcher is also responsible for checking on officers periodically after they have been assigned to an incident.

The CAD operators monitor the Metropolitan Police CAD system. The Metropolitan Police usually sends them messages with rail related incidents. If these incidents require police assistance, the incident is logged into the BTP system and send to the radio dispatchers. The CAD operators require a comprehensive knowledge of the Metropolitan Police region and incident-type codes which are different from those used by the BTP.

METHODOLOGY

The field study combined a number of different approaches that were used to triangulate our understanding of the operational processes, situation awareness and decision making. Field observations, job task analysis and in-depth interviews were conducted to determine scenarios in which CB could affect performance during critical incidents. Internal documentation and performance reports were also reviewed.

We observed four radio dispatchers for 15 hours while they were working. We aimed to determine the frequency with which they shifted their attention between the different computer displays, and to assess the possibility that they could miss a visual change. We were not allowed to introduce any electronic equipment making it difficult to measure precisely the number of monitor shifts.

We also carried out a series of in-depth Cognitive Task Analysis interviews using the Critical Decision Method (CDM) [22, 23]. Four radio dispatchers, two CAD operators, two call handlers and one supervisor were interviewed. The radio dispatchers and CAD operators had between 5 and 8 years of experience, the call handlers one year of experience, and the supervisor 15 years of experience in the control room.

The CDM is a retrospective cognitive analysis method that employs a set of cognitive probes to critical incidents that require expert judgement or decision making [22, 23]. The CDM is a knowledge elicitation method which concentrates on a memorable incident experienced by the domain expert. The set of probes used for this study are presented in Table 1. The interviewee describes a memorable incident in detail which the interviewer uses as a framework for eliciting knowledge, decision strategies and behaviours. The method has been highly successful in eliciting perceptual cues and details of judgment and decision strategies that are generally not captured with traditional reporting methods, although the memories for such events cannot be assumed to be perfectly reliable [24].

Table 1. Cognitive Probes for the BTP.

Goal Specification *	What was the most important thing to accomplish at this point in the incident?
Cue Identification *	What features were you looking for when you formulated your decision? Have there been any situations in which you missed some information that became relevant in a later stage?
Influence of uncertainty	At any stage, were you uncertain about the reliability or relevance of the information that you used to formulate the decision? At any stage, were you uncertain about the appropriateness of your decision?
Information Integration *	What are the action lists that you use? How and where did you get this information, and from whom? What did you do with the information?
Situation Awareness Situation Assessment *	What information did you have available at the time of your decision? Did you use all the information available to you? Was there any additional information that you might have used to assist in the formulation of your decision?
Basis of choice	Do you think that you could develop a rule, based on your experience, which could assist another person to make the same decision successfully?
Mental model	Did you imagine the possible consequences of his action? Did you create some sort of picture in your head? Did you imagine the events and how they would unfold?

Our focus was on cue identification and situation awareness and assessment in order to be able to determine if change blindness could have a negative effect on operational performance.

RESULTS

Missing Changes Estimation

To form an upper bound on the number of changes an operator might miss, we counted the frequency of monitor shifts. Almost three changes of monitor per minute occurred during low/medium workload periods. According to previous laboratory research, this diversion of attention from one display to another should create opportunities for changes to occur on unattended screens, resulting in failure to notice time-critical information [2, 3].

Being a dynamic environment, determining the qualities of a “change” for the FCRL was crucial for our study. The type of “changes” that we were looking for are events that could be presented visually and were relevant to the dispatchers’ work such as incoming calls, new incidents, incident updates, incident location, and unit locations.

From our observations and interviews, we infer that even though radio dispatchers work with several monitors and during high workload periods they are heavily loaded with visual search, they did not seem to be affected by change blindness. This might be because the interface specifically draws attention to the changed information.

Information Systems Overview

In general, the systems were easy to navigate and present adequate use of visual and auditory modalities. Important notifications were well highlighted or colour coded when necessary. Visual cues, notifications and alarms were considered a solved problem.

Since operators are heavily loaded with voice communications, the radio and telephony communication functions are presented visually on a single operator touch screen that integrates the control of the communication subsystems.

The command and control system (C2) supports operators in creating, monitoring and recording incidents. It is a text-based

application, but it seems to provide an efficient platform to establish the available resources required and coordinate the deployment of resources.

The Metropolitan Police Computer Aided Dispatch System is used as a means of exchanging operational information with the Metropolitan Police. This system is a command line system that does not communicate with the C2 system used by the BTP.

Finally, GIS information is obtained from two sources: the “Gazetteer”, an online tool, and the Blueworld mapping tool, which is connected to the C2 system. However, Blueworld is hardly used. Although all operators reported the importance of geographical knowledge, all confirmed that the mapping tool is extremely cluttered and therefore, difficult to use. None of them was familiar with it. They normally use the “Gazetteer” only if directions needed to be given.

CDM Findings

The analysis of the data collected with the CDM was divided in four categories: Cue Identification; Situation Awareness; Integration of Information; and Information Recovery.

Cue Identification

There are mainly visual and auditory cues that are important for their job. Among the auditory cues, call takers, radio dispatchers and CAD operators agreed that the voice of the caller will determine the urgency of the incident.

Among the visual cues, the command and control system provides operators with several operational queues: the action queue displays incoming incidents that need to be resourced and colour-codes new incoming jobs in red and presents them first in the list. Incident updates are presented in blue for easy identification. The system also allows operators to label ongoing incidents into a ‘tagging list’ in order to facilitate search of immediate and prompt incidents (Figure 2).

Incident No.	Grade	Description	Location	Units
345	IMM	TRAIN DERAILMENT	STEVENAGE	BK345,B056,B43,B67,A2761
432	PM	DISTURBANCE	VICTORIA	IS 34, LX512
123	IMM	FATALITY	LEICESTER SQUARE	LA456,L56,LS 756,L95
458	RO	TICKETING	KING'S CROSS PLATFORM 18	BD345,B6456
234	PG	DRUNKNESS	BOURNEMOUTH STATION	
543	IMM	POLAC	GAINBOUROUGH ROAD	L35, L0987,J645, L61..
765	DE	DRUG COLLECTION	SOUTHAMPTON	H67, H57

Figure 2. Schematic representation of the “Tagging List”.

Operators constantly screen the queues for incident and units’ locations. If immediate assistance is required, it would be unacceptable for officers to go, sirens blaring, to the wrong location. Due to the possibility of traffic injuries, it is essential that they have the right location. Unit location, especially for officers on foot, is important because if an officer presses the emergency button, assistance needs to be sent immediately.

Situation Awareness

Radio dispatchers maintain their awareness by constantly screening the ongoing incidents even when they are doing a different task.

Radio dispatchers maintain what they call “environmental awareness” by using their “control ears”. They are able to listen to what the call takers or CAD operators are saying so they know that something is coming to them.

When asked, none of the interviewees reported difficulties when recovering from an interruption nor did they report

missing critical information. They keep their awareness by constantly screening their action queue and the tagging list.

If a major incident was declared, the constant change in priorities, scale of operations, information sources, and systems being used affected operators’ situation awareness. Inherent in the definition of Situation Awareness (SA), but *not* in Change Detection, is a notion of what is important: “relevance of the elements may be determined by the spatial, functional or temporal relationships of elements to goals” [15].

Integration of Information

Radio dispatchers often have to determine if multiple calls are from the same incident. They don’t want to tie up units unnecessarily in reacting to duplicated calls. Usually they wait for an officer to arrive to the scene to confirm such multiple incidents.

The CDM revealed that the task of the operator is inherently spatial but the way that operators perform the task is currently non-spatial. Unit locations (Figure 3), availability, specialization and geographical specifications are listed in codes.

Call Sign	Pw Ref	No	Location	Proximity	Activity code	Comment	Beat id	Resource type	Deployed	EndTime
IHM543	7890	1	Victoria		01	CAT		PCSO		23:00
IBR890	4356	1	King’s Cross		02	CORE TEAM		PCSO	Deployed	15:00
LE51	0039	1	West Ham		51	DUTY OFFICER		Duty Sergeant	Deployed	09:00
B76		1			01			Vehicle		14:00
AV54		1			09			Drug dogs		23:00
THA43	1209	1			01			Police Officer	Deployed	15:00

Figure 3. Schematic Representation of the “Resource List”.

Even though geographic knowledge is extremely important, operators hardly ever use their mapping tool. Instead, they use the lists provided by the C2 system to find units’ and incidents’ locations. Unfortunately, updating the location of police officers on foot is done manually. Police officers call radio dispatchers constantly to update their location requiring radio airtime. When an officer presses the emergency button, radio dispatchers are able to identify the individual officer but not the specific location. It then takes time to analyse where the officer has been deployed and where he or she may be. This analysis is all done using the lists provided by the C2 system.

Additionally, members of the public could call railway stations by their local names. Radio dispatchers must be familiar with the name of the locations, their colloquial names, and critical places near the incident location. A map could help them to visualize locations and other critical information.

CAD Operators also have to integrate the codes used by the Metropolitan Police with the ones they use within the BTP. There is no standard for classifying types of incidents, geographic locations and unit specializations.

Interruption Recovery

If the controller’s attention is drawn away due to an interruption, detection of variations in the visual field was simply the case of reading the incident log that is immediately available to them. The “tagging list” provides them with an excellent tool to visualise rapidly the current situation.

CONCLUSIONS

Based on the literature, we hypothesised that change blindness occurs in real-world operational environments, and by understanding how and when change blindness occurs in the real-world, we would have been able to inform design of interfaces that could tackle this problem.

Contrary to our predictions, operators at the FCRL seemed not to miss critical changes. Our second hypothesis was rendered moot, because change blindness did not affect their SA. Finally, although CB could have occurred during interruptions, operators have developed a series of strategies to recover rapidly. Additionally, their system provides them with the necessary cues and notifications.

Thorough logging gives them access on demand to the history of the incidents. The incident logging system provides BTP operators with an excellent tool to recover from interruptions and inform them about incident related changes that have been reported. In this case, unnoticed visual changes do not appear to be a significant problem.

Operators do, however, need better support for SA. BTP operators' critical challenge is the integration of information with the need for spatial understanding to maintain SA and not only the detection of visual changes. The incompatibility of the operators' inherently spatial task and the rendering of this information could be improved by creating display designs that show the units' locations, access routes and incidents' locations rather than listing them in a table.

This work has been an exploratory first step. It might be possible that change blindness could occur in a different operational environment, for instance, one that has a graphical-tactical display as their main workstation. A comparison with other similar operational environments would be ideal. Unfortunately obtaining clearance and access to such command and control centres is extremely difficult. Future work will include the evaluation of the hypotheses relating to the possibility of change blindness in graphical-tactical displays. The BTP's text-based interfaces allowed a type of logging that seemed ideal for specifically drawing attention to changed information. However, change blindness might occur in operational environments more reliant on monitoring spatial displays.

ACKNOWLEDGMENTS

Our thanks to the British Transport Police for allowing us to conduct this study.

REFERENCES

- [1] Mancero, G., Wong, W. and Amaldi, P. Looking but Not Seeing: Implications for HCI. European Conference on Cognitive Ergonomics; 2007, 28–31 August; Covent Garden, London: Brunel University; 2007. Pp. 167–174.
- [2] Durlach, P. Change Blindness and its Implications for Complex and Control Systems Design and Operator Training. *Human Computer Interaction*, 2004, 19, 423–451.
- [3] DiVita, J., Nugent, W., Obermayer, R. and Linville, J. Verification of Change Blindness Phenomenon while Managing Critical Events on a Combat Information Display (Attentional Processes). *Human Factors*, 2004, 46(2), 205–218.
- [4] Rensink, R. When Good Observers Go Bad: Change Blindness, Inattentional Blindness, and Visual Experience. *Psyche*, 2000, 6(9).
- [5] O'Regan, J. K. Change Blindness 2000 [Retrieved November 7 2006]; Available from: <http://nivea.psycho.univ-paris5.fr/ECS/ECS-CB.html>.
- [6] Simons, D. J. In: Sight, Out of Mind: When objects representations fail. *Psychological Science*, 1996 September, 7(5), 301–305.
- [7] Simons, D. J. Current approaches to Change Blindness. *Visual Cognition*, 2000, 7(1/2/3), 1–15.
- [8] Simons, D. J., Franconeri, S. and Reimer, R. Change Blindness in the absence of a Visual Disruption. *Perception*, 2000 July 14, 29, 1143–1154.
- [9] Muthard, E. K. and Wickens, C. D. Change Detection after Preliminary Flight Decisions: Linking Planning Errors to Biases in Plan Monitoring. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 2002, 46, 91–95.
- [10] Wickens, C. D. and Muthard, E. K. Factors that Mediate Flight Plan Monitoring and Errors in Plan Revision Planning under Automated and High Workload Conditions. 12th International Symposium on Aviation Psychology; 2003; Dayton, OH: University of Illinois Aviation Human Factors Division, 2003.
- [11] Wickens, C. D., Muthard, E. K., Alexander, A. L., VanOlffen, P. and Podczerwinski, E. The Influences of Display Highlighting and Size and Event Eccentricity for Aviation Surveillance. In: HFES, (Ed.). 47th Annual Meeting of the Human Factors and Ergonomics Society; 2003, Santa Monica, CA, 2003. P. 5.
- [12] Pringle H. I., Irwin D. E., Kramer A. F. and Atchley, P. The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin & Review*, 2001, 8(1), 89–95.
- [13] Werner, S. and Thies, B. Is Change Blindness Attenuated by Domain-specific Expertise? An Expert-Novices Comparison in Change Detection in Football Images *Visual Cognition*, 2000, 7, 163–173.
- [14] Archambault, A., O'Donnell, C. and Schyns, P. Blind to Object Changes: When Learning the Same Object at Different Levels of Categorization Modifies its Perception. *Psychological Science*, 1999 May, 10(3), 249–255.
- [15] Endsley, M. R. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1995, 37, 32–64.
- [16] Rensink, R. Change Detection. *Annual Review Psychology*. 2002, 53, 245–277.
- [17] Mack, A. and Rock. I. Inattentional Blindness: An Overview. *Psyche*, 1999, 5(3).
- [18] Simons, D. J. and Levin, D. Failure to Detect Changes to People in Real-World Interactions. *Psychonomic Bulletin and Review*, 1998, 5, 644–649.
- [19] Simons, D. J. Inattentional blindness. *Scholarpedia*, 2007, 3244.
- [20] Scholl, B. J., Noles, N. S., Pasheva, V., Sussman, R. Talking on a cellular telephone dramatically increases 'sustained inattention blindness'. *Journal of Vision*, 3(9), 156–.
- [21] HMIC. British Transport Police – Baseline Assessment. 2006 [Retrieved 2009 July 19]; Available from: <http://inspectorates.homeoffice.gov.uk/hmic/inspections/baseline-assessments-nonho-0607/baseline06-btp.pdf?view=Binary>.
- [22] Klein G., Calderwood R. and Macgregor D. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics*, 1989, 19:462–472.

Session 3: Coping with Emergency Situation

[23] Wong, W. The Critical Decision Method. In: Karwowski, W. (Ed.). International Encyclopedia of Human Factors and Ergonomics: CRC Press, 2006. Pp. 3067–3073.

[24] Hutton, R. J. B., Miller, T. E. and Thordsen, M. L. Decision-Centered Design: Leveraging Cognitive Task

Analysis in Design. In: Hollnagel, E. (Ed.). Handbook of Cognitive Task Design. Mahwah, NJ: Lawrence Erlbaum Associates, 2003. Pp. 383–416.

Collaborative Situational Mapping during Emergency Response

Lucy T. Gunawan¹, Augustinus H. J. Oomes¹, Mark Neerincx^{1, 2},
Willem-Paul Brinkman¹, Hani Alers¹

¹Delft University of Technology, Man-Machine Interaction group, Delft, The Netherlands

²TNO Human Factors, Soesterberg, The Netherlands

{l.t.gunawan, a.h.j.oomes, m.a.neerincx, w.p.brinkman, h.e.b.al-ers}@tudelft.nl

ABSTRACT

During emergency response, individuals observe only part of the picture, sharing of information is needed to get the required complete picture. The aim of our study is to get insight in the collaborative mapping process in order to derive requirements for a map-sharing tool. First, we analyzed the domain to assess the mapping processes, to identify general problem areas of the assessed processes. Subsequently, we conducted a laboratory experiment to systematically investigate the identified problem of collaborative map construction by individuals who observed an incident from different perspectives.

This paper discusses an experiment, which showed that the individual maps are sometimes better than the jointly constructed map, among other things due to the collaboration biases of unbalanced relations and uncertainty about oneself. Thus based on this experiment, the collaborative mapping tool should support joint map construction and help to prevent the identified collaboration biases.

Keywords

collaboration, user-centered design, distributed, situation map, disaster response, emergency

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces (D.2.2, H.1.2, I.3.6) – User-centered design, Prototyping.

INTRODUCTION

The effectiveness of emergency response team heavily depends on their response speed, and on how well they can manage their resources. Their aim is to reduce the resulting damage and the impact on human lives. In order to react promptly, the emergency response team needs to have a reliable overview of the disaster situation. This may include accessibility of the road network, the condition of damaged infrastructures, and the status of available resources. This overview, usually provided in the form of geospatial information, is useful to get a clear

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

mental image of the disaster area. Unfortunately, a situation map as such is often difficult to construct. Moreover, it is often the case that the scope of the incident is only understood after several days.

Geo-information technologies such as Geographic Information Systems (GIS) have the potential to provide critical support during disaster response. Situation-maps demonstrated to be an essential collaboration tool in crisis situations [1]. However, this technology is sometimes inefficient in supporting emergency response teams. Occasionally, wrong interaction modalities are used to convey information. For example, many current practices use verbal communication over the phone to describe geo-spatial information, which may result in misunderstandings and inaccurate positioning of objects and events. Furthermore, many geo-information technologies are too complicated to be used without prior training [2], and are usually designed without supporting collaboration in a team.

We aim to investigate the possibility of constructing a shared situation map using a collaborative distributed mechanism. By supporting collaboration among distributed information-sources, it is expected that the first hand information can be easily collected, checked, and shared. Thus, eliminating intermediate communication chains, which in turn may result in faster and more accurate situation maps. The shared situation map can also be used as a communication tool among the actors involved.

The setup of this paper is summarized as follows: first, we describe our field study by observing our users in their working environment and participated in disaster management exercises. Second, we describe our laboratory experiment setup in order to explore the idea of the joint situation map. Based on these things, we formulate guidelines for the collaborative situation map for emergency response.

OBSERVATION IN THE SAFETY REGION OF ROTTERDAM-RIJNSMOND, THE NETHERLANDS

In the Rotterdam area of the Netherlands, the crisis and disaster management is organized by Safety Region Rotterdam-Rijnmond. The participating agencies are municipalities, the fire service, the ambulance service, the medical emergency service, the police, the dispatch center, the Rotterdam port authority, and the environment protection agency.

We observed the work of this team in their exercises, both at the regional level and nationwide. One of the recent exercises is 'Waterproef', a nationwide disaster exercise in which a flood situation in the Netherlands was simulated. This exercise was held from 30 October to 7 November 2008.

The scaling of incident or disaster is regulated by a national agreement called Coordinated Regional Incident Control Procedure (GRIP). GRIP stages regulate the structure of collaboration of the agencies, based on the scope of the incident. The two main groups are the Incident Command Post (CoPI) and the Operational Team (OT). The CoPI works at the location of the incident while the OT works at the command center in the World Port Center Rotterdam. The CoPI reports the development of the situation to the OT, and OT updates any strategy changes during the incident. The internal structure of the CoPI and the OT is similar, consisting of representatives from the agencies mentioned above.

Current Information Sharing

Although members of the CoPI team work together closely and share information, they only report back to their own superior in the OT. For example, the leader of the CoPI reports to the leader of the OT, the police officer in the CoPI reports to his superior in the OT, and so on. The reporting is mainly done by phone. Any geo-spatial information received by the OT member is drawn on a paper map. These maps are collected by a plotter who draws and maintains a shared situation map that can only be shown among the OT. The information sharing of the team can be seen in Figure 1.

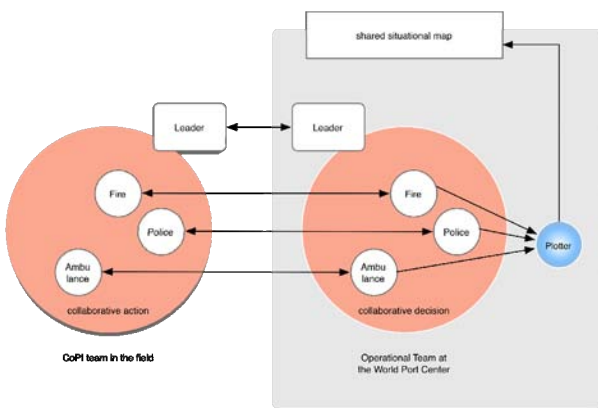


Figure 1. The current information flow and sharing. The collaborative agents are not limited to the fire service, the police, and the ambulance service but including municipalities, the medical emergency service, the dispatch center, the Rotterdam port authority, and the environment protection agency.

Problems with Current Information Sharing

The plotter often encounters difficulties in putting these reports into the system. This is mainly caused by the complexity of the system. The system consists of several unconnected geo-information technologies running on different terminals, as shown in Figure 2. The plotter needs several hours of training before he can use the system optimally, and it is hard to maintain his high performance level due to the infrequent use of the system. Thus, the use of this system in the field is considered as no option due to its complexity. Another reason is that many errors are made by other members of the OT, due to use of verbal communication to convey the geo-spatial information, which result in the need to constantly update the map. For example, as shown Figure 3, the location of the CoPI team on the paper map was first drawn incorrectly, and then a correction was made. It was an error in distance of around 2.75

km. Furthermore, in some cases, the members of the OT are heavily occupied with their business and forget to relay any geo-spatial information to the plotter, which can result in outdated information shown in the situation map.



Figure 2. Problem with current information sharing several systems that were used by the plotter to support his tasks.

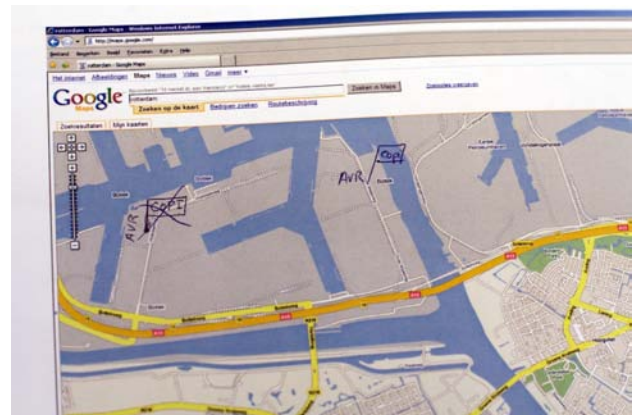


Figure 3. Error and correction on a paper map due to conveying geo-information over verbal communication.

Overall, the current information sharing is inefficient due to the many chains of information processing. This results in information sharing delays, and even unshared situation maps among the CoPI team and the OT. As it is clear that there are problems with current information sharing, there is a need for collaborative mapping among these teams. We assume that by targeting the collaboration activities will result in a more effectively shared map than the support systems currently in use [3, 4].

EXPERIMENT

Based on our observation at the Rotterdam-Rijnmond Safety Region, we are convinced that the distributed collaborative map has the potential to help this kind of user. By sharing information across organizations, both collaborating agencies in the field or in the command center can work jointly to overcome the system limitations. However, at this point, we still do not know how best to accomplish this goal. Thus we designed an experiment to get a better insight on the process of collaborative mapping. We were interested in gathering some observational data on how people collaborate in making a joint map. This first step can help us understand the basic characteristics

of making a collaborative map. We would like to use the experimental results to explore potential problems in constructing joint maps.

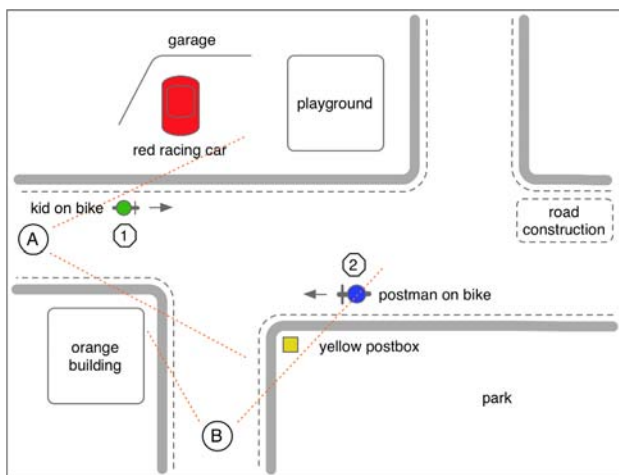
Procedure

We constructed an experiment in which two persons collaboratively made a simple map together. Each experiment lasted between 30 to 40 minutes. In the first phase, the two participants were shown different photo series of 20 pictures depicting an incident, where each picture was displayed for 5 seconds on the screen. Each photo series contained pictures taken from a different vantage point, thus some events were concealed from one of the participants, and vice versa. The photo series of the incident scenario were created by taking photographs of a miniature world populated with Playmobil toy sets. The Playmobil toys were chosen because they offered a good balance between simplicity and flexibility to be used for this purpose. After watching the photo series, in the second phase, the participants were asked to make their own sketch map of the depicted situation. Afterwards, as the third phase, they were asked to compare and discuss the differences in the maps they created individually, and then make a new joint map together.

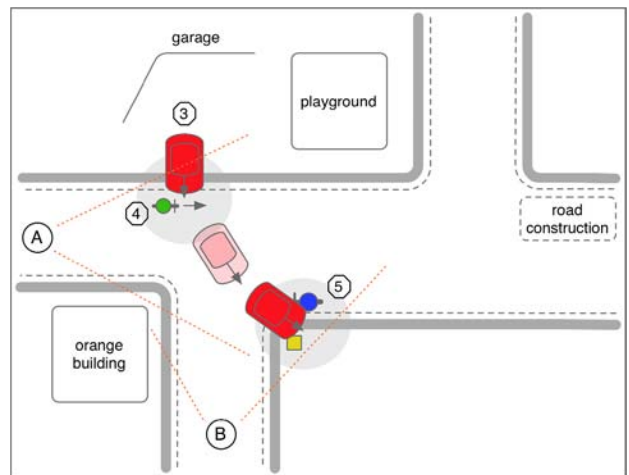
Scenario

The photo series used in the experiment describe an incident scenario in which the following events took place.

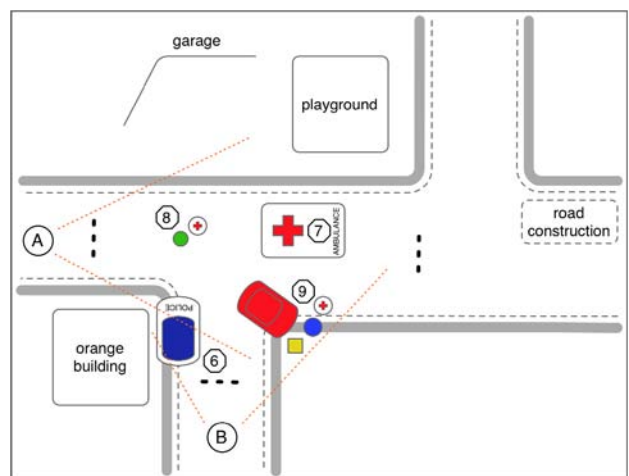
(1) A child on a bike was talking on his mobile phone without paying attention to the traffic ahead (2) At the same moment, across the street, a postman was riding his bike towards a yellow postbox (Figure 4a). (3) A red racing car abruptly hurtled out of a repair garage while being worked on by mechanics. (4) The car ran over the child, and injured him badly. (5) After hitting the child, the car continued cross the street, hit both the postman and the postbox, and then it stopped. (Figure 4b) (6) Shortly thereafter, the police arrived and closed down the area of the incident. (7) An ambulance with two paramedics arrived at the scene of the incident a while later. (8) One of the paramedics treated the child with the help of a bystander. (Figure 4c) (9) While the other paramedic provided first aid treatment to the postman with the help of another bystander. (10) The child was then transported by the ambulance to the nearest hospital. (11) The postman appeared to have no serious injuries, and did not require further treatment. (12) Finally, the police cleared the incident area and opened the street again to traffic. (Figure 4d)



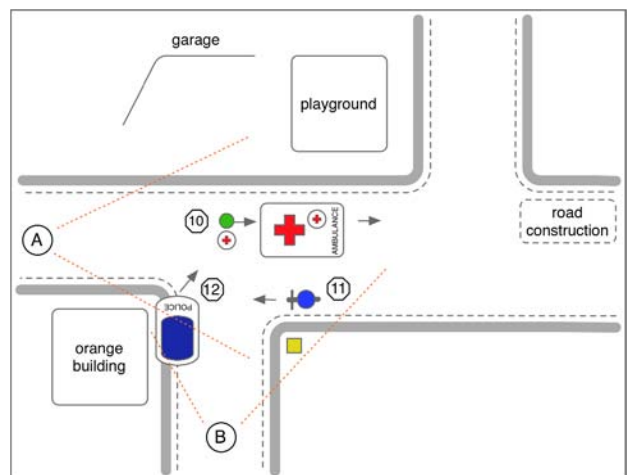
(a)



(b)



(c)



(d)

Figure 4. The overview of the incident scenario used in the experiment, (a) before the accident, (b) the accidents, (c) the response, and (d) the completion. A is the viewpoint of first participant, and B is the viewpoint of the second participant.

As mentioned earlier, the viewing angles were chosen in such a way, that some events were hidden or concealed from one of

the participants. The goal was to stimulate the exchange of information, requiring them to collaborate in order to figure out the complete scenario. Examples of these scenes can be found in three pairs of pictures below (Figures 5, 6, and 7).



first participant's point of view



second participant's point of view

Figure 5. Scene of red racing car hitting the child on a bike.

In the scene of Figure 5, the first accident took place, where the red racing car hit the child. Both participants were able to see the accident, but the first participant had clearer view of the accident.



first participant's point of view



second participant's point of view

Figure 6. Scene of red racing car hitting the postman.

The first participant could not see what happened after the car ran over the child, while the second participant clearly saw that the red car continued to hit the postman.



first participant's point of view



second participant's point of view

Figure 7. Scene of the child was carried into the ambulance.

The first participant could not see what happened after the car ran over the child, while the second participant clearly saw that the red car continued to hit the postman.

The first participant could see that the child was loaded into the ambulance, but he could not see what happened to the postman. On the other hand, the second participant was not able to see that the child was taken by the ambulance, but he had a clear view of the postman being treated.

Participants

In this experiment, we had 10 participants, who were divided in five pairs. The participants sample consisted of researchers and master students at the Delft University of Technology.

Results

In the third phase of the experiment, the collaboration phase, the participants were instructed to compare their individual maps and, based on their discussion, to construct a new map out of their combined recollections. However, choosing which detailed steps they needed to follow in order to achieve that goal was left entirely to the participants. By observing the collaboration phase in the five sessions, we noticed a pattern of steps being used repeatedly:

1. Telling each other their account of the scenario by using the individual maps they created in order to figure out overlaps and differences in their stories:
 - a. They start by stating many landmarks and stationary objects such as: garage, playground, orange building, construction road, postbox, cones, etc.
 - b. Next was the orientation step, where the participants tried to figure out their relative positions on the map.
 - c. Thereafter, they started to tell each other the events in chronological order.
2. Resolving differences and unclear facts.
3. Adding complementary information which were only known by one of the participants
4. Reaching agreement on the complementary information
5. Drawing the information of their combined accounts in a new map. Participants achieved that either by both drawing the map at the same time, or by allowing one of them to do the drawing while the other adding complementary information.

Table 1. List of activities.

No	Activities
1	A kid on a bike was talking on his mobile phone
2	Across the street, a postman was riding his bike.
3	A red racing car abruptly hurtled out of a repair garage
4	The car ran over the kid, and injured him badly
5	The car crossed the street, hit the postman and the postbox
6	The police arrived and closed down the incident area
7	An ambulance with two paramedics arrived
8	One of the paramedics treated the kid
9	The other paramedic treated the postman
10	The kid was transported by the ambulance
11	Postman appeared to have no serious injuries
12	The police cleared the incident area and opened the street

Table 2. Completeness of the individual and joint maps.

Acti- vities	Pair #1			Pair #1			Pair #3			Pair #4			Pair #5		
	a	b	ab	c	d	cd	e	f	ef	g	h	gh	i	j	ij
1			V			V			V			V			V
2			V			V			V			V			V
3			V			V			V			V			V
4			V			V			V			V			V
5			V			V			X			V			V
6			V			V			V			V			K
7			V			X			X			V			K
8			V			V			V			V			K
9			V			V						V			K
10			V			V			K			V			V
11			V			V						V			K
12			V			V						V			

In Table 2 above, the white boxes are events that were not drawn on any of the maps while the grey boxes refer to activities that were drawn on an individual map. Activities that were drawn on the joint map are represented by V. The red boxes X represent two types of activities. Those of which were wrongly drawn on the joint map even though they did not occur in the incident scenario, and the activities that were known to one of the participants but became unclear or less certain as a result of the collaboration. Finally, the yellow boxes K refer to activities that took place in the scenario and were known by at least one of the participants, so they could have been on the joint map but they were not.

We measured the performance of the collaboration by comparing the individual maps to the joint map as summarized in Table 2. A positive performance was achieved when the participant filled each others missing information, and thereby correcting wrong facts (depicted by the green boxes). While the negative performance is when the joint map was worse or less complete than one of the individual maps (depicted by the yellow and the red boxes).

Four out of the five pairs were able to correctly identify their relative positions on the joint map. The process of understanding orientation and relative position was important and necessary to ensure a smooth collaboration. The one pair that failed to complete that step correctly (Pair #5), faced considerable confusion in the discussion process. This hindered their ability to identify certain events in the scenario. As a result, they failed to draw these known events on the map.

Two out of five pairs resulted in a positive performance, as shown in Table 2 (Pair #1 and #4). They managed to piece together all the events of the incident scenario and drew them on their joint map. In both cases, we observed certain collaboration elements that helped improve their performance. These elements include the participants' mechanism of constantly re-checking the story facts, their willingness to listen and to learn from each other, the equality of their standing during the discussion process, and whether they have a prior history of collaboration.

Session 3: Coping with Emergency Situation

Below is a transcript of a conversation that took place during the experiment where the participants repeatedly re-check.

B:	<i>"actually, I saw there was this car, but I thought it was going to run over the kid, but then afterwards I saw a picture of the postman also being run over ..., with the letters on the ground"</i>	
A:	<i>"okay, the postman was also hit?"</i>	A was being informed that the postman was hit by the car, and clarifying this fact
B:	<i>"yea, I think so, yea"</i>	
A:	<i>"oh, hmm..., interesting"</i>	
	<i>"what I remember there was this bike, from my point of view, cycling here"</i>	
B:	<i>"he eh, was it a kid?"</i>	In this discussion, participant A and B were verifying that the first victim was a child on a bike.
A:	<i>"it was a kid"</i>	
B:	<i>"yea I saw the kid too here on the bike"</i>	
A:	<i>"so it was a kid then on the bike, sure"</i>	
B:	<i>"I think so"</i>	
A:	<i>"and then there's a car coming from the playground and that's run over the kid"</i>	
B:	<i>"well then the car runs over the both the kid and the postman"</i>	The conclusion of the discussion
A:	<i>"wow, that's impressive"</i>	
B:	<i>"but the postman didn't go to the ambulance"</i>	
A:	<i>"okay"</i>	

Collaboration provided negative performance in three out of five pairs (Pair #2, #3, and #5). From the observations, it seems that doubt about the observed events can cause hesitation in the collaboration process. In order to overcome the uncertainties, participants resort to adding extra information or omitting events they already had on their individual maps for the sake of reaching a consensus, as can be seen in the example below.

The red box **X** in pair #2 of Table 2, represent a faulty conclusion that resulted from their discussion. They concluded that there were two ambulances instead of one. The pair was indecisive in their discussion and took longer time to draw their joint map compared to the others. They often expressed their hesitation by using words such as 'maybe' and 'probably' in their conversation. Additionally, this was also caused by a mistake in one of the individual maps, where the ambulance was drawn in a wrong position. As a result, they were trying to overcome the confusion by proposing extra events that did not belong to the scenario. The conversation snippet below shows the part where an extra ambulance was added to the joint map.

B: *"you have the ambulance here, and I have the ambulance there"*

A: *"maybe I'm a little bit fuzzy with that"*

B: *"maybe there are two ambulances"*

Some collaboration biases could have been resulted of an unbalanced relationship between the participants, where a stronger personality or a more senior position can allow one participant to dominate the discussion process. These biases can cause some known facts to be discarded from the weaker participant. Examples of these biases are discussed one by one below.

The red boxes **X** in pair #3 of Table 2, represent the introduction of doubt over events which were believed for certain to be facts before the discussion. In this case, the second participant saw two accidents while the first participant, who seems more dominant, only saw one of the accidents. The following conversation shows the second participant, who was right at the beginning, being influenced by the first participant and then becoming unsure about the two accidents, and consequently left the unsure facts out of the joint map.

A: *"the green guy on the bicycle got run over by the red car coming from the garage, and the car then move crash into the thingy or something"*

B: *"I think it was a postbox and a postman"*

"I saw two accident actually, one it hit this guy on the bike in the middle of the street cross, then he hits the postman near the postbox"

A: *"Yea that could be something, I didn't see the postman getting hit, I thought he was still there at the end of the slideshow"*

B: *"okay"*

"okay, But we are sure that at least one accident"

"the second accident we don't know"

In the session of pair #5, a senior researcher was paired with a young master student. The student is represented in Table 1 as participant j. After viewing the photo series, she had almost all the events of the complete incident scenario drawn on her individual map. Unfortunately, the senior researcher (participant i) was uncertain of many facts. The discussion led to a worse joint map than the one originally drawn by the student. This was caused by her hesitation to speak up to the senior participant and being too polite to indicate that he was wrong. Therefore, many events in the student's account did not come out during the discussion and were not drawn on their joint map. These failures are represented by the yellow boxes in Table 2, pair #5.

GUIDELINES

Based on our field study observations and the laboratory experiment, we formulated our guidelines for collaborative situational mapping during the emergency response as follows:

The system should be easy to learn and easy to use, which is necessary to support any time critical operation under chaotic circumstances.

Time is a dimension that needs to be incorporated into the joint map. Both the chronological order of the events, as well as the time stamp indicating when it was added. All changes to the records should be kept as a history log that can be recalled when needed.

The orientation of the agents working in the field should be recorded together with their headings in order to help other collaborating parties understand their relative positions and their viewing direction.

All information entered to the system should be accompanied by some degree of certainty.

There should be a mechanism that evaluates the accuracy of the submitted information and protects information that is believed to be genuine from being overwritten.

The system should provide a simple mechanism to continuously recheck whether the joint map is accurate and up to date.

CONCLUSIONS

A collaborative joint map has the potential to support the disaster response team. However, the system should adhere to a specific set of requirements in order to take advantage of the submitted contributions. If not, then the system may result in the loss of invaluable data or introduce false events to the response team, and thereby hinder the rescue services rather than help them.

The requirements will need to be implemented in the system prototype, and tested in a subsequent experiment. This process will be repeated iteratively in order to refine the system.

The use of toys, Playmobil in our case, as quick prototyping material proved to be adequate in achieving our goal. It was possible to use the setup to easily simulate a modeled incident. On the other hand, since all Playmobil human pieces have a

standard design with a smiling face, photos taken of the incident model may need further editing to convey more appropriate emotions.

ACKNOWLEDGMENTS

We gratefully acknowledge the funding for our work from the IOP-MMI Programme that is run by SenterNovem, an agency of the Dutch Ministry of Economic Affairs.

REFERENCES

- [1] MacEachren, A. M., Fuhrmann, S., McNeese, M., Cai, G. and Sharma, R. 2005. Project highlight: GeoCollaborative crisis management. Proceedings of the 2005 National Conference on Digital Government Research, Atlanta, Georgia.
- [2] Cutter, S. L. 2003. GI Science, Disaster, and Emergency Management, *Transactions in GIS*, 7(4), 439–445.
- [3] Woods, D. D. and Hollnagel, E. 2006. *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*, Taylor and Francis, Boca Raton, Florida.
- [4] Klein, G., Woods, D. D., Bradshaw, J., Hoffman, R. R. and Feltovich, P. J. 2004. Ten Challenges for Making Automation a “Team Player” in Joint Human-Agent Activity, *IEEE Intelligent Systems*, November/December, 91–95.

Analysis of Work Demands of Multi-agency Emergency Response Activity for Developing Information Support Systems

Leena Norros
VTT Technical Research
Centre of Finland
P.O. Box 1000
02044 VTT, Finland
Leena.Norros@vtt.fi

Robert Hutton
BAE Systems Advanced
Technology
FPC 267 PO Box 5
Filton, Bristol, BS34 7QW, UK
Rob.Hutton@baesystems.com

Patrick Grommes
University of Dublin,
Trinity College
College Green Dublin 2,
Ireland
grommesp@tcd.ie

Nick Colford
BAE Systems Advanced
Technology
FPC 267 PO Box 5
Filton, Bristol, BS34 7QW, UK
Nick.Colford@baesystems.com

Marja Liinasuo
VTT Technical Research Centre of
Finland
P.O. Box 1000
02044 VTT, Finland
Marja.Liinasuo@vtt.fi

Paula Savioja
VTT Technical Research
Centre of Finland
P.O. Box 1000
02044 VTT, Finland
Paula.Savioja@vtt.fi

ABSTRACT

This paper reports first results of analyses concerning emergency response (ER) activity executed as a joint effort of three agencies: fire services, ambulance services and the police. The challenge to be tackled in the study is to understand the cognitive, operational and collaborative demands that characterize on-site responding to a complex emergency situation. The concept of Common Operational Picture (COP) is used to indicate one of the distributed cognitive functions of the multiagency ER personnel. The process of the adopted usage-driven design approach is described, and some central methodical solutions explained. Tentative results concerning the formation of COP indicate that a communication-oriented (semiotic) approach provides a possibility to empirically analyse the formation of COP and to understand the cognitive patterns that actors, environment and artefacts jointly form for tackling unanticipated and complex situations.

Keywords

emergency response, common operational picture, common ground, communication, usage-driven design

ACM Classification Keywords

H.1.2 Human factors, *User/machine systems*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

INTRODUCTION

The study is targeted to achieve significant improvement in command and control performance, reliability and cost of multiagency emergency response (ER) activity. This will be aimed at in EU FP7 funded research project called the Common Operational Picture Exploitation (COPE). The design-oriented study (design study) should achieve a step change in information flow both from and to the first responder in order to increase situational awareness across agencies and at all levels of the command chain. The development work in the project concentrates on providing support to the incident commander (IC) and his support personnel in complex emergency situations.

Only partial success in the functionality and user acceptance has been achieved in previous attempts to develop technologies in support of first responders. One reason for failure appears to be insufficient understanding of the actual work of emergency responders. In addition, often times the involvement of end-users in the design process has been too sporadic. Hence, the leading principle in the COPE work is to accomplish a *usage-driven* design process in which technology developers, human factors specialists and end-users collaborate in a systematic way. The design process is driven by the demands arising from the ER activity.

USAGE-DRIVEN DESIGN APPROACH

Three different human factors teams in three countries, together with end-user participants of the project, took the responsibility to accomplish the analysis of emergency response work and to discuss achieved findings with the technology developers. As a broad frame, a design model from Crandall et al. [1] was used. According to this model, a usage-driven design comprises of five knowledge creation functions that are supposed to provide specific human factors outcomes (indicated by the arrow below):

- Preparation => Domain understanding
- Knowledge elicitation => Cognitive work challenges
- Analysis and representation => Leverage points

- Application design => Design concepts
- Evaluation => Impact estimates.

While adopting this frame we emphasize, first, that our usage-driven design focuses on analysis and design of the work system and its intrinsic constraints and demands. We find that focusing on singular users' tasks and experiences is too restricted for design that aims at improvements in complex work. The work system perspective takes into account and portrays the global goals of work and the functions that are needed to their fulfilment, and to reaching outcomes. Portraying these global, and not immediately visible aspects of work, enables understanding the meaning of singular actions and operations that may otherwise appear to the observer as mere routines or even reflexes.

A second specification that we make regarding the usage-driven design approach refers to how the design target is comprehended. We see that the system to be developed is a joint human-technology system, sometimes called the "joint cognitive system" [2]. In a broader sense the joint system approach means that technology and human practices are mutually dependent on each other so that changes take place in both simultaneously. In more narrow sense the joint system approach means that the specific solutions that we are aiming at in the design are composed of coordinated functioning of both technical and human elements. In our study the joint system under design is the Common Operational Picture (COP).

COMMON OPERATIONAL PICTURE

There is no single definition of Common Operational Picture. Originally the concept was used in military context to refer to an integrated information architecture infrastructure that supports interoperability of technology systems. But the concept is also loaded with human factors connotations. It can be linked to several notions that are used frequently today to characterise complex cognitive functions or patterns of real life decision making, so-called naturalistic decision making. Concepts like Situation Awareness, Team Decision Making or Team Sense Making may be mentioned [see e.g. 3]. The concept of Common Ground is a further notion that has attracted our attention. It focuses on communicational processes and denotes the need for shared basic assumptions and understanding of responsibilities as basis for coordinated activities [4].

In COPE project, Common Operational Picture (COP) was defined as follows: "COP is the description in time of the emergency situation that supports the emergency responders within and between different agencies to act appropriately." We have taken the primary functions of the COP as: supporting the development and maintenance of common ground and the support of coordinated action across actors. We consider further that the COP is a pool of information

- that is registered and stored in a database
- concerns past, present and expected future events
- that is available for presentation in a user interface suitable for emergency responder work
- the form of presentation of which is consistent and unambiguous, but not necessarily similar to all stakeholders
- the content of which is structured around operational processes of the emergency responder
- that needs to be interpreted and acted upon by the emergency responders

- that is meaningful in the context of emergency responder work.

This initial definition was targeted to guide the design so that improved COP could be formed later. It was foreseen that the conception of COP would improve as a result of the analysis of the present emergency response practices and during the actual design work.

ACCOMPLISHING THE DESIGN STUDY

Domain Understanding

In the first preparatory phase of our design study the end-user organisations of the project played a central role in acquainting the human factors and technology experts with the emergency response domain. Several visits were paid to the end-user organisations and a number of discussions took place with substance matter experts from police, fire and ambulance services. As the study takes place in a European multicultural context, the challenge was to find a joint understanding of what is meant by the "Emergency Response" activity in each of the 7 countries participating, and could there be a shared conception of it. Equally problematic appeared in the beginning to find a joint understanding of how to define the focus of the study, i.e. which actors' work the project's research and design should support. The focus was decided to be is on the activity of the Incident Commander. IC is the person in charge of multiagency emergency operations in the accident site. From this position IC is in command of several multiagency first responder units or sectors and is reporting to possible higher level command units that operate in the headquarters. Figure 1 illustrates the ER activity under study.

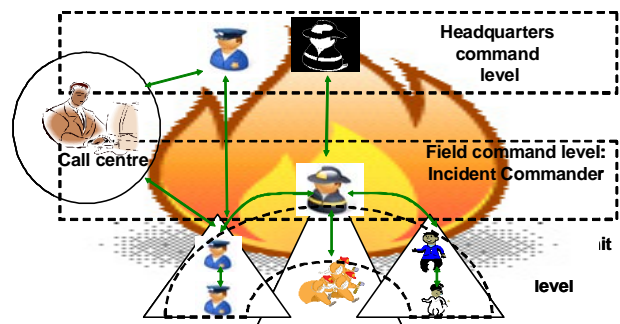


Figure 1. Multiagency emergency response activity (police, fire, ambulance services). The study takes the point of view of the Incident Commander.

Cognitive Work Challenges to Be Portrayed in User Requirements

The second phase of the usage-driven design process was the human factors data elicitation in the field. The aim was to identify the cognitive work challenges. The field work was accomplished in three countries, i.e. UK, Ireland and Finland. A joint method for data collection was designed.

The interview method used in the studies was the Critical Decision Method [see 1]. According to the method the interviewer attempts to elicit so-called "lived-incidents" from the interviewees. A lived-incident means one that the interviewee was personally involved. In preparing the interview it is important to select the interviewee appropriately and to build rapport. The interviewee is explained that the intention is to

understand the actors working environment, the challenges they face, etc, but it is not to evaluate or judge their actions.

The interview process is described in terms of 'sweeps' or multiple accounts / iterations of the same incident, focusing on different levels and types of information each time. Four sweeps are identified:

First sweep – Initial account: The interviewer asks the first responder to tell about a situation in which his skill as a fire fighter was put to the test.

Second Sweep – Identify Events and Decision Points (DP) on a Timeline. In this phase the first responder is asked to describe the incident, identifying a concrete beginning and end to the incident, and all the critical events and decisions that he had to make.

Third Sweep – Deepen on Each DP. The goal is to deepen DP and to identify cognitive demands and patterns in the first responder's professional practice. Questions like "What was your assessment of the situation at this point in time? What were you trying to achieve? etc. are asked.

Fourth Sweep – Technology Implications? The idea to get even more out of the interview, she/he is also asked to identify potential technology that would support in emergency response.

The human factors work group accomplished 20 critical incident interviews among first responders. Observations and demonstrations of currently used artefacts were also made when visiting in the end-user organisations. The data was either recorded as field notes which were transcribed to incident descriptions. Part of the interviews were recorded and transcribed into written protocols for detailed content analysis (see below).

Another very important source for understanding the first responder activity is the standard operating procedures. The procedures were in the first place used as description of the tasks, and tactical or operative decision making schemes of the first responders in different types of emergency situations. Later we found it necessary to revisit the issue of procedures.

The analyses of the end-user interviews and of the standard operating procedures were constrained by the needs of the technology developers. According to the ideas of the usage-driven design, technology should not push development of work. Instead the human factors experts should deliver descriptions of user requirements which would help in finding meaningful ways of using technology and in steering technology development. The "design artefact", typically applied here is the user requirement document.

A standard type of a user requirement document was generated by the COPE human factors work group. In this report we defined the Incident Commander's generic tasks in some detail and inferred a large number of requirements that the technology should fulfil in order to support accomplishing these tasks.

It is clear however, that no list of requirements can be exhaustive, because activity does not lend itself of being captured in such a way that engineering purposes would seem to require. A real challenge for usage-driven design is to find further ways of making analyses of work useful in design. We believe that authors who have proposed the idea of so-called joint cognitive systems are on the right track. In this line of thought the unit to be designed is the joint system and its intelligent and adaptive reacting in different kinds of situations. Such a system needs to be designed in a dialogical way so that technological and human factors *thinking* meet and merge, and bring creative solutions. In the present design practice just the

outcomes of either one are delivered. In the present project, working groups were formed within which human factors specialists dismounted to work with technology developers and to deal with the challenges of the four main design areas of the project, i.e. command and control systems, decision aids, first responder wearable technology and wireless sensor networks (WSN). In the following we shall refer only to the work that was accomplished with regard to the WSN. In this working group an attempt was made to concretise the idea of joint cognitive systems design in the specific design task.

Cognitive Work Challenges: Description of Human-technology Joint Functioning in ER Activity

When adopting a joint cognitive system perspective in design there is a need to make clear that both the target of design and the design practice need to be reframed. It may be claimed that the target is prior to the design practice, because it as an anticipated result, together with the constraints of reaching it, shapes the structure of the activity that produces it.

COP as a Joint Cognitive System

In our case the target of design is the joint function called the Common Operational Picture. Instead of focusing on the various task-based requirements of the IC we shifted our attention to the essential function of forming, via communicating and acting, a common ground and interpretation of the emergency situation. Above we have described the conception of COP that we held as a starting point. At this point a theoretical idea was added to interpret the COP formation process as a semiotic process. This idea was embedded in the empirical analysis of COP (see below) but cannot be developed theoretically in this short presentation (see for the theory the paper of Salo et al. in this conference). We chose to consider COP as a semiotic, or communicative phenomenon and an emerging cognitive structure that expresses and communicates meaning of the environmental signs. Via continuous perception-action cycles actors develop anticipatory structures which enable considering the effects of own actions and their possible outcomes. The environment, artefacts and actors are all part of the COP process.

We made an attempt to concretise COP in practice by analysing how COP is formed in the present practice of fire fighters who act in a multi-agency situation. Two types of empirical material were used. We first analysed the interview data elicited among Finnish fire fighters with the help of the Critical Decision Method. In the analysis of the fully transcribed data we focused on the way the interviewees described their decision making and how they conceived the role of different artefacts in the decision making. The analysis indicated, and thus verified some earlier results from the ER domain, that the first responders do not like to name specific decision making instances in the flow of response actions. The possible peculiar features of the situations were interpreted as "natural consequences" of the situation. The means of structuring the experiences were the learned routines that are schematised in the standard operating procedures. The interpretation of the environment and the division of responsibilities run smoothly with the aid of these routines. The role of Emergency Response Centre was emphasised in the formation of a picture of the situation and the role of a multiagency TETRA radio network, used in Finland among agencies, was the most important medium for communication. It was however interesting that the communicative role of some physical artefacts was also mentioned, e.g. one interviewee explained the important role of the water hose in communicating with the fellow fire fighter during smoke diving.

To gain more detailed empirical insights of the COP, and of the added value of the WSN, we conducted a design experiment at the Emergency Services College in Kuopio, Finland. Important for gaining insights of COP was that now we faced a particular fire fighting situation. It was a traffic accident in which three vehicles collide, one of them being a truck carrying hazardous substance, ammonium. The collision leads to a leak of ammonium that, mixed with extinction water, leaks into nearby environment, and in the form of an ammonium cloud spreads even further. The situation made it possible to analyse and model the aims and purposes of the activity, the first responder functions that are needed, and the information and resources that are available. The functional modelling approach we used does not prescribe a particular task sequence but makes explicit the possibilities and constraints of the situation. This model formed the background against which the actual actions of the first responders could be portrayed. Our analysis focused on the Incident Commanders work. In the analysis of the videotaped course of the incident and the voice recordings of the TETRA communication, which were transcribed to protocols, it was possible to construct how the COP was formed in this situation. In the analysis the following model of the COP process was used as a background (see Figure 2).

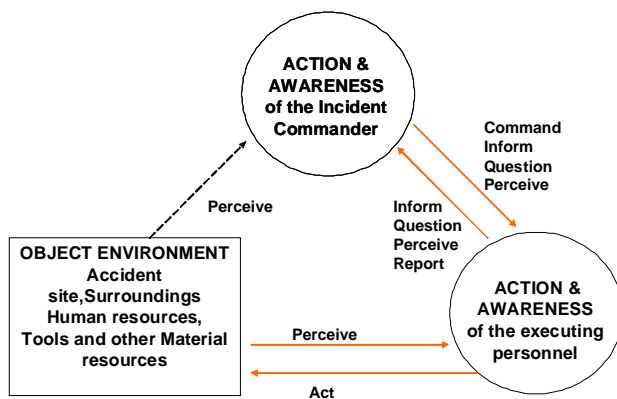


Figure 2. The process of formation of Common Operational Picture from the viewpoint of the incident commander.

The analysis revealed how, as a result of action-perception cycles the incident commander developed a COP. The participants in the communication structure became evident and also which parts of the environment and which media were involved. In the present practice it appears that the command vehicle is the overall locus of the COP. The IC did hardly leave his command position in the vehicle. Two forms of representation were dominant, the incident commander's memory and the on-line dialogue with the command vehicle driver who remained in the vehicle during the incident and offered continuous help to the IC. The main communication medium was TETRA radio.

Design Practice Innovations

Designing joint systems assumes innovations in the design process. In our case the WSN designers and the human factors experts agreed to accomplish a design experiment in which the added value of WSN technologies was considered. Testing was accomplished in real-like environments to gain understanding of the feasibility of the technology and realistic view of a larger scale accident, the leak of hazardous substance. Three fire fighter units were involved and one ambulance unit. In this exercise at the practice field of the Emergency Services College 20 first responders were involved. They acted under the command of one incident commander.

The specific innovation of the tests was the new testing scheme. It was called the Parallel Augmented Exercise. We organised two ER activities that were run simultaneously. The first one represented the "present practice" and was accomplished in complete form applying well learned methods (see also previous section). This activity provided a reference to the parallel second group, the "augmented practice" group. The task of two experienced Incident Commanders was to apply the WSN technology against the actual events, and to make on-line observations of the applicability of the WSN in the situation. With the help of this procedure we were able to create a design space where it was possible to observe and imagine changes in both the technology and the practices. After the tests the trainer evaluated the successfulness of the fire fighting in the "present practice" and the ICs of the "augmented practice" were interviewed about their experience of the new WSN technology.

DISCUSSION

The communication-oriented analysis of the COP appeared a promising possibility to capture COP empirically. As this approach provides a possibility to connect human, environment and technology in a same structure it appears one possibility to define joint cognitive systems.

REFERENCES

- [1] Crandall, B., Klein, G. A. and Hoffman, R.R. 2006. Working minds: A practitioner's guide to cognitive task analysis. Cambridge, MA: MIT Press.
- [2] Woods, D. and Hollnagel, E. 2006. Joint cognitive systems – patterns in cognitive systems engineering. Boca Raton: Taylor & Francis.
- [3] Schraagen, J.-M., Klein, G. A. and Hoffman, M. 2008. The macrocognition framework of naturalistic decision making. In: Naturalistic Decision Making and Macrocognition. Schraagen, J.-M. et al. (Eds.). Ashgate: Aldershot. Pp. 3–25.
- [4] Clark, H. H. and Brennan, S. E. 1991. Grounding in communication. In: Perspectives on socially shared cognition. Resnik, L. B., Levine, J. M. and Teasley, S. D. (Eds.). American Psychological Association: Washington, DC.

Decision Making with a Time Limit: The Effects of Presentation Modality and Structure

Yujia Cao
Human Media Interaction
University of Twente
P.O. Box 217, 7500 AE,
Enschede, The Netherlands
y.cao@utwente.nl

Mariët Theune
Human Media Interaction
University of Twente
P.O. Box 217, 7500 AE,
Enschede, The Netherlands
m.theune@utwente.nl

Anton Nijholt
Human Media Interaction
University of Twente
P.O. Box 217, 7500 AE,
Enschede, The Netherlands
a.nijholt@utwente.nl

ABSTRACT

In this study, a user experiment was conducted to investigate the effects of information presentation factors (modality and structure) on decision making behavior, using a time-limited task. The time constraint required subjects to develop heuristic strategies to substitute the defined normative strategy. The two presentation factors have been shown to significantly affect the decision making performance, assessed by time efficiency and accuracy. The modality factor mainly influenced the time efficiency, due to its impact on the efficiency of information perception. By analyzing the subjective reports and the error distribution, the structure was shown to influence the selection of heuristic strategies. Consequentially, it affected both the time efficiency and the accuracy of decision making. The interaction between the time constraint and the presentation effects was also observed.

Keywords

information presentation, modality, structure, decision making, time stress

ACM Classification Keywords

H.5.1 [Information Interfaces and Presentation (e.g., HCI)] Multimedia Information Systems; H.1.2 [Models and Principles] User/Machine Systems – *Human information processing*.

INTRODUCTION

The influence of information presentation formats on decision making processes has been an important research topic in various fields, such as human-computer interaction, user interface design, economics and marketing. Information presentations are neither only input signals to human cognitive processes nor only extensions of human memory. They guide, constrain, and even determine cognitive behavior [17]. It has been shown that decision makers tend to adapt their manner of information acquisition and their decision making strategies to the way the task is presented, such as the use of modalities and the spatial layout (structure) of the presentation. The adaptation is believed to

be guided by a cost-benefit analysis, compromising between the desire to minimize cognitive effort (cost) and the desire to maximize the accuracy (benefit) [5, 8].

Comparing the presentation of a dataset using tables and graphs, Speier [12] showed that graphs could better assist the acquisition or evaluation of precise data values, as well as the holistic analysis of data relationships and trends. This effect was especially strong when the task was complex. Schkade et al. [10] used numbers and words to present equivalent numerical information, and found that words required more processing effort than numbers. In addition, when words were used, subjects conducted more compensatory and arithmetic activities and less information search activities. Stone et al. [13, 14] investigated the effects of modality on risk-taking decisions. The risk of using a cheaper but less safe tire and a safer but more expensive tire were presented with different modalities. Results show that presenting risk information graphically (with images or graphs) as opposed to numerically (with numbers) increases risk-avoiding decisions, because images and graphs highlight the number of people harmed, thus enhancing the perception of risk.

The spatial structure of the presentation also has been shown to influence decision making [2, 5, 10, 15]. A commonly investigated task is the multi-attribute choice task, which is to select one alternative from several, where each alternative has several attributes. Information can be presented by alternatives or by attributes, using a table or a list. Most studies consistently found that when information was organized by alternatives, subjects tended to process an alternative before considering the next alternatives; when information was organized by attributes, subjects tended to compare all alternatives on a single attribute before considering the next attribute. Schkade [10] shows that the decisions were made faster with the by-attribute structure, and the accuracy was not affected. In contrast, the by-alternative led to more accurate and time efficient decisions in [15].

Previous findings were commonly obtained under a condition where no time limitation was set to the decision making task. However, decision making is very often time-limited in real-life situations. Studies on time-limited decision making behavior suggest that decision makers tend to focus on the general outline of the problem instead of in-depth analysis when time stress sets in [3]. Using the multi-attribute choice task in particular, a strategy switch was observed from being more alternative-based (depth-first) to more attribute-based (breadth first) [8]. In addition, decision makers are also prone to selectively use subsets of the information, adopt simpler modes of information processing and base their decisions on certain important 'cues' [4, 6].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

In this study, a user experiment was conducted to investigate the impact of information presentation on time-limited decision making, when information can only be partially processed and heuristic rather than normative decision making strategies⁴ are applied. We used a multi-attribute choice task with a clearly defined normative strategy and outcome. On the one hand, the time limitation made the application of heuristic strategies necessary; on the other hand, the selection of heuristic strategies was constrained by the requirement of reaching the same outcome as the normative strategy. The task was embedded into a crisis medical rescue scenario in order to create a context motivating the time limitation. However, it was not our intention to have a realistic medical rescue setup, nor did we expect subjects to have knowledge about medical treatment. We intended to observe the effect of presentation modality and structure on the decision making performance, assessed in terms of time efficiency and accuracy. In addition, we were also interested in subjective perceptions of the different presentation formats, and the influence of information presentation on the subjects' choice of decision making strategy. Finally, we looked into the effect of information presentation format on tasks with different levels of difficulty, where time constraints play a bigger or smaller role.

PRESENTATION AND TASK

The decision making task was set up using an earthquake crisis scenario where the number of wounded people exceeded the capacity of medical resources (equipment and staff). Therefore, the order of treatment needed to be determined as fast as possible, based on the evaluation of the patients' injuries.

Presentation Materials

A pair of patients was presented at a time. The injury condition of a patient was described by five injury categories (derived from [9]): heart failure, respiration obstruction, blood loss, brain damage and fracture. The first three categories were described as more threatening, and thus more important than the last two. The severity of each injury category was described at one of four levels (derived from [11]): severe, moderate, mild or none.

The two presentation factors were *modality* (text or image) and *structure* (by-injury or by-severity), resulting in four different presentation conditions. In the two text conditions, the injury categories and severity levels were presented with English text. In the two image conditions, injury categories were presented by icon images of the affected organs (e.g. an icon image of a brain referred to the 'brain damage' item), and severity levels were presented by color rectangles (red for 'severe', orange for 'moderate', yellow for 'mild' and green for 'none').⁵ The injury information of two patients was organized in a table. The table could be structured by the injury categories or by the severity levels. When using the by-injury structure, the more important three injury categories were located on top of the less important two. The injury column was fixed for all tasks and the severity values varied. When using the by-severity structure, the four severity levels were ranked from high to low. A higher severity level was located more on top of the table. The severity column was fixed for all tasks and the

⁴ Normative strategies apply a careful and reasoned examination of all alternatives and attributes. Heuristic strategies are simple and fast rules of thumb [8].

⁵ Strictly speaking, color rectangles are not images. However, in this experiment, we use "image" to generally refer to non-verbal visual modalities.

locations of injury categories varied. Figures 1–4 demonstrate the four presentations of an identical patient pair.

Injury	Patient 1	Patient 2
Heart failure	mild	severe
Blood loss	severe	mild
Respiration obstruction	none	moderate
Brain damage	none	none
fracture	severe	none

Figure 1. A patient pair presented in the text modality and the by-injury structure.

Severity	Patient 1	Patient 2
Severe	Blood loss fracture	Heart failure
Moderate		Respiration obstruction
Mild	Heart failure	Blood loss
None	Respiration obstruction Brain damage	Brain damage fracture

Figure 2. A patient pair presented in the text modality and the by-severity structure.






Injury	Patient 1	Patient 2
	Yellow	Red
	Red	Yellow
	Green	Orange
	Green	Green
	Red	Green

Figure 3. A patient pair presented in the image modality and the by-injury structure. The text of colors was added here to ensure the readability in a grayscale printing.











Severity	Patient 1	Patient 2
Red	 	
Orange		
Yellow		
Green	 	 

Figure 4. A patient pair presented in the image modality and the by-severity structure. The text of colors was added here to ensure the readability in a grayscale printing.

Task

The subjects played the role of a medical manager in the emergency medical treatment center. The task was to compare the injuries of pairs of patients and select the more seriously injured patients to be treated first. Based on a pilot study, the time limit for each decision was set to 20 seconds. A speech reminder was given after 15 seconds. The information was removed from the screen when time was up so that a decision was forced to be made even if the analysis was not yet completed.

The Normative Strategy

The normative strategy evaluates the overall injury level of a patient by a linear utility function. The attributes of the function are severity values of the five injury categories. The severity level *severe* was described as three times as important as *mild*, and *moderate* twice as important as *mild*. Therefore, the severity values can be considered as 3, 2, 1 and 0 for *severe*, *moderate*, *mild* and *none*, respectively. Moreover, the attributes have different weights, because the more important three injury categories (heart failure, blood loss and respiration obstruction) were considered to be twice as important as the other two. Finally, the over-all injury evaluation of a patient was quantified as equation 1, where ‘Se’ refers to severity value:

$$Injury_{Norm} = 2 \times Se_{heart} + 2 \times Se_{blood} + 2 \times Se_{respiration} + Se_{brain} + Se_{fracture} \quad (1)$$

When comparing two patients, the one with the highest injury value should be treated first. For the patient pair in Figure 1, the injury value is 11 ($2 \times (1 + 3 + 0) + 0 + 3$) for patient 1 and 12 ($2 \times (3 + 1 + 2) + 0 + 0$) for patient 2. Therefore, the correct decision is to treat patient 2 first. To quantify the processing load of this strategy, the number of elementary information processing operations (EIPs, described in [8]) was calculated. This strategy requires 10 read EIPs (acquiring the values), 8 addition EIPs (summing up the values), 6 product EIPs (weighting operations) and 1 comparison EIP (identifying the larger value between two).

Heuristic Strategies

The 20 seconds time limitation requires subjects to be fully engaged in the task. In most cases there will be insufficient time to apply the normative strategy (equation 1). All intermediate outcomes of the calculation should be kept in the short-term memory which also increases the cognitive load of the normative strategy. Therefore, simpler heuristic strategies are likely to be applied. Various heuristic strategies with different levels of accuracy could be developed for this task. Unbiased heuristic strategies always lead to the correct outcomes, and thus are efficient and accurate decision making “shortcuts”. However, biased heuristic strategies might enhance the time efficiency but lead to wrong decisions.

Figure 5 shows an example of an unbiased heuristic strategy which uses compensatory eliminations to reduce the amount of calculation needed. The method is to identify two injury items that 1) are from different patients; 2) have the same severity level; and 3) belong to the same priority group. Such two items have the same contribution to the comparison of the two injury values, and thus can be eliminated from the calculation. When all possible eliminations are done, the remaining items are calculated for a final choice. Note that “none” items have a value of 0 and can be ignored as well. In this example, the moderate respiration obstruction of patient 2 has a value of 4 (2×2) and the severe fracture of patient 1 has a value of 3. Therefore, patient 2 is the correct choice. In total, there are 10 read EIPs, 1 product EIP and 3 comparison EIPs (two

eliminations and one final choice). The total number of EIPs (14) is only 56% of using the normative strategy (25). The unbiased heuristic strategy was introduced to the subjects in the introduction session as an inspiration. They were informed that they could freely apply their own strategies to reach the correct decisions in time.

Biased strategies might be developed for this task as well. For example, one might ignore the injury categories with the lower priority and only consider the most important three injury categories. One could also ignore the priority rules and treat all five injury categories equally. These biased strategies can reduce the calculation load but cannot guarantee a correct outcome.

Injury	Patient 1	Patient 2
Heart failure	mild	severe
Blood loss	severe	mild
Respiration obstruction	none	moderate
Brain damage	none	none
fracture	severe	none

} ×2

Figure 5. An example of an unbiased heuristic strategy.

EXPERIMENT

Experimental Design

We used a within-subject 2×2 design. The two independent factors were *modality* at two levels (text or image) and *structure* at two levels (by-injury or by-severity). Therefore, all subjects performed four experimental trials, namely the ‘text & by-injury’, ‘image & by-injury’, ‘text & by-severity’ and ‘image & by-severity’ trial. The trial order was counterbalanced with a size 4 Latin Square. Each trial contained 12 tasks which were identical for all four trials but randomly ordered.

Dependent Measurements

The decision making performance was measured by two dependent variables, namely *time efficiency* and *accuracy*. The *time efficiency* of a decision refers to the time interval between the moment when a task is presented and the moment when the decision is made (in seconds). A decision is *accurate* if it is identical to the outcome from the normative strategy. These two measurements can be retrieved from the log files of subjects’ mouse clicks.

Subjective experience was obtained by questionnaires. The questionnaire consisted of two parts. The first part asked subjects to perform four binary comparisons on the cognitive demand of the task as shown in Table 1. A sample of the four presentation conditions (Figures 1–4) was provided as a reference to the questions. The question was for example: “Comparing condition 1 and 3, in which condition did the task require less cognitive effort (was easier)?”

Table 1. The four binary comparisons on the cognitive demand of the decision making task.

Comparisons		Reference
Between modalities (Text vs. Image)	structure = by-injury	Fig. 1 vs. 3
	structure = by-severity	Fig. 2 vs. 4
Between structures (By-injury vs. By-severity)	modality = text	Fig. 1 vs. 2
	modality = image	Fig. 3 vs. 4

In addition, the subjects were also asked to indicate which presentation conditions they found the easiest and the most difficult. The second part of the questionnaire is related to the decision making strategies. Subjects were asked to orally describe the strategies they had used in each presentation condition. Those were written down by the experimenter during the description.

Subject and Procedure

32 university students (graduate students and PhD students) volunteered to participate in the experiment. All of them were fluent English speakers and none of them had a medical background.

The experiment contained three sessions: an introduction session, a training session and the experimental session. The introduction described the rescue scenario, the task and the four presentation styles, the normative decision making strategy and the unbiased heuristic strategy. In the training session, subjects practiced 20 tasks, 5 tasks for each presentation style. No time limit was used. Feedback on the decision accuracy was given after each decision was made, via speech. After training, subjects were required to perform four experimental trials of totally 48 tasks (4×12). A performance summary was given after each trial, announcing how many correct decisions had been made. After the four trials were all finished, subjects were required to complete the questionnaire. The time duration of the experiment was about 40 minutes.

Hypotheses

According to the cognitive fit theory [16], presentation manners that provide a better cognitive fit to the nature of the task can better assist the making of more accurate and less effortful decisions. The modality factor certainly has an impact on the information perception effort and quality. Regarding the representative strength, text is suitable for conveying abstract information, such as the relationships between events; while images are suitable for describing concrete concepts and information of a highly specific nature, such as concrete objects [1]. Therefore, the images of organs were expected to be more suitable than text for presenting the injury categories. Furthermore, shapes and colors have great salience to human information processors due to the sharp contrast they are able to create [7]. Compared to text, the color coding was expected to be better able to reflect the difference in the severity levels and assist comparisons.

One key step of this task is to separate the two priority groups. Only when this separation is clear, the weight and the elimination method (Section: "Heuristic Strategies") can be applied. When information is presented with the by-injury structure, this separation does not require any effort since the more important three injury categories are located above the other two. In contrast, the by-severity structure does not particularly support this priority separation, which consequentially complicates the application of weights and eliminations. Therefore, we expected the by-injury structure to be more cognitively compatible with the task than the by-severity structure.

We assumed that the cognitive advantage of a certain modality/structure over another is particularly pronounced when the decision making task is time-limited. Accordingly, the following two hypotheses were built:

1. The modality factor has an effect on the decision making performance. The time efficiency and accuracy are both higher in the image conditions than in the text conditions.

2. The structure factor has an effect on the decision making performance. The time efficiency and accuracy are both higher in the by-injury conditions than in the by-severity conditions.

RESULTS

Due to the within-subject design, we applied repeated ANOVAs with *modality* and *structure* as two nested independent factors, on the *time efficiency* and the *accuracy* variable, respectively. The trial order was treated as a between-subject variable and was shown to have no significant effect on either of the two dependent variables.

Decision Making Performance

Time Efficiency

The average time spent on one task (in seconds) in each condition is shown in Figure 6. Subject performed the fastest in the 'image & by-injury' condition, and the slowest in the 'text & by-severity' condition. Repeated ANOVA results showed that 1) there was no significant interaction between the two factors, $F(1, 31) = 0.38, p > 0.5$. This indicates that the effects of these two factors on the time efficiency are independent from each other; 2) the modality factor had a significant effect on the time efficiency, $F(1, 31) = 48.31, p < 0.001$. Subjects performed significantly faster in the image conditions than in the text conditions, regardless of how the information was structured; and 3) the structure factor also has a significant effect on the time measurement, $F(1, 31) = 27.84, p < 0.001$. Subjects performed significantly faster when the information was sorted by injury categories than by severity levels, regardless of which modality was used. Thus, for both modality and structure our hypotheses regarding time efficiency were confirmed.

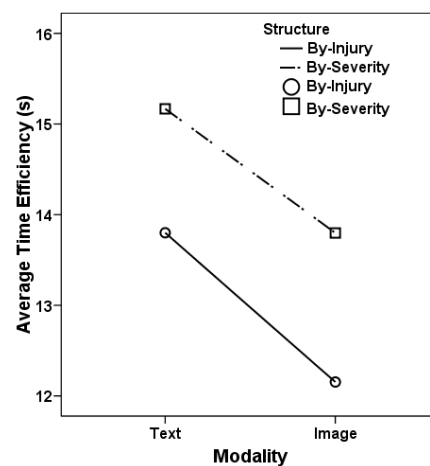


Figure 6. The average time efficiency of four conditions.

A post-hoc test (Bonferroni test) further revealed five significant pair-wise effects, as shown in Table 2. Significant differences in the time efficiency variable occurred between all pairs of conditions, except between the 'text & by-injury' condition and the 'image & by-severity' condition. One of them has the more suitable modality (allows faster performance) but the less suitable structure; the other one has the more suitable structure but the less suitable modality. For each of the two conditions, the disadvantage counteracts the advantage, leading to a comparable average time efficiency (see Figure 6).

Table 2. Pair-wise comparisons on the time efficiency measurement by Bonferroni test (only significant results).

Pair-wise effects	Sig.	Factor involved
Lower: Text & By-Severity Higher: Image & By-Severity	$p < 0.001$	modality
Lower: Text & By-Injury Higher: Image & By-Injury	$p < 0.001$	modality
Lower: Text & By-Severity Higher: Text & By-Injury	$p < 0.001$	structure
Lower: Image & By-Severity Higher: Image & By-Injury	$p < 0.01$	structure
Lower: Text & By-Severity Higher: Image & By-Injury	$p < 0.001$	modality & structure

Accuracy

The average number of correct decisions made in each trial is shown in Figure 7. Subjects made the most correct decisions in the ‘image & by-injury’ condition, and the least correct decisions in the ‘text & by-severity’ condition. ANOVA results show that 1) there was no significant interaction between the two factors, $F(1, 31) = 0.07, p > 0.5$, indicating that the effects of modality and structure on the decision accuracy were independent from each other; 2) the modality factor did not have an effect on the accuracy measurement, $F(1, 31) = 2.26, p > 0.1$; and 3) the structure factor had a significant effect on the decision accuracy, $F(1, 31) = 4.16, p < 0.05$. Subjects made significantly more correct decisions when the information was structured by injury categories than by severity levels, regardless of which modality was used. Thus, our hypotheses regarding accuracy were only confirmed for the structure factor, but not for the modality factor.

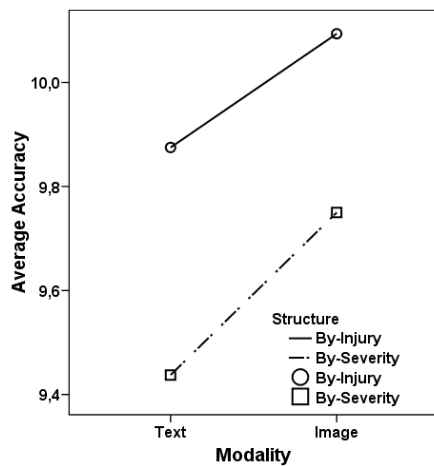


Figure 7. The average decision accuracy of four conditions.

One significant pair-wise effect was found in the post-hoc test. Significantly more correct decisions were made in the ‘image & by-injury’ condition than in the ‘text & by-severity’ condition. This indicates that when the more suitable modality and the more suitable structure were combined, subjects performed significantly more accurately than when the less suitable modality and the less suitable structure were combined.

In summary, the two performance measurements reveal a significant modality effect as well as a significant structure effect. The modality factor influences the time efficiency. Image allows faster performance than text. The structure factor

affects both time efficiency and accuracy. The by-injury structure allows faster and more accurate performance than the by-severity structure. There is no interaction between these two presentation factors.

Subjective Comparisons

The results of subjective comparisons of the cognitive demand of the task under different presentation conditions are summarized in Figures 8–10. Generally speaking, these subjective judgments are consistent with the results of the performance measurements. The ‘text & by-severity’ condition was considered as the most difficult condition by 19 (59.4%) subjects and the ‘image & by-injury’ condition was considered as the easiest condition by 21 (65.6%) subjects (Figure 8). Twenty-six (81.3%) subjects found the task less demanding in the image conditions than in the text conditions, regardless of the structure factor (Figure 9). Twenty-two (68.8%) subjects preferred the by-injury structure to the by-severity structure, regardless of the modality factor (Figure 10).

Apart from the majority preferences, 4 subjects preferred text to image (Figure 9) and also pointed out one of the text conditions to be the easiest one (Figure 8). Further looking into their performance, we found that their decision accuracy was indeed higher in the text condition than in the image conditions (with the same structure). We noticed that three people out of these four have a daily research topic that is clearly text or speech oriented. Although lacking of solid experimental evidence, this observation still suggests that in addition to the generally applicable guidelines (such as ‘images are more suitable than text to present concrete information’), the professional background might also be a useful reference for the usage of modality, especially when the interface is designed for a specific user group.

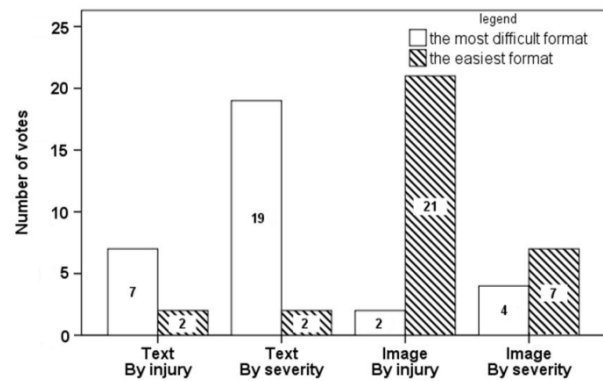


Figure 8. Subjective reports of the easiest and the most difficult presentation conditions.

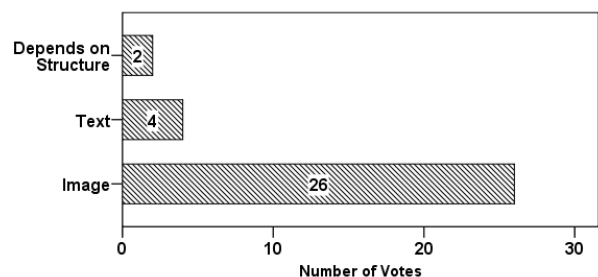


Figure 9. Voting results of the cognitive load comparisons between a text condition and an image condition.

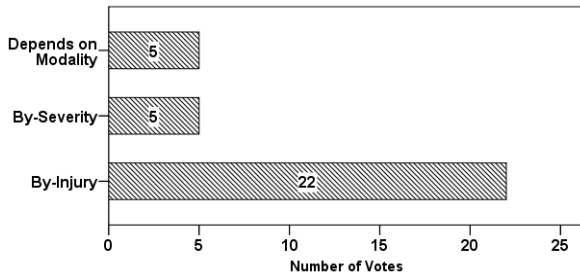


Figure 10. Voting results of the cognitive load comparisons between a by-injury condition and a by-severity condition.

Figure 10 also shows that there are 5 subjects who preferred the by-severity structure instead of the by-injury structure. Again, their performance data shows that they indeed made more correct decisions in their favored structure than in the other (when modality was the same). Therefore, it seems that the subjective preference can very well reflect the decision making accuracy. As a part of the future work, we need to compare all subjective judgments with the associated performance data, aiming at identifying the correlations between them. On the one hand, we have generally applicable guidelines which can lead to a standardized optimal design for a majority of users. On the other hand, in cases where the task performance of every single user is critical (such as when crisis managers work with a realistic crisis support system), additional customized references might be needed in order to enhance the general guidelines. The individual subjective preferences seem like a promising direction to look into.

Strategy Analysis

Subjective Reports

Most of the subjects were able to clearly describe how they processed the information to reach the final decision. Two general trends could be easily recognized from these subjective strategy descriptions. First, the normative calculation model was never applied, meaning that subjects all tried to apply heuristic methods to solve the tasks. Second, the unbiased heuristic strategy introduced to the subjects was generally accepted and applied. However, if the remaining time was perceived to be insufficient, subjects sometimes turned to biased strategies in order to reach a real quick decision. The development of biased strategies was influenced by the structure factor rather than the modality factor.

In the by-injury structure, since the separation of priority group was clear, subjects commonly mentioned that they carried out a careful analysis of the high-priority group, and only if there was still some time left, then they also had a quick glance at the low-priority group. But in most cases, this quick processing didn't change the decision. Basically, this biased strategy (BS1) bases the decision on the more important three injury categories only, as shown in equation 2. Since these three categories do contribute more to the overall injury value, the outcome has a good chance to be correct, but not always.

$$Injury_{BS1} = Se_{heart} + Se_{blood} + Se_{respiration} \quad (2)$$

In the by-severity structure, realizing the complication to separate the priority groups and apply elimination method, subjects mentioned that they first had a look at the distribution of the injury items and compared which patient's injuries were located more towards the top side of the table (the more severe side). Then started to carefully identify the priorities and apply

eliminations. However, very often they could not completely finish so that the primary decision remained. This biased strategy (BS2) ignores the priority definition and treats all injury categories equally, as shown in equation 3.

$$Injury_{BS2} = Se_{heart} + Se_{blood} + Se_{respiration} + Se_{brain} + Se_{fracture} \quad (3)$$

Quantitative Analysis

Based on the analysis of subjective report, we took one step further to look for quantitative evidence of these two biased strategies being applied. The attempt was also to further confirm the influence of the strategy factor on the development of strategies. First of all, we applied these two strategies to all 12 tasks. Results showed that each of them could reach 8 correct decisions out of the 12. Accordingly, we defined the following four task groups:

1. BS1-Correct group: 8 tasks
2. BS1-Wrong group: 4 tasks
3. BS2-Correct group: 8 tasks
4. BS2-Wrong group: 4 tasks.

Second, for each presentation condition, we calculated how many correct decisions were actually made by the subjects within each group (in percentage). The average from all subjects is presented in Table 3.

Table 3. The average percentage of correct decisions made within each group under each presentation condition.

Presentation Condition	Percentage of correct decisions (%)			
	BS1 Correct	BS1 Wrong	BS2 Correct	BS2 Wrong
Text & By-Injury	91.4	64.1	80.8	85.2
Image & By-Injury	92.4	63.5	84.4	83.6
Text & By-Severity	82.4	71.1	87.9	63.9
Image & By-Severity	83.2	77.3	86.7	69.3

Next, we compare the results between the BS1-Correct group and the BS1-Wrong group. According to the average values in Table 3, we can see that the difference is larger in the two by-injury conditions than in the two by-severity conditions. Results of T-tests further confirmed that when the by-injury structure was used, significantly more correct results were made within the BS1-Correct group than the BS1-wrong group (Table 4). However, no such effect was found when the by-severity structure was used.

When applying the same comparison between the BS2-Correct and the BS2-Wrong group, reversed results were obtained (Table 4), indicating that significantly more correct decisions were made within the BS2-Correct group when the by-severity structure was used, but no effect was found in the by-injury conditions.

A tentative conclusion that can be drawn from these results is that the decision accuracy was influenced by the application of BS1 when the by-injury structure was used and the application of BS2 when the by-severity structure was used. This in turn means that the structure factor indeed to some extent influenced the development of decision making strategies.

Table 4. T-tests result for identifying the application of BS1 and BS2 in all presentation conditions.

Presentation Condition	T-test pairs	
	BS1-Correct vs. BS1-Wrong	BS2-Correct vs. BS2-Wrong
Text & By-Injury	$p < 0.001$	$p > 0.1$
Image & By-Injury	$p < 0.001$	$p > 0.5$
Text & By-Severity	$p > 0.1$	$p < 0.05$
Image & By-Severity	$p > 0.1$	$p < 0.05$

Time Constraint: Low vs. High

The results presented so far have already shown that the presentation factors, modality and structure in particular, had an effect on the time-limited decision making performance. However, we were interested in further exploring the interaction between different levels of time constraint and the presentation effects. Since our experiment setup did not include multiple levels of time limitation, this interaction cannot be directly investigated. However, the 12 decision making tasks were not identically difficult. The difficulty level of a task was assessed by the difference in the overall injury values of a patient pair (calculated by equation 1). The larger the difference is, the easier/quicker it is to identify which patient has more severe injuries. Therefore, the time constraint could be considered as weaker for easier tasks and stronger for more difficult tasks. In this case, if the time efficiency and accuracy are analyzed separately for tasks at different difficulty levels, we might be able to indirectly observe the interaction between the time limit and the presentation effects.

The 12 tasks were assigned into two groups. For the 8 tasks in the more difficult group, the difference between the two overall injury values is below 3. In the relatively easier group, the difference is between 5 and 10 for the 4 tasks. The time efficiency and accuracy were re-calculated respectively for the two groups (Figure 11).

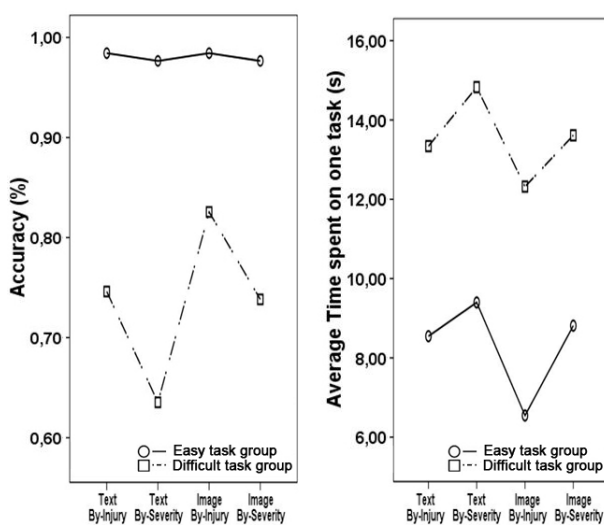


Figure 11. The accuracy (left) and time efficiency (right) calculated respectively for the easy and difficult task group.

As expected, in the relatively easy task group the performance was both more accurate and faster. There were only about 2% of errors among the easy tasks, and most of them occurred in the by-severity conditions. There was no significant modality or structure effect on accuracy. ANOVA on the time efficiency measurement did show a modality effect ($F(1, 31) = 22.5, p < 0.001$) and a structure effect ($F(1, 31) = 16.2, p < 0.001$). This indicates that when the time constraint was relatively weak, the decision accuracy was almost unaffected by the quality of presentation, since the subjects could take their time to make the correct decisions. However, the cognitive benefit of good presentations was still reflected by the time efficiency of decision making.

When the tasks were more difficult, the time allowed to make a decision was no longer sufficient to complete the unbiased but more demanding decision making processes, resulting in a general decrease of accuracy in all presentation conditions. In such a situation, the presentation factors showed an even stronger impact on the decision making performance, since they influenced both the accuracy and the time efficiency. When the presentation manner is more cognitively compatible with the task, the decisions are made faster and more accurate. In addition, it can be observed from Figure 11 (left) that the accuracy showed different levels of tolerance towards the increase of the task difficulty. The better the presentation condition is, the less the accuracy drops between the easy and difficult task groups.

CONCLUSIONS AND FUTURE WORK

In this study, we investigated the effects of information presentation on time-limited decision making, focusing on the modality and the structure factors. The decision making performance was assessed in term of time efficiency and accuracy. The subjective judgments of various presentation formats were also obtained. In addition, we also investigated the influence of presentation factors on the subject's choice of decision making strategy. Finally, we looked into the interaction between presentation effects and the time constraint, by analyzing the performance of tasks at different levels of difficulty, where time constraints play a bigger or smaller role.

Regarding the modality factor, our result is in line with the previous studies and confirms that modality has an impact on the decision making performance. Additionally, we suggest that the modality factor influences the time efficiency more than the accuracy. A suitable modality accelerates the decision making process by decreasing the effort and increasing the quality of information perception. However, this does not necessarily lead to a higher accuracy, because the selection of decision making strategies is not determined by the usage of modality. Generally, modality selection should aim at providing a cognitive fit to the perception task. When visual search among different types of objects is required, images are usually more suitable than text for presenting those objects. When different levels of severity (or urgency, importance etc.) need to be perceived, colors can be a very effective presentation option.

The structure factor has been shown to have a significant impact on both the time efficiency and the accuracy of decision making. This is mainly because of its influence on the selection of strategies. When the time constraint does not allow the most accurate but demanding strategy to be used, subjects develop heuristic strategies in order to make a decision in time. When a structure does not provide a good cognitive fit to the task, more cognitive effort is needed to perform the task. Then, less effortful strategies are more likely to be chosen, which are

normally also less accurate. Therefore, the presentation structure should assist the application of unbiased decision making strategies. If several information items are required to be considered as a group, they need to be spatially clustered. If a table is used, locate the more critical information items more on the top.

Regardless of the level of time constraint, the presentation factors always have an impact on the cognitive demand of the decision making task. However, this impact is stronger when the time constraint is stronger. In this experiment, for the relatively easier group of tasks, only the time efficiency was influenced by the presentation factors; while the accuracy stayed high. However, for the group of difficult tasks, both time efficiency and accuracy showed a presentation effect. The decrease of accuracy was less when the presentation format was more cognitively compatible to the task.

Our future work involves three aspects. First, as mentioned in Section: “Subjective Comparisons”, the relation between performance measurements (especially accuracy) and subjective judgments needs further investigation. Second, in order to directly observe the interaction between the time constraint and the presentation effect, this experiment needs to be replicated without the time limit or a new experiment needs to be carried out with multiple levels of limits. Third, we noticed that subjects commonly didn’t make a full use of the 20 seconds that were offered to them. When the ‘5 seconds remaining’ warning was delivered, some subjects appeared very stressful and they made their choices immediately after the warning speech started. It seems that the level of time stress was perceived to be higher than it really was, and this perception was individually different. However, none of our measurements allowed the assessment of stress. Therefore future work is needed to obtain a deeper understanding of the perceived stress induced by the time constraint.

ACKNOWLEDGMENTS

This research is part of the Interactive Collaborative Information System (ICIS) project, which is sponsored by the Dutch Ministry of Economic Affairs, grant nr: BSIK03024. We thank the 32 participants for their effort and time.

REFERENCES

- [1] André, E. The Generation of Multimedia Presentations. *Handbook of Natural Language Processing*, 2000, 305–327.
- [2] Bettman, J. and Kakkar, P. Effects of Information Presentation Format on Consumer Information Acquisition. *Journal of Consumer Research*, 3(4), 1977, 233–240.
- [3] Dörner, D. and Pfeifer, E. Strategic Thinking and Stress. *Ergonomics*, 36(11), 1993, 1345–1360.
- [4] Keinan, G., Friedland, N. and Ben-Porath, Y. Decision Making Under Stress: Scanning of Alternatives Under Physical Threat. *Acta Psychologica*, 64(3), 1987, 219–228.
- [5] Kleinmuntz, D. N. and Schkade, D. A. Information Displays and Decision Processes. *Psychological Science*, 4(4), 1993, 221–227.
- [6] Levin, I. P., Huneke, M. E. and Jasper, J. D. Information Processing at Successive Stages of Decision Making: Need for Cognition and Inclusion–Exclusion Effects. *Organizational Behavior and Human Decision Processes*, 82(2), 2000, 171–193.
- [7] Lurie, N. H. and Mason, C. H. Visual Representation: Implications for Decision Making. *Journal of Marketing*, 71, 2007, 160–177.
- [8] Payne, J. W., Bettman, J. R. and Johnson, E. J. *The Adaptive Decision Maker*. Cambridge University Press, New York, 1993.
- [9] Porter, K., Shoaf, K. and Seligson, H. Value of Injuries in the Northridge Earthquake-Technical Note. *Earthquake Spectra*, 22(2), 2006.
- [10] Schkade, D. A. and Kleinmuntz, D. N. Information Displays and Choice Processes: Differential Effects of Organization, Form, and Sequence. *Organization Behavior and Human Decision Processes*, 57, 1994, 319–337.
- [11] Shoaf, K., Seligson, H., Peek-Asa, C. and Mahue-Giangreco, M. Standardized Injury Categorization Schemes for Earthquake Related Injuries. *UCLA Center for Public Health and Disasters*, Los Angeles, CA, 43, 2001, 1–43.
- [12] Speier, C. The Influence of Information Presentation Formats on Complex Task Decision-making Performance. *International Journal of Human-Computer Studies*, 64(11), 2006, 1115–1131.
- [13] Stone, E. R., Sieck, W. R., Bull, B. E., Yates, F. J., Parks, S. C. and Rush, C. J. Foreground: Background Salience: Explaining the Effects of Graphical Displays on Risk Avoidance. *Organizational Behavior and Human Decision Processes*, 90(1), 2003, 19–36.
- [14] Stone, E. R., Yates, J. F. and Parker, A. M. Effects of Numerical and Graphical Displays on Professed Risk-Taking Behavior. *Journal of Experimental Psychology Applied* 3, 1997, 243–256.
- [15] Tabatabaei, M. An Experimental Analysis of Decision Channeling by Restrictive Information Display. *Journal of Behavioral Decision Making*, 15(5), 2002, 419–432.
- [16] Vessey, I. and Galletta, D. Cognitive Fit: An Empirical Study of Information Acquisition. *Information Systems Research*, 2(1), 1991, 63–84.
- [17] Zhang, J. The Nature of External Representations in Problem Solving. *Cognitive Science*, 21(2), 1997, 179–217.

Posters

Levels of Automation and User Control in Use of an Automatic Turbine System

Jonas Andersson
MSc, PhD Student
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
jonas.andersson@chalmers.se

Anna-Lisa Osvalder
Associate Professor, PhD
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
alos@chalmers.se

ABSTRACT

This paper describes an empirical field study performed in a nuclear power plant simulator control room and presents how the use of an automatic turbine system (ATS) affects nuclear power plant turbine operators' ability to stay in control during different levels of automation. The paper also presents how the operators cope with the automation interface and how their problem solving ability is affected by varying levels of automation. The Contextual Control Model (COCOM) was used to analyse the turbine operators' work with the ATS. The aims were to investigate how the ATS design support the turbine operators in their work in terms of monitoring and controlling the turbine process, and to identify possible improvements in the ATS user interface design. Seven turbine operators were interviewed during the simulator training session. The results of the interviews point out that automation related problems such as out-of-the-loop performance and loss of skills exist in the control room setting. The use of COCOM as a means for analysis provided explanations to these results and implied that time for evaluation is an important factor for effective performance. Finally, improving the visibility of the underlying program logic was found to be the most important measure to enhance the ATS interface.

Keywords

human factors, levels of automation, control room

ACM Classification Keywords

J.7 [Computers in other systems] Command and control;
H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *user-centered design*.

INTRODUCTION

Automation technology has changed human activity in process control from manual work on the plant floor to distant supervisory control in control room environments. In the case of turbine operation in nuclear power plants, automation plays an important role. Automation offers many advantages in terms of stable control and production efficiency and it also facilitates

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

control room work by relieving the operator of continuous manual actions. However, it has also been shown that automation can create a number of problems such as “out of the loop” performance problems, degradation of skills and inappropriate trust [1]. These effects all reduce operators' ability to stay on top of their working situation and are directly connected to the design and level of automation applied in the technical system.

The purpose of this study was to examine how operator performance is affected by different levels of automation in nuclear power plant turbine operation. The aims were to investigate how the ATS design support the turbine operators in their work in terms of monitoring and controlling the turbine process, and to identify possible improvements in the ATS user interface design.

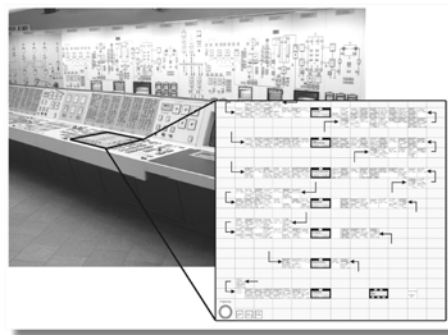


Figure 1. The Automatic Turbine System (ATS) interface.

The automatic turbine system consists of a series of sequences that can direct the turbine equipment from an axis standstill to full operation, where the generator produces electricity to the grid (Figure 1). The automatic system can also be used in the reverse order to bring the turbine to a standstill. It is mainly used during turbine start-up and shut-down. This process takes place through seven main steps that contain approximately ten sub-steps each. The system can be utilized using three different levels of automation; manual mode, step-mode and fully automatic mode. Manual operation corresponds to separate control of each object in the sequences, whereas in step-mode the automation is used to perform sequences although the operator has to acknowledge each step. In full automation the operator only monitors the ATS user interface. Figure 1 shows a picture of the ATS interface and its placement in the nuclear power plant control room.

THEORY

The human information process is usually described as a linear model starting with detection of signals that are processed and leads to a response action. Linear information processing models have many advantages, but they are incomplete since they

depict human actions basically as a response to something that happens [2]. Many human actions are however performed due to expectations of future events, in the near or far future. This is not explicitly accounted for in sequential information processing models although a feedback loop from action back to perception is present. However, process control in general mainly consists of anticipatory actions to avoid unfavourable process states. Therefore, the contextual control model (COCOM) is appropriate when analysing human supervisory control. COCOM has its origin in Neisser’s perceptual cycle [3] and describes how current understanding determines what actions are taken to reach a specific goal. The action in turn produces feedback that modifies the understanding (Figure 2). A central concept in COCOM is time. According to Hollnagel and Woods [4], time is central to stay in control and can be used to model control performance. The main components to model control are; time to evaluate, time to select action and time to perform action. The full set of time components can be found in [4].

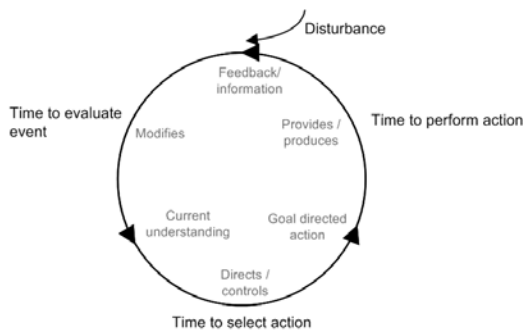


Figure 2. The Contextual Control Model (adopted from [4]).

The description of control is simplified to four different control modes, ranging from scrambled via opportunistic to tactical and strategic control. *Scrambled* control refers to the state where control is practically lost and there is no correspondence between actions and situation. In the *opportunistic* control mode, planning may be limited due to lack of time or poor understanding of the situation. Opportunistic control is typically applied when mental models are inappropriate. The *tactical* control mode refers to situations where performance follows known procedures. In *strategic* control a longer time horizon is present and high level goals can be aimed for. A thorough description of the four control modes can be found in [4].

A number of effects arise when automation is introduced in human-machine systems. In this study, three typical automation related problems were found; out-of-the-loop problems, skill degradation and trust in automation.

Out-of-the-loop performance problems are characterised by how humans find it difficult to detect automation failures and revert to manual control [5]. This depends upon a number of factors. Firstly, automation may reduce feedback from the process. The feedback that exists is also different from when manual control is used. Another factor is that automation puts the operator in passive observation of the process which puts higher demands on operator vigilance. Automatic control also means that the operator can engage and focus on other activities, which makes it even harder for the operator to observe all process feedback. Another cause for out-of-the-loop problems is that the operator has an inadequate mental model that gives false expectations. Altogether, the origin of out-of-the-loop unfamiliarity comes from disrupted feedback that

reduces situation awareness, resulting in false expectations and makes shift to manual control difficult.

Skill degradation refers to how operators tend to loose knowledge and manual skills in highly automated processes [5]. This increases the demands on adequate training and effective procedures to avoid problems in case of an automation failure. When simple physical tasks are replaced by automation and difficult tasks that are too hard to automate are left to human operators, the work becomes more demanding. Automation also makes it possible to handle more tasks simultaneously, which further increases the cognitive workload.

Operators’ trust in automatic systems affects how and if automatic functions are used. [5]. If operator trust does not match the automation’s capabilities, problems with misuse and disuse can occur [6]. If the operator does not trust the automation to perform what is expected in an appropriate way, automation is likely to be abandoned and the advantages of the automatic system are lost. Over-trust on the other hand, occurs when the automation is believed to be more reliable than it actually is, resulting in complacent behaviour. To avoid these effects, trust has to be calibrated so that it matches the actual capabilities of the automation [7].

METHOD

This study was performed using a field study approach, where seven turbine operators from a Swedish nuclear power plant were interviewed. The field study approach was chosen because it gives a realistic view of the working conditions in the control room [8]. The operator crews conducted two eight-hour long training shifts during two days at the training facility. These shifts were divided into different parts where the studied simulator session was one part. The simulator session lasted for approximately three hours and included handling of the ATS as an integrated part of the problem scenario. The operators monitored and performed actions on the ATS at irregular intervals throughout the session as a part of their problem solving. The operators’ working experience varied from being under education to become licensed operators to more than twenty years of experience. Every less experienced operator worked together with an experienced operator in each shift team. The data collection was made using a qualitative approach since the study focused on the operators’ opinions and their collected experiences from working with the turbine automation equipment. The interviews were performed in a semi-structured manner after the completion of the simulator sessions. Each interview lasted for approximately one hour and was focused on the use of the ATS and how varying levels of automation affects the operators’ work and what automation related problems they had experienced. All questions used the recently performed training session as a starting point. The operators were then asked how they *would have been* affected by changes in the level of automation in the specific situation. This was followed by an individual description of difficulties and situations they had encountered while using the ATS in their daily work. After that, discussions followed regarding how the interface design supports operator work.

RESULTS & ANALYSIS

Manual Mode

In manual mode, the operators perform actions without using the ATS interface. Instead, individual objects are manoeuvred according to written procedures to reach the intended process state. Procedures are available for both planned actions and

trouble shooting situations and are used to ensure high reliability and to avoid human error. During normal situations, the operator usually has enough time to analyse the situation, plan actions and choose the appropriate procedure. The implementation of procedures then directs actions towards the intended goal. In manual tasks, the relationship between action and feedback is clear. Verification of the process response using the procedure is also easy since individual objects can be followed closely. This strengthens the operator's knowledge and understanding of the systems functionality and enhances the feeling of control (Figure 3).



Figure 3. Manual mode.

During anomalies, the operator's knowledge and experience is the foundation for understanding what type of situation that has emerged, although procedural support is available also for state recognition. In manual mode, the understanding of what has gone wrong in an anomalous event is facilitated by the clear relationship between previous actions and process response. The operator has been able to follow the process closely prior to the anomaly, with enough time to confirm that actions have been successful. When the series of successful actions is broken, it is easier to trouble shoot the problem compared to when using step- or automatic mode. However, the operators have to trade off the improved control to the relatively slow manual actions. Manual handling also requires intensive focus on the ongoing task which removes attention resources from the over all monitoring task.

Step Mode

In step mode, the operators use the ATS interface to start automatic sequences and follow the actions until the sequence has been fulfilled. The operator then evaluates that the desired process state is reached before the next sequence is initiated. During normal conditions this cycle is performed until the intended process state has been reached (i.e. either the generator is producing electricity to the grid or the turbine axis has come to a standstill).

The *current understanding* part in Figure 4 is affected by the turbine automation since feedback is reduced when the automation performs the actions. The operator is thereby distanced from the process. The ATS interface only shows whether the sequence's sub-steps have been successful or not, without any information on how the automation actually performs its actions (i.e. how the program logic conditions control the actions). If the operators can not easily see how the automation functions, understanding will be affected. In step mode, the operator can remain in the attended tactical control mode since he/she can pause to interpret, check and evaluate before initiating the next sequence. The available time to evaluate then compensates for the poor visibility of the program logic.

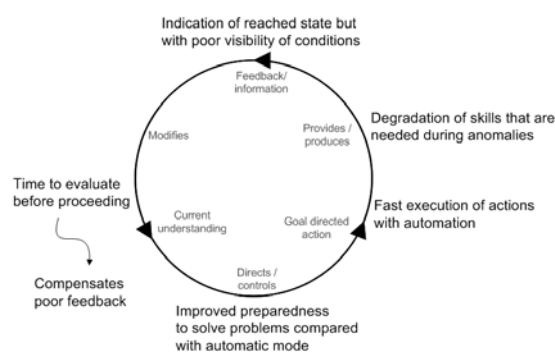


Figure 4. Step mode.

When an anomaly occurs, the ATS interface indicates at what step in the sequence the automation has stopped. The operator then has to consult paper based logic schemes to identify the cause of failure. Due to the complexity of the program logic, the problem identification can be very time consuming and delay the restart of the automation. When the ATS is operated manually, the operators' knowledge regarding how the automatic program sequences function is maintained. However, the use of automation degrades this knowledge – knowledge that is needed especially during extraordinary events.

Automatic Mode

In automatic mode, the turbine operator initiates the ATS program sequences and the program is executed without operator interference. It stops when it has finished its tasks, or if it encounters sequence conditions that are unfulfilled or when program orders fail to execute. Feedback on what sequences that are accomplished is continuously displayed in the ATS-interface, similar to the presentation in step mode.



Figure 5. Automatic mode.

The use of procedures is a cornerstone of nuclear power plant operations. Although actions are performed automatically, operators read the procedures to keep a high awareness of what is about to happen. Since the ATS processes information and executes actions much faster than human operators, it is very difficult to keep up to date with procedures in automatic mode. The operators feel that the automation is *running away* from them, which creates reluctance of using the automatic mode.

In the COCOM circle, the avoidance can be explained by the lack of time to evaluate each action that the ATS performs (Figure 5). Without time to check and think, the operator's role is changed from active processing of information to passive monitoring which can create out-of-the loop performance problems. Similar to the step mode, poor visibility also makes it difficult to recover from failures leading to degradation of manual skills due to lack of practice.

DISCUSSION

The use of COCOM as a framework for analysis proved to be useful as a tool for analysing supervisory work. The model provided explanations as to why operators stayed in, or lost control using different levels of automation by pointing to the important time aspects. In automatic mode, the lack of time for evaluation between feedback from the previous action and the initiation of the next action compromised efficient use of the ATS. COCOM also directs the search for solutions since solutions need to address the availability of time.

The aims of this study were to investigate how the ATS design support the turbine operators in their work in terms of monitoring and controlling the turbine process, and to identify possible improvements in the ATS user interface design. To improve control when using the ATS, the time for evaluation is the most critical aspect to consider. To increase the time available for evaluation of the automatic actions there are seemingly two ways to proceed. The first obvious measure is to stop the automatic sequence to give the turbine operator time to evaluate the previous actions, as applied in the step mode. The advantage of this approach is that it allows the operator to pace his/hers own work. However, the problem with poor visibility of sequence conditions remains and neither is the ATS utilised at its full potential. Another measure is to improve the ATS interface by increasing the observability of the automatic actions. Showing how the sequence conditions affect objects in the process together with information on what is needed to fulfil a condition would improve the operator's ability to quickly initiate trouble-shooting activity. A timeline that presents the past, present and future ATS activities would also strengthen the operator's awareness of the process state. By connecting the timeline to classical process schemes, awareness of what is happening during automatic actions would be further improved.

In a screen based automation interface, the use of a timeline together with an improved presentation of the underlying program logic could be combined to create an integrated automation interface that captures the advantages of manual actions and reduces the drawbacks of the opaque ATS interface. Using process schemes to indicate where automatic sequences are performing actions will enable the operators' ability to recognize object functionality that will facilitate prediction of future actions thus reducing the feeling of the automation running away. However, displaying objects that are physically distant but active in the same sequence requires skilful interface design to avoid an excessive number of new screen images.

One limitation to the study was the number of participating operators and their varying degree of experience. The identified problems were, however not present only among the inexperienced operators but within the experienced group too. Also, the operators with the longest experience tended to be the most careful when using the ATS equipment, probably because of experienced mishaps. In future studies, the number of participants should be increased to improve validity, and it would be preferable to observe each mode of operation individually in combination with interviews on previous experience. Nevertheless, the operator comments on their previous experiences with the ATS gave surprisingly rich data.

CONCLUSIONS

The results of the interviews regarding the turbine automation interface point out problems with out-of-the-loop, loss of skills and trust in relation to the use of the ATS. The use of COCOM as a means for analysis provided explanations of the results found through the interviews. Especially, time for evaluation was shown to be very important to the operators' ability to maintain control in all three levels of automation used in the turbine operations. To support control and address the automation related issues in future designs will be a challenge since complex interdependencies have to be presented. Increasing the observability of conditions and underlying program logic would be one measure to reduce the problems identified. Relevant information concerning the ATS status has to be provided together with the history of the automation as well as future actions. This has the potential to improve how the turbine automation user interface supports the operators' ability to take over control when shifting from automatic to manual mode in case of a failure.

ACKNOWLEDGEMENTS

Thanks are due to the turbine operators participating in the study and the KSU instructors Krister Walldem and Ronny Andersson for their support. This work has been financed by the Nordic Nuclear Safety Research (NKS).

REFERENCES

- [1] Sheridan, T. B. Humans and automation: system design and research issues. HFES issues in human factors and ergonomics; 3, 2002, Santa Monica, California: Wiley in cooperation with the Human Factors and Ergonomics Society (HFES).
- [2] Bye, A., Hollnagel, E., Hoffmann, M. and Miberg, A. B. Analysing automation degree and information exchange in joint cognitive systems: FAME, an analytical approach. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. 1999.
- [3] Neisser, U. Cognition and reality: principles and implications of cognitive psychology. A series of books in psychology, 99-0108656-6. 1976, San Francisco: Freeman.
- [4] Hollnagel, E. and Woods, D. D. Joint cognitive systems: foundations of cognitive systems engineering. 2005, Boca Raton, CRC Press.
- [5] Lee, J. D. Human Factors and Ergonomics in Automation Design. In: Salvendy, G. (Ed.). Handbook of Human Factors and Ergonomics, 2006, John Wiley & Sons, Inc.
- [6] Parasuraman, R. and Riley, V. Humans and automation: Use, misuse, disuse, abuse. Human Factors, 1997, 39(2), 230-253.
- [7] Lee, J. D. and See, K. A. Trust in automation: Designing for appropriate reliance. Human Factors, 2004, 46(1), 50-80.
- [8] Mumaw, R. J., Roth, E. M., Vicente, K. J. and Burns, C. M. There is more to monitoring a nuclear power plant than meets the eye. Human Factors, 2000, 42(1), 36-55.

Use Errors and Usability problems in Relation to Automation Levels of Medical Devices

Lars-Ola Bligård
Design and Human Factors
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
lars-ola.bligard@chalmers.se

Jonas Andersson
Design and Human Factors
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
jonas.andersson@chalmers.se

ABSTRACT

This paper is a reflection on the role of levels of automation in medical device usability. The reflection is based on a comparison of results from two independent usability evaluations on dialysis machines (one newer and one older). Usability problems were found that could be traced back to the use interface design and automation level of the machines. The comparison showed that there was a difference in usability problem types at different levels of automation and a conclusion was that the usability aspect becomes even more important as the level of automation increases.

Keywords

usability problems, use errors, automation

ACM Classification Keywords

H.5.3 [User-centered design]; K.4.3 [Automation].

INTRODUCTION

In today's healthcare, the use of advanced and complex medical devices that offer better treatment and more functions is increasing. The importance of appropriate usability of medical devices to ensure safety has been emphasized in many publications [1–5]. The level of automation is also increasing and the devices now need less direct user involvement in the treatment, both from medical staff and patients.

In other domains that utilize complex and advanced technology with a high level of automation, problems with automation have been acknowledged for a long time. For example, the use of adaptive cruise control in modern vehicles can cause automation surprises as the vehicle increases its speed when turning off a highway. This happens since the car no longer senses a vehicle in front of it, but is unable to define the situation as an inappropriate time to accelerate [6]. In process control, the use of automation can cause the operators to lose their manual skills since the skills can be left unpracticed for a long time. Difficulties of knowing what the automatic controllers are actually doing are also common in this domain, leaving the operators outside of the control loop. This compromises

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

effective manual take-over in case of a failure in the automatic system [7]. In the aviation domain, numerous examples have been described where automation related problems have led to severe accidents with many casualties [8]. Although medical devices are unlikely to kill as many people at once as an airplane crash, the severity of mistakes can have fatal outcomes – outcomes to which the design and use of automation is a potential contributor.

This paper is a reflection on the role of levels of automation in medical device usability. When the results from two independent analytical usability evaluations on dialysis machines were compared, interesting indications regarding the relation between level of usability and level of automation were found.

THEORY

Usability Problems and Use Errors

Usability of a device describes how easy it is to understand how the device works and how to make it perform its' intended tasks [9]. Characteristics that make this difficult are described as a usability problem. Consequently, a usability problem is any aspect of the design that is expected, or observed, to cause user problems with respect to some relevant usability measure (e.g. learnability, performance, error rate, subjective satisfaction) and that can be attributed to the design of the device [10].

As a result, usability problems in a device can increase the occurrence of use errors [11]. A use error is defined according to IEC (p 17) as an "*act or omission of an act that has a different result than intended by the manufacturer or expected by the operator*". For medical equipment the prevention of use error is important since such errors can result in injury to the patient [12].

In a usability evaluation of medical devices, where safety is the main objective, both usability problems and use errors are often investigated to achieve a more comprehensive analysis of the interaction.

Automation and Automation Problems

The level of automation in a device or technical system is increased when tasks, both manual and cognitive, are transferred from the human user to the device. This transfer is made to increase efficiency and quality of the task/process and to make the task/process less restricted by human limitations. Since automation makes tasks easier to perform, the automation can be argued to increase the device usability.

However, an increase of the level of automation also has negative implications for the human's role in socio-technical systems. Automation changes work from manual tasks being actively performed to passive monitoring of the automation's

actions. This can create a distance to the controlled process, irrespective of whether it is an operator monitoring a nuclear power plant or a nurse taking care of a patient. A physical distance results since automation offers possibilities to centralize the monitoring work, where several units can be supervised at the same time. This can be positive since resources are freed and can be used more efficiently. On the other hand, the rich feedback that comes with physical presence is lost and the feedback becomes limited to the capability of the sensors of the automatic system and to how the information from the sensors is presented to the operator or nurse.

Furthermore, the human being is not well suited to perform monitoring work. The vigilance of humans performing close monitoring has been shown to be significantly degraded after about 30 minutes [13]. Also, the amount of information that has to be taken into regard for safe operation of a complex technical system often exceeds the human information processing capacity.

METHOD

Usability Evaluation

The analytical usability evaluations were made on two types of dialysis machines. One old hemodialysis machine and one newer acute dialysis machine were compared in the study. A hemodialysis machine is used in dialysis wards on patients with chronic kidney deficiency, and an acute dialysis machine is used in intensive care units on patients with temporary kidney deficiency.

The main aim of the evaluation was to detect and identify presumptive usability problems and use errors. This knowledge was then used to increase the safety of forthcoming dialysis machines. The methods used were Generic Task Specification (GTS)[14], Enhanced Cognitive Walkthrough (ECW)[15] and Predictive Use Error Analysis (PUEA)[16]. The two evaluations were performed independently with approximately a year in between.

In both of the analytical evaluations, the user played by the evaluator was a nurse (intensive care and dialysis nurse) with adequate knowledge in treatment and use of dialysis machines. However, in the evaluation it was assumed that the users had never used the specific dialysis machine before. In the evaluations, the identification of presumptive problems and errors was made in relation to the intended use and intended user. This information was gathered by user interviews in a preparation step of the evaluation. For example, the difference in domain knowledge of the users for the two dialysis machines was considered in the evaluation. The intended users of the hemodialysis machine had more knowledge about dialysis treatment than the users of the acute dialysis machine. Both the evaluations were performed with an intended user as part of the evaluation team.

Qualitative Comparison

Some time after the second usability evaluation was performed; the authors discussed the relation between automation and usability. In the discussion, the dialysis machines came up as an example and this triggered a qualitative comparison of the two usability evaluations.

To begin with, the differences in the design of user interfaces were investigated. Secondly, the main usability problems and use errors were elicited and compared. Thirdly, the differences in the use and user of the machines were considered. Finally, these three aspects were analysed in relation to the level of automation.

RESULT

Differences between the Machines

In relation to the use and user, the old machine had fewer functions and a lower level of automation since tasks were handled manually and less help was given to the users via the interface. The new machine had more functions and a relatively higher level of automation since more tasks were executed automatically and help regarding the operating procedure was provided through the interface.

The design of the new dialysis machine had thus implemented two main principles. The first, to transfer knowledge in the user's head to knowledge in the world [17]. The second, to transform manual tasks to tasks performed by the machine with less user involvement. These two principles ought to increase the usability by demanding less knowledge and skill of the user.

Usability Problems and Use Errors

The identified usability problems in the old dialysis machine were foremost due to the need of device specific knowledge to operate it, because of hidden functionality and lack of feedback from the user interface. In the new dialysis machine, the main causes of usability problems were the text and symbols used in the display and on the device. These did not guide the user sufficiently. A usability problem found in both machines was inappropriate alarm messages. Due to insufficient information in the message, the users found it difficult to handle the alarm situation. The main presumptive use errors for both machines were action errors when connecting tubing lines and fluid containers.

Use and User

As described earlier, there is a large difference in the user and use of the two dialysis machines. Although this was considered and compensated for when comparing the results, the users of the old hemodialysis machine need a relatively greater amount of knowledge to use the machine effectively compared to the users of the new acute dialysis machine.

ANALYSIS

User Knowledge

The first obvious result of this comparison is that it was easier and faster for novice users to use the new machine, since it was very self-instructive. These users never had to learn and understand how the machine actually functioned. In comparison, the users of the old machine continuously learned new things when using the machine since the machine was not self-instructive and the users were forced to learn by doing, asking colleagues or consulting the manual.

Usability Problems and Use Errors

The characteristics of the usability problems found reflect the design of the user interfaces. In the old machine, many problems arise from the machine demanding specific knowledge to be used, while the new machine had its problem in guiding text and markings. The higher level of automation seems to have changed the character of the usability problems. In the new machine, the problems were related to the information given from the machine. An interesting aspect was that the character of the use errors had not changed in the same way. This is probably due to that the main manual tasks of the machines (connecting tubing lines and setting treatment), did not change their level of automation. The most common use errors in both

machines were the connection of connecting tubing lines and setting treatment.

Extraordinary Use

As mentioned above, both machines had usability problems concerning insufficient alarm messages. When considering the user knowledge needed to operate each machine, a difference between the machines emerged in the extraordinary operating condition. When the device malfunctioned and caused an alarm, it was easier to cope with the old machine since the users had extensive knowledge of the machine's functionality and the treatment. The new machine was instructive when everything was working normally, but during a malfunction, the information given from the device was insufficient. Further, the users of the new machine did not have the extensive knowledge of the functionality and the treatment that was needed to use the old machine.

Summary

To summarize, the analysis showed a tendency that the higher level of automation in the new acute dialysis machine (compared to the old hemodialysis machine), made it easier to use in the ordinary case, but harder to use in the extraordinary case.

This occurred since the users working with the acute dialysis machine usually had less knowledge of the functionality and the treatment. This is an indication that the users of the new machine are more dependent on the usability since they do not acquire in-depth knowledge while using the machine. The new machine is therefore more vulnerable to usability problems and use errors, since the users are dependent on the information provided by the machine's interface.

DISCUSSION

The comparisons raised some interesting issues. The effort made by the designers of the new acute dialysis machine to make it easier to use in ordinary tasks by using automation had some negative aspects. It made it harder for the users to understand whether the device was working normally or not, and it also made it more difficult to use in extraordinary cases. In contrast, the old hemodialysis machine required much more knowledge from the users, but they were better prepared to monitor the treatment and to handle any extraordinary case. This is interesting since the effort to achieve high usability in the common task, by applying automation, decreased the usability for the out-of-the-ordinary tasks that neither the user nor the machine were prepared for.

The comparison of the old hemodialysis machine and the new acute dialysis machine show how automation can prevent knowledge acquisition. This knowledge is needed in extraordinary situations where in-depth understanding of machine functionality is required to predict the outcome of an intended action. This problem has similarities with the loss of skills and knowledge that is an effect of the user being "out-of-the-loop" [18]. In this case, the knowledge is however not lost in the sense that the user has acquired it and then lost it due to lack of practice. Instead, the automation prevents the learning of how to handle extraordinary events from taking place at all.

This could be denoted as a type of Bainbridge's classical "Ironies of automation" [19] since the designers' efforts to improve ease of use by providing good usability and automation, actually undermine the user's ability to handle extraordinary events that fall outside of the automation's capability. In many domains the problematic effects of automation has been acknowledged for a long time and measures are taken to avoid

negative outcomes. For example, in the aviation domain, many companies encourage their pilots to fly manually from time to time to avoid loss of manual skills. In many process industries, fully automatic control is avoided because of the operators' feeling of control being lost.

The findings also resemble the *left-over* principle of function allocation [20], where the easy tasks are automated and the hard tasks that could not be automated are left to the human user to deal with. The rationale for using this principle is that automation can perform tasks faster, more reliably and precisely than humans can. However, this principle is only preferable if all conditions and situations that the technical system can encounter can be accurately predicted. This is rarely the case in human-machine systems that act jointly. Instead, the left-over principle often leads to deskilled users that have to deal with difficult situations. Here, the nurses using the new acute dialysis machine experienced a hard time performing adequately during extraordinary situations due to the normal tasks being automated.

The comparison shows how usability problems relate to and worsen classical automation problems. However, when automation problems such as out of the loop symptoms and loss of skills are discussed in literature, usability is seldom mentioned. In all interaction with machines, usability is important to ensure understanding and ease of use, irrespective of domain. In the case of usability of automatic systems, such as the new acute dialysis machine, the usability perspective becomes even more important and there is a need to incorporate the effects of automation in a longer time perspective. The user needs to understand both the user interface of the machine *and* the functionality of the automation to be able to follow what the autonomous functions are doing.

If the user cannot understand the information presented through the interface the probability of committing use errors increases since the user may be acting upon incorrect beliefs. If the user does not understand the autonomous activities behind the interface, the probability of out-of-the-loop problems will increase accordingly. Therefore, for the ordinary case it is important that the machine present information that helps the user to stay in the loop and enable the users to gain knowledge and prevent loss of skill.

For the users of the dialysis machines to be able to cope with the extraordinary, it is essential that users during ordinary use receive information about the functionality and the treatment. This information enables them to gain the in-depth understanding needed to handle unexpected events. Furthermore, it is in the extraordinary case that the user is in the greatest need of good usability. The designer therefore has to anticipate what information is needed to help the user, and how it should be presented in an appropriate way.

As the level of automation increases in medical devices, the usability aspect becomes even more important since the probability of automation-related effects such as out-of-the-loop problems increases with the introduction of automation. Use errors can also become more severe at a high level of automation since an accurate understanding of the device and safe handling go hand in hand.

REFERENCES

- [1] Ward, J. R. and Clarkson, P. J. An analysis of medical device-related errors: Prevalence and possible solutions. *Journal of Medical Engineering and Technology*, 2004, 28(1), 2–21.

- [2] Kaye, R. and Crowley, J. Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management. 2000, Rockville: US Food and Drug Administration.
- [3] Crowley, J. J. and Kaye, R. D. Identifying and understanding medical device use errors. *Journal of Clinical Engineering*, 2002, 27(3), 188–193.
- [4] Obradovich, J. H. and D. D. Woods, Users as designers: How people cope with poor HCI design in computer-based medical devices. *Human Factors*, 1996, 38(4), 574–592.
- [5] Lin, L. et al. Applying human factors to the design of medical equipment: Patient-controlled analgesia. *Journal of Clinical Monitoring and Computing*, 1998, 14(4), 253–263.
- [6] Norman, D. A. *The design of future things*. 2007, New York, NY: Basic Books.
- [7] Endsley, M. R. and Kiris, E. O. The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 1995, 37(2), 381–394.
- [8] Billings, C. E. *Aviation Automation: Principles and Guidelines*. 1996, NASA Ames Research Center: Moffett Field, California.
- [9] Norman, D. A. *Emotional design: why we love (or hate) everyday things*. 2004, New York: Basic Books.
- [10] Nielsen, J. *Usability engineering*. 1993, Boston: Academic Press.
- [11] Bligård, L.-O. *Prediction of Medical Device Usability Problems and Use Errors – An Improved Analytical Methodical Approach*. 2007, Chalmers University of Technology: Göteborg.
- [12] IEC, IEC 60601-1-6:2004 Medical electrical equipment – Part 1–6: General requirements for safety – Collateral standard: Usability 2004, Geneva: IEC.
- [13] Parasuraman, R. Human-computer monitoring. *Human Factors*, 1987, 29(6), 695–706.
- [14] Bligård, L.-O. and Osvalder, A.-L. *Generic Task Specification – A Framework for Describing Task Demands and Mental/Physical Work loads in a Human-Machine System*. In: 2nd International Applied Human Factors and Ergonomics 2008. 2008: Las Vegas.
- [15] Bligård, L.-O. and Osvalder, A.-L. *Using Enhanced Cognitive Walkthrough as a Usability Evaluation Method for Medical Equipment*. In: IEA. 2006. Maastricht, Netherlands
- [16] Bligård, L.-O. and Osvalder, A.-L. *An Analytical Approach for Predicting and Identifying Use Error and Usability Problem*. *Lecture Notes in Computer Science*, 2007. 4799. Pp. 427–440.
- [17] Norman, D. A. *The Design of Everyday Things*. 2002, New York: Basic Books Inc.
- [18] Lee, J. D. *Human Factors and Ergonomics in Automation Design*. In: *Handbook of Human Factors and Ergonomics*, Salvendy, G. (Ed.). 2006, John Wiley & Sons, Inc.
- [19] Bainbridge, L. Ironies of automation. *Automatica*, 1983, 19(6), 775–779.
- [20] Hollnagel, E. and Woods, D. D. *Joint cognitive systems: foundations of cognitive systems engineering*. 2005, Boca Raton, FL: CRC Press.
- [21] Lee, J. D. and K. A. See, Trust in automation: Designing for appropriate reliance. *Human Factors*, 2004, 46(1), 50.

Mobile Air Ticket Booking

Ivan Burmistrov

Moscow State University, Department of Psychology

11-5 Mokhovaya

Moscow 125009 Russia

ivan@interux.com

ABSTRACT

Online air ticket booking is a cognitively complex task even on fully-functional internet-access devices such as desktops, representing a repetitive multi-parametric search in the flights database and then browsing long lists of flights found, consisting of different carriers, prices, dates and times, to create an optimal combination of outbound and inbound flights. We present the results of research into prospective users of mobile air ticketing, a set of domain-specific user interface design guidelines, and a wireframe design for mobile air ticket booking application.

Keywords

M-commerce, mobile usability, air ticket booking, guidelines

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation] User Interfaces – *User-centered design*; K.4.4 [Computing Milieux] Computers and Society – *Electronic commerce*.

INTRODUCTION

Travel is by its very nature mobile, and “going mobile” is currently considered by many travel specialists as a central part of travel agents’ strategy for growth, even a do-or-die decision for travel industry [1]. Although proliferation of mobile applications in air travel is not so fast as expected by service providers and mass media, a number of airlines currently provide mobile solutions for such activities in the airline travel activity chain as checking flight status, flight schedules, mobile check-in, mobile boarding pass, seat availability and seat selection, and making changes for flights. However, applications for mobile air ticket booking are relatively rare as there are serious barriers to mobile airline ticket purchasing. First of all, the very nature of air ticket booking task makes its mobile implementation highly questionable.

Online air ticket booking is a cognitively complex task even on fully-functional internet-access devices such as desktops and laptops, representing a repetitive multi-parametric search in the flights database and then browsing long lists of flights found, consisting of different carriers, prices, dates and times, to create an optimal combination of outbound and inbound flights.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Performing a booking task may well take tens of minutes on a full-fledged PC, it is taxing on attention and mental load, and requires considerable text input (especially when entering passenger data and payment details). These characteristics of the task themselves make their implementation barely suitable for a mobile device because mobile use contexts are not tolerant to long cognitively-loaded interactions and massive text input.

In this practically oriented paper we describe our experience in developing mobile user interface for this highly challenging task, the analysis of target user characteristics, provide a list of domain-specific guidelines for mobile air ticket booking applications, and present a wireframe design for mobile air ticket booking system.

THE PROJECT

The development of m-ticketing system for airline travel became a joint effort between one of the national largest mobile operators and popular online travel agency *Bilet Online*. The system was planned as not a standalone application but to become a component of a bundle of various mobile applications being marketed by the mobile operator to its subscribers.

The parties were aware that usability engineering would be the key to their system’s success and hired specialists from the Laboratory of Work Psychology, Moscow State University that had rich experience in online travel usability and conducted long-term research on multitasking and interruptions in human-computer interaction, a field directly relevant to mobile usability. The expected deliverables from our participation in the project were: (I) the user interface guidelines for mobile air ticket booking, (II) a wireframe design of the user interface, and (III) usability testing of a working prototype implemented on a mobile phone. Below we present the results of stages (I) and (II)⁶.

Target Users

From the very beginning the parties realized that a mobile air ticket booking system has no chances to become a mainstream, mass-market product to be used by everyone, so studios efforts had been made to define the target group of prospective users of the air m-ticketing system.

Air travelers are generally classified into two groups – business and leisure travelers – with main difference between them that business travelers being primarily concerned with the exact date and time of travel and are less concerned with its costs while leisure travelers usually seek for the cheapest flights and are more flexible with dates. Other differences between these two groups include frequency of travel and therefore air travel experience, travelling mainly alone or in a group, and predictability of the moment and destination of the next travel.

⁶ Unfortunately, by the time of writing this article we were not able to test our design with users in real use situations because working prototypes of the software were not yet developed by programmers.

Market research conducted by the travel agent's marketing department on their existing client base showed that in normal situations leisure travelers more than likely will use non-mobile channels of purchasing tickets (either online at an e-commerce website, via a phone call or alternatively at a physical travel outlet) and would almost never use a mobile phone (especially in the case of family vacation planning). This finding is in concordance with opinion expressed by travel market specialists that "while the capability obviously exists, there is not yet a universal demand for using a mobile to book the next family holiday" [1]. Alternatively, active business travelers seemed to be potentially more grateful users of the m-ticketing technology, for example, in situations of rush travel. However, business travelers noted as well that mobile alternative will only be used when PC internet is unavailable. A survey of passengers of two airlines in South Africa [11] showed that (1) respondents perceive more value in *receiving information* via their mobile devices than they do in *making bookings*, and (2) business travelers exhibited significantly more "mobile readiness" than leisure travelers.

The resulting prospective customer profile that guided further development of the guidelines and interface design looked as follows: (a) frequent business traveler, the moment of next travel sometimes unpredictable, (b) experienced in air travel and familiar with its peculiarities, (c) mainly travelling between limited number of familiar airports, (d) travelling mainly alone than in group, (e) high degree of technology self-efficacy individuals.

Mobile Usability

The most recent study of mobile usability published by Jakob Nielsen in July 2009 *Alertbox* [13] showed that mobile web user experience is still miserable, main usability problems remaining the same as ten years ago: small screens making almost all interactions hard; awkward and error-prone input, especially for typing; download delays, even with 3G service; and mis-designed sites not optimized for mobile access. Website use on mobile devices, even on touch phones that offer full-featured browsers does not offer PC-level usability [13]. In contrast to web-based mobile applications, device-based mobile applications utilizing client side processing and location context are able to achieve objective performance and subjective usability measures comparable to those of the PC-based versions, despite the limited input and display capabilities of the mobile device [15].

The decision not to use WAP and instead build on a device-based rich media platform such as Java was not ours, but we appreciated it a lot due to a number of reasons. In particular, device-based mobile applications provide sophisticated interaction styles beyond the simple navigation model of web based applications. They also offer a more immediate experience since they are not so heavily bound by request/response cycles inherent in web based design [15]. Furthermore, device-based applications also give more opportunities for visual design aesthetics having significant impact on perceived usefulness, ease of use, and enjoyment, which ultimately influences users' loyalty intentions towards a mobile service [5].

Existing Applications

Our search for existing air ticket booking solutions revealed that a number of airlines offer WAP-based booking services, but it yielded only a few rich media solutions. All but one of these solutions were solutions for a single air company and not included search requests to global distribution systems (GDS) accumulating hundreds of air carriers, the case we dealt with.

The only GDS-powered solution was in fact 1:1 replica of a corresponding PC website transported into mobile device and obviously represented an inappropriate design decision to a mobile task. Among existing systems, only one has been designed by the usability specialists [8], so our work combining rich media, GDS access and usability engineering looked as a pioneering one.

GUIDELINES

The guidelines below adapt general recommendations for mobile applications [6, 7, 9, 10] and recommendations for air travel websites [4, 17, 18]. These guidelines are oriented to regular mobile phones that account for the vast majority of the market: devices with a tiny low-resolution screen, a numeric keypad, a joystick or four-directional buttons, two soft-buttons below the screen, and the ability to run Java applications. Due to space limitations we mention only a subset of most general guidelines we developed.

Define the Target User

Mobile applications strongly require a clear understanding of the motivations and circumstances surrounding mobile device use and adoption from the perspective of the consumers [16]. In particular, culture is an important determinant of mobile device use and adoption since different cultures developed different traditions for mobile services use. For example, in Japan on *All Nippon Airways*, 5% of all domestic flights are booked on mobile phones [2], and this high percentage is unbeaten anywhere in the world. The explanation of this fact is because of the length of their commute to work, people in Japan use a mobile to surf the web, making more surfing on mobiles than on PCs. Another interesting motivation for mobile device use, in the case of a major city in Thailand, was the fact that people are often stuck in their car due to frequent traffic jams [16].

This means that investigation into characteristics of prospective users, contexts of use and technology adoption factors must be a starting point in developing mobile applications for such a complex and specific task as airline m-ticketing.

Make Mobile Application a Supplement to a Website

There are serious reasons to implement and market mobile version as not a standalone application but a satellite to the "main" travel agent's website. Firstly, this will avoid heavy text input aspects of the air ticket booking task because the mobile application can use data from the user's profile (passenger and payment information) entered via website and stored on the central server. Secondly, this will temper the user fears about safety of mobile transactions since no sensitive information will be transferred through mobile channels. Thirdly, close integration with the website will allow users to make urgent changes and cancel flights booked via the website thus seriously increasing the usefulness of mobile application to the user. Fourthly, the history of previous travel and user preferences can be borrowed from the central server to mobile in order to pre-fill the fields in the mobile interface with smart defaults thus minimizing text entry.

Reduce Functionality to an Absolute Necessary Minimum

The standard set of flight search parameters on a travel agent websites includes: (a) roundtrip/one-way flights, (b) from/to

destinations, (c) departure/return dates, (d) preferred departure/return time of day, (e) number of adults, children and infants, (f) economy/business class, (g) flexible dates, (h) direct flights only, and (i) preferred airlines. For a mobile application, we recommend to reduce the search options to (a), (b) and (c) only.

When displaying the flight search results page, it is recommended to avoid cluttering the screen with redundant information except (1) date and time, (2) price, (3) number of stops for transfer flights, (4) next day arrival, and (5) operating carrier (the airline that uses its aircraft for the flight). (More detailed information on the selected flight may be presented on the flight summary screen.)

Provide Support for Multitasking and Interruptions

The nature of mobile computing requires user interaction design to pay special attention to multitasking and interruptions [11]. Mobile contexts are typically public and dynamic rather than private and stable, and mobile users must permanently switch back and forth between the mobile tasks and external sources, temporarily leaving the switched-from tasks on hold or slowing them down [14]. Tasks with interruptions take longer to complete on a mobile device compared to a desktop computer, due to a smaller screen, limited input interaction and high demands on attention [11].

A semi-naturalistic field study of users performing mobile web tasks while moving through typical urban situations [14] demonstrated the impulsive, fragmented, and drastically short-term nature of attention in mobility. Continuous attention to the mobile device fragmented and broke down to bursts of just 4 to 8 seconds, and attention to the mobile device had to be interrupted by glancing the environment up to 8 times during waiting a web page to be loaded.

Our earlier research [3] revealed that re-orientation in the main task after attention switch-away is mainly responsible for performance degradation in interrupted tasks, and this case requires the development of less attention-demanding user interfaces and support for quick resumption when switching back to the mobile task.

Recommendations for the support of task switching include: (1) breaking the interaction into small pieces – typically one operation per screen, and (2) providing an attention cues enhancing recognition to direct the user to a place in the suspended task (for example, a highlight may be presented around a text box as an attention indicator for a specific point in a task) [11].

Make Application Location Aware

Location awareness is a clear advantage of mobiles over desktops. Even without using GPS sensors, there are technical possibilities to detect the user's current location at least with geographical region precision and provide the user with relevant information and smart defaults to reduce text input.

DESIGN

Figures 1–3 present a wireframe design for the mobile user interface following the canonic scheme of online air ticket booking process: flight search (a–g in Figure 1), search results

(h, j), flight summary and flight confirmation (i, k), booking and payment (l–n), purchase confirmation (o)⁷.

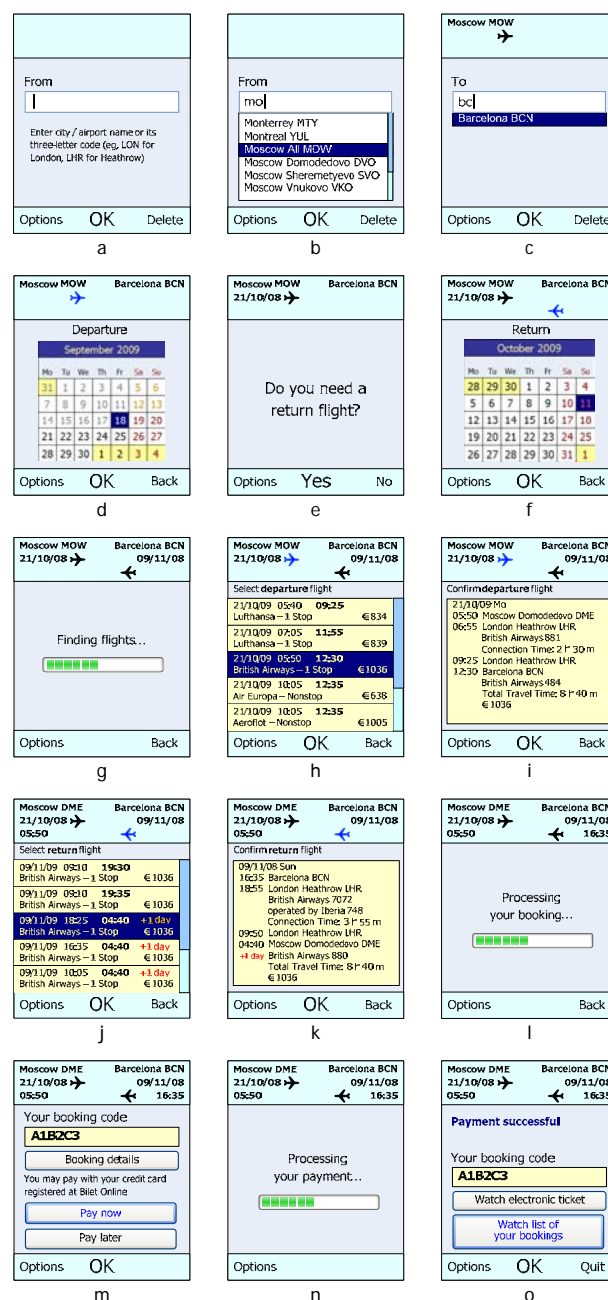


Figure 1. Booking process.

Figure 2 presents the screen area organization and Figure 3 shows the “slide-push” transition between consecutive screens.

⁷ One may notice that our design in many aspects resembles the design described in [8]. However, we came to a similar design independently, borrowing only one idea from [8] – a push-left/right screen transitions between pages. Similar problems, similar solutions.



Figure 2. Screen organization.



Figure 3. Push transition.

ACKNOWLEDGMENTS

This work was partially supported by the grant № 08-06-00284a from the Russian Foundation for Basic Research to the Laboratory of Work Psychology, Moscow State University.

REFERENCES

[1] Abacus 2008. Going mobile will soon be a do-or-die decision for travel industry. *TravelSmart-Asia Watch* 45 (Oct. – Nov. 2008), 1–6. http://www.abacus.com.sg/trvlsmartasia/trvlsmartwatch/Lists/News/Attachments/139/TSA_Eng_Issue9.pdf.

[2] Bicknell, D. 2007. The whole world in their hands. *Travolution* 13.0 (Dec. 2007), 22–26.

[3] Burmistrov, I. and Leonova, A. 2003. Do interrupted users work faster or slower? The micro-analysis of computerized text editing task. In: *Proceedings of HCI International 2003, Vol. 1*. Heraklion, Crete, Greece, Jun. 22–27, 2003. Lawrence Erlbaum Associates, Mahwah, NJ. Pp. 621–625.

[4] Chariton, C. and Choi, M. 2002. User interface guidelines for enhancing usability of airline travel agency e-commerce web sites. In: *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA, April 20–25, 2002). CHI'02. ACM, New York, NY. Pp. 676–677. DOI = <http://doi.acm.org/10.1145/506443.506541>.

[5] Cyr, D., Head, M. and Ivanov, A. 2006. Design aesthetics leading to m-loyalty in mobile commerce. *Inform. Manag.* 43(8), Dec. 2006, 950–963.

[6] Duda, S., Schiebl, M. and Hess, J. M. 2002. *Mobile Usability Report*. Eye square, Berlin. <http://eye-square.com/documents/Mobile-Usability-eyesquare-english.pdf>.

[7] Gong, J. and Tarasewich, P. 2004. Guidelines for handheld mobile device interface design. In: *Proceedings of the 2004 Decision Sciences Institute Annual Meeting* (Boston, Massachusetts, USA, Nov. 20–23, 2004). Decision Sciences Institute, Atlanta, GA. Pp. 3751–3756.

[8] Hjerde, M. 2007. Case Study and Discussion: Booking Airline Tickets from a Mobile Phone. *Sender 11: A Blog about Mobile Interaction Design*. May 09, 2007. http://sender11.typepad.com/sender11/2007/05/case_study_and_.html

[9] Jones, M. and Marsden, G. 2006. *Mobile interaction design*. John Wiley & Sons Ltd.

[10] Lee, Y. E. and Benbasat, I. 2003. Interface design for mobile commerce. *Commun. ACM*, 46(12), Dec. 2003, 48–52. DOI = <http://doi.acm.org/10.1145/953460.953487>.

[11] Lubbe, B. and Louw, L. 2009. The perceived value of mobile devices to passengers across the airline travel activity chain. *J. Air Transport Manag.* DOI = <http://dx.doi.org/10.1016/j.jairtraman.2009.02.002>.

[12] Nagata, S. F. 2003. Multitasking and interruptions during mobile web tasks. In: *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. Denver, Colorado, USA, Oct. 13–17, 2003. Human Factors and Ergonomics Society, Santa Monica, CA. Pp. 1341–1345.

[13] Nielsen, J. 2009. *Mobile Usability*. Alertbox, July 20, 2009. <http://www.useit.com/alertbox/mobile-usability.html>.

[14] Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Portland, Oregon, USA, Apr. 02–07, 2005. CHI'05. ACM, New York, NY. Pp. 919–928. DOI = <http://doi.acm.org/10.1145/1054972.1055101>.

[15] Ryan, C. and Gonsalves, A. 2005. The effect of context and application type on mobile usability: an empirical study. In: *Proceedings of the Twenty-Eighth Australasian Conference on Computer Science – Vol. 38*, Newcastle, Australia. Australian Computer Society, Darlinghurst, Australia. Pp. 115–124.

[16] Sarker, S. and Wells, J. D. 2003. Understanding mobile handheld device use and adoption. *Commun. ACM*, 46(12), Dec. 2003, 35–40. DOI = <http://doi.acm.org/10.1145/953460.953484>.

[17] UsabilityOne 2007. *Website Customer Experience: Airline Websites*. UsabilityOne. <http://www.usabilityone.com/usability-resources/newsletters/may07/may07-2.php>.

[18] Webcredible 2009. *Flights Online – Ensuring Your Site Takes off: Online Travel Usability Report 2009*. Webcredible. <http://www.webcredible.co.uk/user-friendly-resources/white-papers/travel-usability-2009.pdf>.

Towards Co-creation of Service Research Projects – A Method for Learning in the Network

Minna Halonen
VTT Technical Research
Centre of Finland
P.O. Box 1000
FI-02044 VTT, Finland
minna.halonen@vtt.fi

Katri Kallio
VTT Technical Research
Centre of Finland
P.O. Box 1000
FI-02044 VTT, Finland
katri.kallio@vtt.fi

Eveliina Saari
VTT Technical Research
Centre of Finland
P.O. Box 1000
FI-02044 VTT, Finland
eveliina.saari@vtt.fi

ABSTRACT

How to combine research knowledge across disciplines is a challenge when studying and developing services in industry and public organizations. This paper presents a new kind of workshop process aiming at co-creation in a research network. We piloted the process at VTT Technical Research Centre of Finland during January – May 2009.

The originality of our approach is in combining the methods of foresight and developmental impact evaluation. Furthermore, the process builds up a collaborative network and its research strategy from below, from above and together with customers and collaborators. Our paper describes the pilot process phase by phase and the first results and experiences from the workshops.

Keywords

service research, co-creation, networking, workshops, foresight, organisational learning, roadmapping, developmental impact analysis, methods

INTRODUCTION

In large research organizations there is a tendency that new research projects originate in research groups or in knowledge silos. Applied research faces specific challenges of how to consider the needs of the customers, scientific knowledge and societally relevant questions in the research projects. Large organizations may also have a communication gap between managers and employees. We claim that collaboration across “the silos”, across hierarchical levels, disciplines and different actors, does not emerge easily. New interdisciplinary research networks require tools and methods for initiating learning, synergy and collaboration.

Services, both as a business and as a science, are a rapidly growing sector, and they have a remarkable influence on processes and operations in companies and other organisations. Although new service business opportunities are facilitated by ICT and other rapidly developing technologies, the development and implementation of services take place at a slower pace. In

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

order to innovate and develop successful services for global markets, we need future-oriented and multi-disciplinary approach, which combines technological knowledge to e.g. behavioral, social and design sciences.

Technical Research Centre of Finland (VTT) has traditionally focused on the development and application of technology. One of its current challenges is to create a new line of research and development in the field of service business and service innovation research. This means developing new kind of expertise, and also combining knowledge in new ways. We had the opportunity to facilitate service science and business network through its first steps. In this paper, we describe the method we developed.

We combined foresight and organisational learning methods, namely roadmapping and developmental impact evaluation. During the workshops VTT researchers and the management were enabled to create a shared understanding of service research strategy at VTT. The workshops were designed to facilitate dialogue between the users of the research, potential collaborators such as universities, the funding agencies and societal actors in the field of service science.



Figure 1. Some of the participants at the end of the last workshop day.

We developed a process called learning by foresighting and evaluating (LIFE). LIFE enables the management of future-oriented networking across organisational borders as a basis for continuous learning and innovation. The process is a potential embryo for a new kind of research culture towards learning in the network, shared and transparent planning of project proposals.

We, as authors of this paper have complementary knowledge and expertise. The first author works currently in the field of technology foresight, managing and facilitating roadmapping projects. The second author has expertise on management and organizations, and has experience on developmental impact evaluation. The third author has studied the dynamics and development of research groups, and has conducted developmental impact evaluation process in two research communities. In the LIFE process all three of us were involved as facilitators and co-constructors of the workshops.

METHOD

The method consists of four phases:

- 1) Roadmapping the future challenges and directions of the service science research
- 2) Developmental impact evaluation of two previous research projects conducted by VTT, which gives an opportunity to question the prevailing research strategy and its strengths and weaknesses.
- 3) Conceptualizing the future vision and strategy of service science research with the help of lessons learnt from the previous projects and the horizon of the future challenges.
- 4) Co-creating the first project proposals and roadmaps according to the new strategy.

The workshops of the LIFE process were designed to help the participants to move forward in their zone of proximal development. (Engeström 1999). This term refers to a situation and terrain of constant ambivalence, struggle and surprise, when the participants are about to create the next actions for the future. In this terrain, the possibility of *expansive transformations* (Engeström 2001) or as we call *creative shifts* may take place. The workshops aimed at creating a learning situation, in which the participants were able to see their research in a wider perspective than before. This expanded horizon meant seeing research projects e.g. from the management's, from the customer's or from the research collaborator's point of view.

Although, the need for creating a research strategy in dialogue between different actors is often called for, it is rarely implemented and achieved in practice. Because learning and creativity flourish in the atmosphere of enthusiasm, our method pays special attention on how to create such inspiring workshops. Therefore metaphors and language of art were utilized to prepare the participants to co-create.

New bearing idea of the process was to bring people "outside of their boxes". This was done e.g. by using art metaphors when building working groups. The groups were named after different art schools: symbolists, cubists, futurists and so on. This measure turned out to be a good ice-breaker between researchers with different expertises. This new identity gave an inspiring and neutral starting point for the collaboration.

We cultivated art metaphors throughout the whole process with the exception of the last workshop which was held by the lake Siikajärvi (Lake Whitefish in English) in the national park of Nuuskio near Helsinki. At this point we switched the metaphor from art to fishing. The last workshop day was coloured by fishing metaphors starting from a picture of a local lake in the background of the day's programme and finishing with a fish soft toy which participants threw from one to another during the final feedback discussion.

SSB-network: 5. työpajan ohjelma	
26.5.2009 Siikaranta, Nuuskio	
9:30	Kurkistus verkon saaliseen Minna Halonen, Katri Kallio ja Evellina Saari, VTT
10:00	Palvelututkimuksen strategia ja palvelulaboratorio Jouko Myllyoja, VTT
10:15	Ryhmätyö: Mitä jäi haaviin, miten kalastetaan jatkossa?
11:15	Palveluinnovaatiot ja -muotoilu Katri Ojasalo, Laurea
12:00	Lounas
13:00	VTT:n haasteet palvelututkimuksessa Kari Neillimo, VTT
14:00	Palvelututkimuksen kalansaatiin perkaus 2:ssa ryhmässä (15 min/altio)
16:00	Pietarin kalansaalis?
17:00	Sano mulkkut!

Figure 2. The programme of the last workshop held by the lake Siikajärvi (Lake Whitefish in English).

The LIFE process is a potential embryo for a new kind of research culture towards learning in the network, shared and transparent planning of project proposals.

LEARNING BY FORESIGHTING AND EVALUATING (LIFE) PROCESS

Learning by foresighting and evaluating process derives from the theories and methods of expansive learning, Developmental Impact Evaluation and Roadmapping. Abbreviation LIFE characterizes also the vivid and interactive process between different stakeholders during the workshops. The process enhances new face-to-face contacts inside and across organizations and hierarchical positions. It creates new "life" for the research area, which is dispersed in the organization.

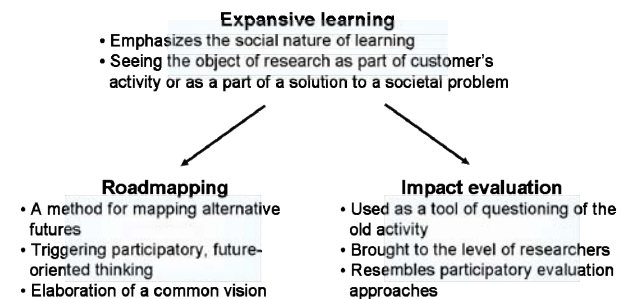


Figure 3. Learning by foresighting and evaluating -process.

Expansive learning emphasizes the social nature of learning. Learning is not taking place only inside individual's mind, but is embedded in the development of activity. It considers learning taking place between people and in the working environment, in its situations, actions, negotiations and using of material artefacts. Expansion refers to the phenomenon of exceeding the initially given context of specific problems and refocusing on the wider context that generates those problems. In the case of a research activity, this means perceiving the object of the research not merely as an opportunity to expand scientific knowledge, but as an integral part of the customer's activity or as a part of a solution to a societal problem.

One key aspect of expansive learning is that the old way of acting has to be questioned and it is used as a starting point for finding new solutions and forming a new kind of activity. In LIFE process, the analysis of two past research projects and

their impact evaluation is used as a tool for questioning. In our process we have brought the concept impact analyses to the level of researchers and research group. Traditionally the impact assessments have been made at the organisational level and reported mainly to the top management.

The other approach used in the LIFE process is based on foresight, or futures studies, which explores alternative futures. It is used to improve decision making with long-term consequences, offering a framework to better understand the present and expand mental horizons. Roadmapping presents one practical and largely used foresight method. Roadmapping is a method for mapping alternative futures. It links the future to the present and helps the elaboration of a vision of the future. It is a participatory process where process itself is often more important than the result i.e. a roadmap, which presents the graphical output of a roadmapping process. In our process we have combined roadmapping to organisational learning methods to foster organisational development and the creation of horizontal networks. In this context roadmapping has been utilised as a method for triggering participatory, future-oriented thinking within the LIFE process with less importance given to the roadmaps themselves.

THE PHASES OF LEARNING BY FORESIGHTING AND EVALUATING (LIFE) PROCESS

During the workshops circa 30 VTT researchers and the management representatives were gathered together to create a shared understanding of service research strategy at VTT and new multidisciplinary service research projects. In addition the workshops were designed to facilitate dialogue between VTT's service researchers and the "outside world": the users of the research, potential collaborators such as universities, funding agencies, and the societal actors in the field of service science. For every workshop we brought an outsider's view to inspire the future research directions and alliances of VTT.

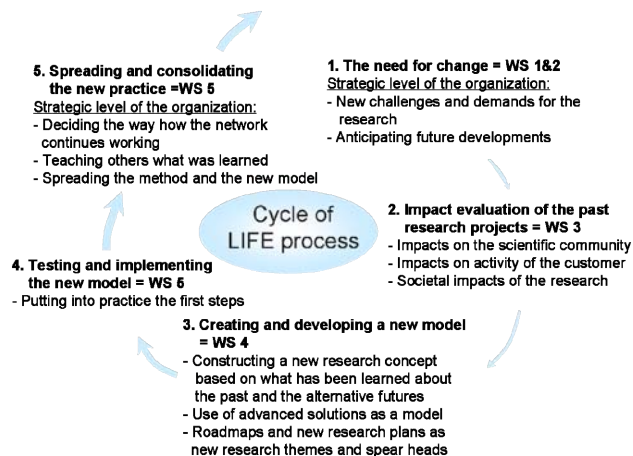


Figure 4. Cycle of LIFE process.

In the following, we open up the phases of learning by foresighting and evaluating step by step. The phases are introduced as they took place in the context of SSB network at VTT from January to May 2009. Methodological principles, such as how impact evaluation, organizational learning and foresighting complemented each other, are described in each phase. Figure 4 describes the method as a learning process or cycle.

The Need for Change – the First Two Workshops

In the *first workshop* the participants were gathered together to initiate the networking between them. The purpose of the workshop was to become acquainted with each other. The management proposed only one workshop session to create a service research network. However the participants realized that one workshop was not enough for creating either new insightful projects or long-term network to VTT.

We described the phases of the LIFE process as a draft to the participants and the management. We explained that this was the first effort also for us as facilitators to combine foresighting and developmental impact evaluation. However, we had conducted these methods and processes separately in different research communities with promising results (e.g. Saari et al. 2008, Ahlqvist et al. 2008). We called the participants for piloting this potential organizational innovation together with us. Our invitation to co-create was received with enthusiasm and we agreed on conducting four workshops during four subsequent months.

The purpose of the *second workshop* was to create a dialogue between managers' vision of service research and concerns of the researchers. The task of the researchers was to listen the message at the management level and then ponder and compare their own ideas about the future developments and concerns relating to its implementation to their daily work. The discussion was facilitated by "a fish bone" team exercise in order to make participants' viewpoints and concerns visible.



Figure 5. A group working on a roadmap during the second workshop.

After the discussion participants were divided in two groups, both of which developed a service research landscape for the next 15 years. Through this first roadmapping exercise we explored opportunities and identified challenges, drawing 'the big picture' of the service research in the future. The aim was to expand mental horizons and acquaint the participants with the method. In the end of the second workshop we chose together with the participants which research projects would represent the past service research at VTT. These were to be used as objects for developmental impact evaluation.

Impact Evaluation of the Past Research Projects – the Third Workshop

In the *third workshop*, two past service science research projects were evaluated from three complementary perspectives: scientific,

societal and the customer's. One of the projects' customers, research collaborators and a representative of the funding agency were invited to the workshop to evaluate the impacts of the projects from their perspective. The researchers observed and analysed the speeches with a help of a conceptual tool. The main purpose of the evaluation was to question the old way of conducting research. In the end of the workshop, the groups created new research themes based on what they had learned from the past (impact evaluation) and the future (roadmapping). These themes formed the basis for the next workshop.

Creating a New Model – the Fourth Workshop

In the *fourth workshop* the participants were supposed to form new project proposals and action plans. The roadmaps (service research landscapes) produced in the second workshop were utilised as a basis for the brainstorming session. In this second roadmapping phase we had smaller groups and more focused themes compared to the first roadmapping exercise. The aim was to scan deeper into the future of service research by refining the knowledge and organising, designing and planning the real life actions. In the end of the workshop each team presented their project embryos. They continued writing of the plans as a real research proposals.

Testing and Implementing the New Research Plans – the Fifth Workshop

In the *fifth workshop*, the new research proposals, which represented the spearheads of the strategic focus areas, were introduced. We invited well known service experts to analyze and spar the proposals in this early phase.

These new project proposals formed the service research strategy and took the first steps into practise. At this phase we stepped away as the facilitators of the network. However, it is important to decide how the network will continue acting in the future. Based on our previous intervention processes we have learned the importance of the middle managers role in continuing and spreading this kind of networked way of operating.

CONCLUSIONS

This process developed the service research network and produced new project initiatives as immediate results. New research collaborations were initiated spontaneously between the participants already during the process. At the end of the process both the researchers and the management expressed their interest to continue on co-creating service research projects and underlined the importance of "good facilitation" in

such an activity. Their high motivation was testified by active participation throughout the workshop process despite tight project schedules of the people involved.

However organizational learning is fragile; it takes time before a new vision constructed together becomes visible in actions. The LIFE process serves also as a management tool for planning and organising research and creating networks. In order to continue this way of learning in the networks, it needs to be adopted as a continuous way of planning and organizing research.

In addition, the LIFE process itself is a promising service concept, which we may market as a service for recently established research networks. The first pilot proved how fruitful it is to combine the analysis of the past to the construction of the future in the same process. By combining these methods we create an opportunity for people to learn and move between these horizons – from past to the future.

ACKNOWLEDGMENTS

We thank the participants of the workshop process for their enthusiasm and broadmindedness in launching into this new, challenging experience. We want to thank the management of VTT for the trust, also in the form of financial resources, given to us facilitators and the network. We are grateful for the highly qualified scholars and professionals who accepted our invitation to share with the network their insights on service research and did it with an exceptional sense of openness.

REFERENCES

- [1] Ahlqvist, T., Bäck, A., Halonen, M. and Heinonen, S. 2008 Social Media Roadmaps. Exploring the futures triggered by social media. VTT Tiedotteita – Research Notes 2454. VTT, Espoo. 78 p. + app. 1 p.
- [2] Engeström, Y. 2001 Expansive Learning at Work: Toward an Activity Theoretical Reconceptualization. Journal of Education and Work, 14(1), 133–156.
- [3] Engeström, Y. 1999 Expansive Visibilization of Work: An Activity- Theoretical Perspective. Computer Supported Cooperative Work 8, 63–93.
- [4] Saari, E., Kallio, K. and Hyytinen, K. 2008 Learning from the users of research: Defences, insights and fragile endeavors as manifestations of expansive learning. Paper presented in: Organizational Learning and Capabilities (OLKC) Conference 28.4.–30.4.2008. Copenhagen.

Exploring the Use of Humanizing Attributions in User Experience Design for Smart Environments

Minna Kynsilehto, Thomas Olsson
Tampere University of Technology
P.O. Box 589, 33101 Tampere, Finland
{minna.kynsilehto, thomas.olsson}@tut.fi

ABSTRACT

In everyday interaction with devices and technology, users tend to give human attributes to technology. They may react socially to technology as if it were a human or make guesses about its' intentions, such as "the computer is plotting against me." Smart environments, being highly ubiquitous and continuously evolving, incorporate characteristics and a new interaction paradigm that might actively trigger such humanizing attributions. Therefore, new approaches in design are required to accommodate the holistic user experience with smart environments, in ways which appreciate the human tendency to react to technology as it were human (i.e. *anthropomorphism*). We present an early research agenda for studying the phenomena related to anthropomorphism in smart environments. Based on a brief summary of background literature of anthropomorphism, smart environments and user experience, we propose few central research questions and problems how the phenomenon might affect the UX of smart environments.

Keywords

anthropomorphism, humanizing attribution, smart environments, user experience, research agenda

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation] User Interfaces – *user-centered design*; H.5.m [Information Interfaces and Presentation] Miscellaneous; H.1.2 [Models and Principles] User/Machine Systems – Human information processing.

INTRODUCTION

Although most people have little or no experience on smart and ubiquitous environments in practice, those environments possess potential to provide rich and lively user experiences. Future smart environments will be embedded with technology that offers location-specific services and resources for ad hoc needs. The possibility to use and exploit the content and resources in our proximity creates a huge potential for building smart environments with adaptive context aware services and seamless interaction. Because of the temporally evolving nature of such services, the technology can act as an agent [15], actively communicating with users and even automatically

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

taking initiative. In smart environments, users will not have manuals or tutorials in their disposal, in other words, users are required to think "on their feet" in order to understand and interact with the environment. Therefore, interaction methods have to rely heavily on intuitive, familiar processes. Human-human interaction is by far the most familiar and intuitive for people, and readily available.

What effect do human-like interaction methods have on user experience? If devices are capable of interacting with methods natural to human communication (voice, touch) what kind of expectations, beliefs and understanding users will have about technology? *Anthropomorphism* is generally understood as assigning human-like features to animals, computers and religious entities. Merriam-Webster online dictionary (www.m-w.com) defines anthropomorphism as "an interpretation of what is not human or personal in terms of human or personal characteristics". Researchers have given various accounts on anthropomorphism and suggested theories and explanations [5, 6, 11, 12, 16]. Four of these are presented in the following section, as they play a role in understanding the phenomena of anthropomorphism in human-computer interaction (HCI).

User experience (UX), on the other hand, is regarded as a comprehensive concept that describes the experience resulting from the interaction with a product or a service. In current literature, UX has been considered to involve both instrumental (e.g. utility, appropriateness, and usability) and non-instrumental elements (e.g. joy, appeal, aesthetics, social and cultural factors, as well as hedonic elements, such as identification and stimulation) [1, 2, 8]. A recent ISO standard proposal [10] defines it as "A person's perceptions and responses that result from the use or anticipated use of a product, system or service." For this research, it is central to identify what elements and characteristics of the services affect the processes of anthropomorphism, and on the other hand how UX elements and user expectations become influenced by the human like characteristics.

The purpose of this research is to investigate how anthropomorphism can help in making natural and intuitive UX and understanding such issues as trust, agency and the user's expectations, for instance. Helping to make sense of functions, processes, and capabilities of technology seems to be a never-ending quest; everyone can easily remember situations where they did not understand what their computer did or what it can do. Therefore, understanding anthropomorphism in smart environments and HCI in general is crucial for developing systems that allow intuitive interaction, utilizing knowledge of human tendency to attribute human like qualities to technology. In design, it is essential to understand, that users *will* anthropomorphize, be it accurate or not, and it's the designers job to see that interaction patterns fits the users' understanding. We believe that the design of future smart environments can drastically benefit from incorporating human like qualities and ways of interaction.

THEORETICAL BACKGROUND

So far, users' social responses to computers have been researched widely [7, 14, 16, 17] but the process of how people use humanizing models to make sense of technology functions has received very little attention.

Some researchers have looked into whether people *believe that there are more or less humane qualities* to computers at least to some extent. From studies described by Reeves and Nass [16] it is clear that people do not consciously consider computers as social actors but instead react to certain kinds of cues automatically. The phenomena they illustrate, occurs when interacting with technology while doing a task, not when they are asked to reflect on the properties of technology. In addition, Farzanfar et al. [5] state that the participants in their study of automated mental health care were not under the illusion of the machine having human qualities at any time during the study. Furthermore, Fussell et al. [6] present a study, which indicates that people differentiate conceptually what kinds of thinking humans and machines are capable of. According to Levin et al. [12] people differentiate human and computer thinking based on intentionality when considering the difference.

One explanation for anthropomorphism is treating computers and devices as *proxies*. According to proponents of this theory, users are thought to react and talk to the programmer or the builder of the device they are using. This kind of explanation has received very little support. For instance, Sundar and Nass [17] found evidence against "computer as proxy" explanation. In their study participants reacted very differently to a computer, when they thought were communicating to a programmer than when they were told they are interacting with a computer. If the proxy – explanation would be true, then the two conditions would have had very little difference.

According to Reeves and Nass [16], anthropomorphism is a natural way to make sense of and understand the environment, and deeply hard-wired into human information processing. The processes of anthropomorphism come about unconsciously, regardless the conscious realization that the object possesses no such capabilities. Furthermore, Nass et al. [14] use term *ethopoeia* to describe this phenomenon of giving human attributes to non-human entities. Brief examples of Reeves and Nass [16] study results will be listed in chapter 2.1.

A related way of addressing anthropomorphism is to talk about *mindlessness*. Mindlessness refers to the state of not reflecting on your actions or environment but rather operating on automatic models and scripts. Johnson and Gardner [11] report a study, where users with greater experience anthropomorphized more than those with less experience. They explain the result with automaticity: the more experience, the more rehearsed and thus less conscious is the interaction process with technology. Not thinking leads to the use of over-learned human interaction scripts.

In conclusion, in human-computer interaction context anthropomorphism is more than describing behavior of technology with terms used to describe human behavior. Anthropomorphism is a process of inference about unobservable characteristics of a nonhuman agent, rather than descriptive reports of a nonhuman agent's observable or imagined behavior [3].

Social Cues in Antropomorphism

In several experiments Nass and Reeves [16] show that people can treat computers as social actors. Assigning human

attributions to computers can happen with minimal social cues (eg. such as labeling a computer a team member or giving it a female or male voice) and irrespective of the conscious realization that computers are not human. These studies showed how people are polite to computers, treat computers with male and female voices differently and that people like computers with imitating a personality resembling their own. Research [7] also indicates that the number of humanizing cues increases anthropomorphism in participants. Luczak et. al [13] maintain that those devices that were seen as friends or helpers, were treated in a more friendly manner than those seen as mere tools.

SEEK Model of Anthropomorphism

Epley et al. [3] propose a SEEK model based on following determinants: Sociality, Effectance, Elicited agent Knowledge to explain the variance in the extent which people anthropomorphize non-human agents. They base their model on research done on social cognition (for a full review, see Epley, et al. [3]). The scope of the social cognition background literature is far too wide to address here, thus we present only the main points of Epley et al's [3] proposal.

Epley et al. [3] propose two motivational mechanisms, need for establishing social connections with others (sociality) and, desire for control over the environment (effectance) which have received support in further studies [4]. In both studies, those high on the need of effectance or sociality, anthropomorphized more.

In addition, Epley et al. [3] also describe a model of cognitive process of induction applied from the work of Higgins [9]. The process of elicited agent knowledge comes about when people encounter entities that elicit anthropomorphic attributions. Firstly, it activates knowledge about humans, when assessing non-human agents. Second, the activated knowledge may or may not be corrected or adjusted to incorporate non-human knowledge. Finally, the possibly adjusted knowledge is applied to a non-human agent.

Our Approach to Anthropomorphism

Concluding the aforementioned, anthropomorphism in our research is defined along the same lines as Reeves and Nass [16]; it is a human tendency to automatically attribute human qualities to technology according the cues it presents. Typically anthropomorphism occurs when the user is "thinking on their feet", i.e. trying to understand what is going on, how to interact with or what to do with technology. Results by Luczak et. al. [13] support this line of thinking as they report talking more to devices when there's a malfunction.

Taking together these lines of reasoning, ethopoeia, mindlessness, social cues and SEEK model, we conclude that the attribution process comes about as a combination of social cues, situational and personal factors.

- Social cues mean any kinds of cues that suggest human qualities or functioning, such as apparent independence, male or female voice and similarity to the user.
- With situational factors we mean factors that are related to context and use situation, such as hurry, experience or competition over attention resources.
- Personal factors are trait-like, stable qualities that usually change only little over time, for example, cultural habits and personality traits, such as sociality and effectance.

To elicit anthropomorphic thinking in users, sufficient cues will prompt human models, and situational and personal factors may affect the correction process in such a way, that the human model is applied without correction. Naturally, this is a simplified working model of the process, and following studies will hopefully shed more light on it.

ANTHROPOMORPHISM AND USER EXPERIENCE IN SMART INTEROPERABLE ENVIRONMENTS

Smart environments set both new and well-known challenges and elements for user experience. The scope of possible UX elements in smart environments in general is far too wide to tackle exhaustively in this paper. Thus we next mention only some salient and topical that might become affected by anthropomorphic processes. From our point of view such features of UX as *predictability and control*, *affordance perception*, *trust and reliability* and *naturalness of interaction* are central. Furthermore, *automation* and *agency* are implicitly present in smart environments, and cannot be neglected as topic of study in this context. In addition to the actual felt experience of the environment, also the expectations of it might become affected by anthropomorphism.

Vihavainen et al. [18] has discussed issues related to *automaticity*, maintaining that automaticity is not always beneficial and can contribute to poor user experience or abandoning the service altogether. Automation should provoke strong anthropomorphic cues; it makes technology seem independent, taking initiative and making observations. Therefore, initiative and context aware agency solutions in UIs require careful design where anthropomorphic theories could be adapted.

Nijholt [15] suggests that embodied *agents* are a useful way to provide user interface for ambient intelligence environments. Agents are strongly associated with anthropomorphism, and thus warrant inquiry of human-like features and effects of them in smart environments. From the agent point of view, considering anthropomorphic thinking about agents is particularly interesting when their apparent independence and initiative comes into play.

Of the UX elements we present here, *predictability and control* describe how much the user can determine the results of one's actions and the user's feeling whether s/he feels being able to control the processes in the environment. Control is in strong contrast with automaticity and visibility of the processes, and thus affects how the user sees the role and nature of the environment. In addition, *trust and security* issues are related to control and automaticity in many ways, for instance, allowing the environment monitor the user or knowing who has access to monitor data.

Second, *affordances*, the interaction cues present in the environment, indicate to the user which actions and resources are possible or exist in the environment. Affordances awaken user expectations; what the environment can do or how to interact with it. Cues that indicate *natural interaction methods*, such as voice, touch or gaze, may prompt knowledge of human interaction and thus awaken expectations in the user.

Altogether, what all this means when combined with the knowledge of anthropomorphism, is that we need to consider the ways of human-human interaction when assessing reliability, trusting, understanding the intentions and qualities of others and transfer this knowledge into the UX design of smart environments.

RESEARCH AGENDA

This research work produces data for a doctoral thesis work on anthropomorphism in smart environments. The high level research questions are as follows:

1. How people use knowledge of human behavior and thinking to understand and explain functioning of technology?
2. How people perceive agency and role of technology in smart environments? What expectations and beliefs it elicits?

Practical goal of this work is to determine situations where and what users anthropomorphize, when it is fruitful to use underlying human patterns to enhance UX and when it is appropriate to underline the non-human nature of technology, in other words, to provide such cues that make users assess the nature of technology correctly.

Explorative Qualitative Study

The first stage in this agenda will be gathering empirical observations on how people apply anthropomorphic explanations on various technological phenomena. As users generally have very little user experience with smart environments and natural interaction methods, we begin with exploring anthropomorphic thinking with computers and other high-tech devices. Especially, we are interested in phenomena that are related to smart environments: agency, automation, control and trust. Main research questions will include the following:

- What are the most typical or frequent kinds of anthropomorphic thinking in technology use?
- What is the function of aforementioned kind of thinking? What are the benefits and drawbacks?

Methods used in the study will be qualitative for the main part: interviews and projective methods. Interviews will be loosely or semi-structured, aiming for gathering qualitative data on how people make sense about technology and what kind of relationships they form with their devices. We seek to answer the following questions:

- How users perceive device agency? What kinds of assumptions are related to agency? Is it similar to human agency?
- What makes users trust technology? What human-like features in technology might affect trust?
- What kind of roles users assign to technology and how does this affect their thinking about it?
- What human-like attributions users employ when interacting with and controlling the technology?

Participants to this study will be selected on the basis of technology experience, balancing for gender and age. As smart environments, as envisioned in the ubiquitous computing research agenda, are very rare, people have no or little experience of them. Therefore, we aim at recruiting users with high experience on mobile devices and services, virtual environments, or novel gaming devices. As we aim for rich qualitative data, the sample size will not be large. Estimated number of interviewees is 10.

The method of analysis will be discourse analytic, aiming to extract common processes and typical categories of anthropomorphic thinking. The main goal of the analysis is to identify the most frequent type of anthropomorphic thinking and the factors that affect the phenomenon.

Further Studies

The purpose of further studies is to assess the way humanizing cues affect anthropomorphism, especially users' expectations and understanding of system functionality. Based on results from the first stage of this research agenda, some factors will be applied in consequent studies. Variable selection will be based on salience, relevance in terms of use cases, and research literature. Due to nature of smart environments, we believe that agency, automation and natural interaction methods will be among the factors in further studies.

Factors are then measured in relation to the suggested variables that affect anthropomorphism: strength and quality of human-like cues, situational and personal factors. In addition to more qualitative methods, the studies will include experimental set-ups, where UX factors will be measured in relation to variable quality and intensity. These results are used to understand anthropomorphism in HCI and in devising methodology for UX design.

ACKNOWLEDGMENTS

This work was supported by the Finnish Funding Agency for Technology and Innovation (TEKES) under the project DIEM (grant 40223/07). We thank Sari Kujala and Veikko Surakka, as well as all the DIEM colleagues, for their insightful comments.

REFERENCES

- [1] Arhippainen, L. and Tähti, M. 2003. Empirical Evaluation of User Experience in Two Adaptive Mobile Application Prototypes, In: Proc. of MUM 2003, Norrköping, Sweden. Pp. 27–34.
- [2] Battarbee, K. and Koskinen, I. 2005. Co-experience: user experience as interaction. *CoDesign*, 1(1), 5–18.
- [3] Epley, N., Waytz, A. and Caccioppo, J. T. 2007. On seeing human: a three factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886.
- [4] Epley, N., Akalis, S., Waytz, A. and Cacioppo, J. T. 2008. When we need a human: motivational determinants of anthropomorphism. *Social Cognition*, 26(2), 143–155.
- [5] Farzanfar, R., Frishkopf, S., Friedman, R. and Ludena, K. 2007. Evaluating an automated mental health care system: making meaning of human-computer interaction. *Computers in Human Behavior*, 23, 1167–1183.
- [6] Fussell, S. R., Kiesler, S., Setlock, L. D. and Yew, V. 2008. How people anthropomorphize robots. *HRI'08*, March, 12–15, Amsterdam, Netherlands.
- [7] Gong, L. 2008. How social is responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24, 1494–1509.
- [8] Hassenzahl, M. and Tractinsky, N. 2006. User Experience – A Research Agenda. *Behaviour and Information Technology*, 25(2), 91–97.
- [9] Higgins, E. T. 1996. Knowledge activation: Accessibility, applicability and salience. In: Higgins, E. T. and Kruglanski, A. W. (Eds.). *Social psychology: Handbook of basic principles*, New York: Guilford Press. Pp. 133–168.
- [10] ISO DIS 9241-210:2009.
- [11] Johnson, D. and Gardner, J. 2009. Exploring mindlessness as an explanation for the media equation: a study of stereotyping in computer tutorials. *Personal Ubiquitous Computing*, 13, 151–163.
- [12] Levin, D. T., Killingsworth, S. S. and Saylor, M. M. 2008. Concepts about the capabilities of computers and robots: a test of the scope of adults' theory of mind. *HRI'08*, March, 12–15, Amsterdam, Netherlands.
- [13] Luczak, H., Roetting, M and Schmidt, L. 2003. Let's talk: anthropomorphization as means to cope with stress of interacting with technical devices. *Ergonomics*, 46(13/14), 1361–1374.
- [14] Nass, C., Steuer, J., Tauber, E. and Reeder, H. Anthropomorphism, agency & ethopoeia: computers as social actors. <http://portal.acm.org/citation.cfm?id=259964.260137>.
- [15] Nijholt, A. 2004. Where computers disappear, virtual humans appear. *Computers & Graphics*, 28, 467–476.
- [16] Reeves, B. and Nass, C. 1996. *The Media Equation: How People treat computers, television and new media like real people and places*. CSLI publications, Cambridge.
- [17] Sundar, S. S. and Nass, C. 2000. Source orientation in human-computer interaction. *Communication research*, 27(6), 683–703.
- [18] Vihavainen, S., Oulasvirta, A. and Sarvas, R. 2009. "I can't lie anymore!": The implications of location automation for mobile social applications. *Proceedings of MobiQuitous*.

Human-Centred Design of Technical Infrastructures of Smart Environments

Marketta Niemelä
VTT Technical Research
Centre of Finland
P.O. Box 1300
33101 Tampere, Finland
Marketta.Niemela@vtt.fi

Thomas Olsson
Tampere University of
Technology
P.O. Box 589
33101 Tampere, Finland
Thomas.Olsson@tut.fi

Eija Kaasinen
VTT Technical Research
Centre of Finland
P.O. Box 1300
33101 Tampere, Finland
Eija.Kaasinen@vtt.fi

ABSTRACT

This paper discusses human-centred design of technical infrastructures of smart environments. Smart environments make our everyday environments computerized by embedding information and communication technology in them. The design of smart environments includes to a great extent design of technical infrastructures that facilitate interaction within and with the environment. Therefore, human-centred design should be extended from individual services to focus already on the technical infrastructures that the services will be built on. This paper describes two project cases. The first case concerns the design of a technical infrastructure that facilitates interaction with the smart environment with personal mobile phones. The latter case reflects the design of a technical infrastructure facilitating interoperability of different devices and services for smart environments.

Keywords

human-centred design, ubiquitous computing, smart environment, user experience

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *user-centered design*.

INTRODUCTION

In the close future, we will not only use and exploit single smart devices, applications, and services. Instead, we will enter and act in smart environments, in which the services and recourses all seamlessly interoperate with each other to benefit the user. For instance, in the vision of the DIEM project [2], devices, services and applications openly share their data to be utilised by other services and finally the user. The notion of smart environment is based on ubiquitous computing, which refers to computerisation of everyday environments – embedding information and communication technology in them.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

From the user's point of view, smart environments open up huge possibilities but also pose several challenges. For instance, how does the user know what embedded information and services are available in the environment, and how to grab and control those? How does the user know which information s/he leaves into the space? Trust, control, and literacy of smart environments, among other issues, essentially affect user experience of smart environment and should be taken into account in the design of smart environments from the very beginning.

Human-centered design [5] is a well-established practice to the design of individual software applications. The approach aims at understanding the needs of users, specifying the potential contexts of use, and evaluating the design solutions against user requirements. Many application features are, however, not defined only in the application itself but by the underlying technical infrastructures such as device platforms, databases, middleware and interface libraries that are typically fixed before the application development takes place. Studying those system-level solutions and their effects at the end-user level is underlined with the coming of smart environments that constitute of different technical infrastructures on which individual applications and services are built on. Human-centred design of the technical infrastructures is needed to ensure that the environment will support such application features and interaction paradigms that provide successful user experience. This approach becomes even more significant when designing smart environments with rich and temporally evolving user experience

We present the approach of human-centred design of technical infrastructures for smart environments by reflecting two project cases, MIMOSA and DIEM.

CASE MIMOSA

We started developing the methodology for human-centred design of a ubiquitous computing architecture [7] in MIMOSA project [10]. The project was developing novel microsystems solutions for wireless sensors and tags as well as a mobile platform that facilitates connecting those microsystems wirelessly to ordinary mobile phones and other personal mobile devices. The architecture (Figure 1) provides the applications with wireless connections to small tags and sensors, thus letting the applications utilize different identification, measurement and context data. Potential application fields include health care, fitness, sports and everyday consumer applications. The assortment of services facilitated by this system level solution is much wider than dealt with in the studies of software architectures [1, 3, 4, 8]. We involved potential users and application field experts in the design process to get information about the type and quality of applications that the architecture should facilitate.

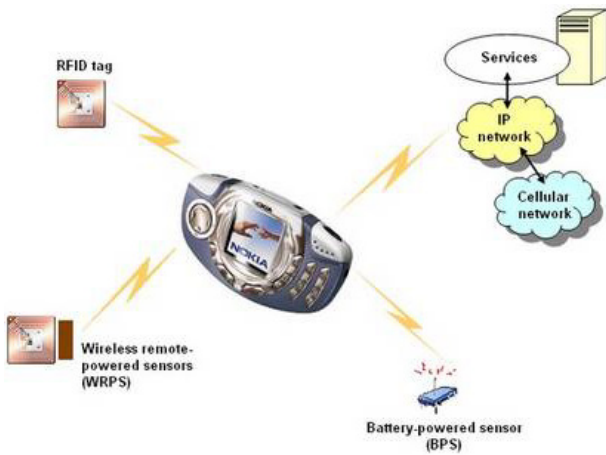


Figure 1. The MIMOSA architecture.

Human-centred design approach was applied in the MIMOSA project at two levels. Firstly, from the end user’s point of view the aim was to identify useful and acceptable application concepts as well as to define user requirements for these concepts both regarding the individual applications and the architecture. Secondly, application developers assessed the applicability of the architecture as a platform for different applications both from application and application development points of view. The objective was to identify the functionalities that future applications would need and to apply human-centred design to ensure that the architecture would enable those functionalities in a usable way.

The human-centred design approach included the following, partially overlapping steps:

- Creation of usage scenarios on the selected application fields to clarify future possibilities offered by the architecture
- Refining the scenarios based on feedback from users, application field experts and technical experts of the project
- Evaluation of the scenarios with end users and application field experts to refine the scenarios
- Analysis of the scenarios down to use cases and further into sequence diagrams to describe in details what kind of communication they indicated between the user, different parts of the architecture and external actors.
- Building proof-of-concept demonstrators to illustrate key look-and-feel features in the applications and to compare alternative solutions
- User and expert evaluations of the proof of concepts
- Defining usage requirements for the architecture based on use case analysis and feedback from users, application field experts and technical experts of the project
- Implementing MIMOSA first generation demonstrators
- Evaluation of the demonstrators and their implementation process.

To identify the implications on the architecture, the use cases and related user feedback from scenario evaluations were analysed together with human factors experts and architecture designers. Special attention was paid on common usage patterns that were repeating from one scenario to another as they were potentially the core features of the architecture.

In MIMOSA project scenarios worked well as design tools. The rich scenario material ensured that the functionality of the architecture could be studied in different usage situations and with diverse use cases. Several common patterns were recognized in the scenarios. The systematic analysis of the scenarios made it possible to identify quite early and with moderate resources several important user and application requirements that had implications on the architectural design.

Analyzing the user and application requirements into requirements for the MIMOSA platform architecture was not a straightforward job. Technical documentation did not help very far but several workshops were needed with architecture and component designers to fully understand the structure and functionality of the architecture, and the boundaries for design solutions.

However, user and application requirements could be traced back to requirements for the architecture. The rich scenario material from the very beginning facilitated identifying common use cases that the architecture would need to support. User evaluations of the scenarios and proof of concepts gave insight how these use cases should work in practice.

CASE DIEM

Based on the experiences of MIMOSA, the human-centred design approach of technical infrastructures for smart environments is being further developed in the DIEM project [2].

The goal of DIEM is to build new kind of smart environments that comprise of ecosystems of digital devices (Figure 2). The DIEM ecosystem is based on an interoperability architecture for devices, services, and information from different domains [9]. Digital devices and systems contain relevant information, physical and digital resources, and services for various purposes. Hence, by being able to extract, combine and integrate all this content on a common innovation platform, new services and applications can be created for ad hoc needs [9]. The possibility to use and exploit the content that can be made available from our proximity opens up a huge potential for building entire ecosystems and smart environments with novel services and seamless interaction. The project focuses on a variety of types of physical contexts: buildings as smart environments (mostly indoors: homes and offices) and public spaces (e.g. piazzas, parks, fitness-centers and marketplaces).



Figure 2. DIEM smart environment based on digital device ecosystem.

The challenges of interoperable smart environments for the user are somewhat different than in MIMOSA. First, the DIEM smart environment is dynamic and evolving by nature: new devices and users can join it, bringing their information content and services in, and then leave the environment. Application designers – also end users as designers – can test new applications,

elaborate them, and then remove them. Second, interaction with and in the smart environment emphasizes the social aspects, as others can participate simultaneously, observe or prompt actions that ask for involvement. In short, the smart environment is characterised by open information sharing, gradual change and evolution, and user involvement by producing content to the environment and easy innovation of new applications.

As the ultimate mission of the project is to create a concept and implementation of a generic and scalable smart space interoperability solution and platform, which then can be adapted to various domains and applications, it is critical to extend human-centred design to the architecture and platform level. A non-predictable number and type of applications and services will depend on the DIEM system level solutions.

In DIEM, we apply human-centred design at the system level by collecting, analysing and interpreting user requirements for the functional architecture and platforms. An essential part of the process has been to identify the usage patterns enabled by the interoperability architecture. These abstract usage patterns form an analysis tool to help analysing the user feedback considering interaction with the smart environment down to the system level to support good user experience.

The usage patterns have been identified in collaboration with the technical system developers of the project as follows:

User-activated Usage Patterns

LOOK AT AVAILABLE SERVICES: User enters the space and wants to see what services are available.

GET DATA: User downloads data from the space

PUT DATA: User uploads data into the space

SYNCHRONISE: User asks device to synchronise its (certain) data (with other devices)

USE SERVICE: User utilises available resource or service

SUBSCRIBE FOR SERVICE/ALERT (TIME/EVENT TRIGGERED): User wants to have service or alert either at certain times or in a certain context

UNSUBSCRIBE: User does not want any more the subscribed service/alert

Automatic Usage Patterns

ALERT/NOTIFICATION: User gets automatic alert (no subscribe needed beforehand)

SITUATIONAL SERVICE: User gets automatic service (no subscribe needed beforehand)

KEEP IN SYNCH: Device synchronises its data automatically (time/event triggered)

GET SUBSCRIBED SERVICE/ALERT: User gets automatically service or alert that has been subscribed before

MONITOR: Device monitors the environment looking for relevant data (or services)

The abstract usage patterns form a framework to interpret user feedback considering the application level down to the system level. We expect to identify user feedback and requirements on critical user experience issues related to the core characteristics of DIEM smart environments (interoperability, multimodality, openness, and evolving nature). As interoperability may make it difficult to perceive where the information is from, how it is built up and can it be trusted, such issues as control, reliability,

and simplicity of interaction should be designed at the most early phase to smart environments. Critical user experience elements are, for instance:

- 1) *Attention division:* the services in the DIEM environment need to grab and hold the user's attention. Because there may be many other potential points of interests, competing tasks, other users and sources of distraction designers cannot rely on having the user's undivided attention.

At the system level: should there be implementation of priorities in which applications or information compete of the user's attention – e.g., when listing available services or monitoring the environment for relevant data

- 2) *Control & Predictability:* Describes the user's feeling of control, whether they feel being able to control the processes in the environment, starting new ones, interrupting them, and preventing those that they wish not to use. Predictability refers to the extent to which the user can determine the results of their actions based on their interactions with the system.

At the system level: should the architecture support or implement generic, systematic control methods or affordances for the application level to make user interaction more predictable and easy to learn.

- 3) *Reliability & Trust:* Reliability refers both to system functions as well as to validity of the information that the system offers to the user. A reliable system has little errors, behaves consistently over time and situations, and prevents leaks of information to unwanted parties. User trust in mobile services includes perceived reliability of the technology and the information and services provided, reliance on the service in planned usage situations, and the user's confidence that (s)he can keep the service under control and that the service does not misuse his/her personal data [6]. The users need to be aware of the risks in using the product and must have reliable information to assess how much one can rely on the technology in different usage contexts and situations.

At the system level: should the architecture include methods to trace the source and check the reliability of information. For instance, reliability of information provided by user-build mash-up applications may be a problem. Furthermore, should the architecture include methods to inform the user where the data s/he shares may end up.

During the project, the usage patterns will be illustrated in different ways and evaluated. At the time of this writing, the usage patterns are being illustrated as scenarios in the form of cartoons (Figure 3).

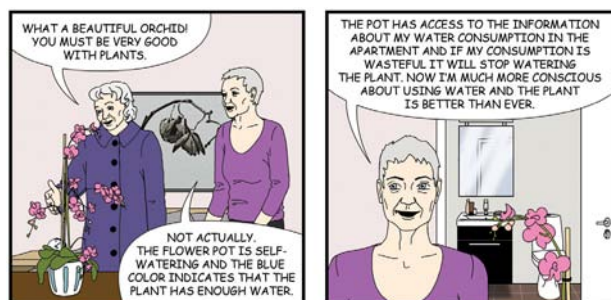


Figure 3. DIEM scenario cartoon clip.

In the scenarios the interaction methods, although multimodal according to the DIEM objectives, may not be innovative as such in purpose: the user attention needs to be directed to the quality of interaction and combinations of interaction (usage) patterns instead of novel interaction methods or tools, to facilitate the analysis of feedback on interoperability platform.

Our next step is to analyse and refine the scenarios with end users and analyse the feedback, which based on the MIMOSA experience mainly concerns the application level, against the patterns. In the scenario evaluations, we will systematically gather feedback on different ways these usage patterns are represented as applications, and analyse the feedback to identify such core issues that should be implemented at the system level. These issues will be defined as user requirements for the interoperability platform. Our goal is to describe in details what kind of communication in the interoperability based smart environment promotes the most positive user experience.

CONCLUSIONS

To make ubiquitous applications become common, appropriate platforms and architectures for the applications need to be available. Well-designed architectures serve two goals: first, they ease application development by providing basic functionalities and second; they ease adoption of the applications to the users as the applications are available on the same platform and they comply with common usage patterns. MIMOSA experiences point out that several architectural design decisions concerning a ubiquitous computing architecture have significant influences on the end user experience. MIMOSA experiences also show that user requirements regarding the architecture can be identified early in the design by illustrating the forthcoming applications to potential users as scenarios and proof of concepts. In DIEM, we are further developing the approach of human-centred design of the system-level solutions with some systematical changes in the approach. In order to facilitate the analysis of user feedback at the system level, we have first defined the usage patterns and only after that built the scenarios to illustrate the patterns. In user evaluations, we look forward to receive feedback on how the usage patterns should be implemented at the system level so that they support positive user experience in the smart environment. In addition to the scenarios, the project will create and evaluate proof-of-concepts and complete service pilots and demonstrators. We are targeting to define efficient and scientifically justified methodology and guidelines for designing good user experience in interoperability-based smart environments.

ACKNOWLEDGMENTS

This work was supported by the Finnish Funding Agency for Technology and Innovation (TEKES) under the project DIEM (grant 40223/07). We thank Minna Kynsilehto for analysing DIEM user experience elements.

REFERENCES

- [1] Bass, L. and John, B. E. 2003. Linking usability to software architecture patterns through general scenarios. *Journal of Systems and Software*, 66, 187–197.
- [2] DIEM project description. 2008. www.tivit.fi/en/device
- [3] Edwards, W. K., V. Bellotti, V., Dey, A. K. and Newman, M. W. (2003). Stuck in the Middle: The Challenges of User-Centered Design and Evaluation for Infrastructure. In: *Proceedings CHI2003*, ACM Press (2003). Pp. 297–304.
- [4] Folmer, E. and Bosch, J. 2004. Architecting for usability: a survey. *Journal of Systems and Software* 70, 61–78.
- [5] ISO 13407:1999. 1999. Human-centered design processes for interactive systems. International standard of International Standardization Organization, 1999.
- [6] Kaasinen, E. 2005. User Acceptance of Mobile Services – Value, Ease of Use, Trust and Ease of Adoption. Doctoral Thesis. VTT Publications 566, Finland, 2005. 151 p. + app. 64 p.
- [7] Kaasinen, E., Niemelä, M., Tuomisto, T., Välikkynen, P. and Ermolov, V. 2006. Identifying user requirements for a mobile terminal centric ubiquitous computing architecture. In: *Proceedings of International Workshop on System Support for Future Mobile Computing Applications (FUMCA 06)*, IEEE (2006). Pp. 9–16.
- [8] Kazman, R, Abowd, G., Bass, L. and Clements, P. 1996. Scenario-Based Analysis of Software Architecture. *IEEE Software*, November 1996, 47–55.
- [9] Lappeteläinen, A., Tuupola, J.-M., Palin, A. and Erikson, T. Networked systems, services and information: the ultimate digital convergence. First NoTA Conference, 2008. <http://www.notaworld.org/files/common/TheUltimateDigitalConvergence.pdf>.
- [10] MIMOSA project home page. 2008. <http://www.mimosa-fp6.com>.

“Show me, how does it look now”: Remote Help-giving in Collaborative Design

Dhaval Vyas¹, Gerrit van der Veer^{1,2}, Anton Nijholt¹, Dirk Heylen¹

¹Human Media Interaction, University of Twente, 7500 AE Enschede, the Netherlands

²School of Computer Science, Open University Netherlands, 6419 AT Heerlen, the Netherlands
{d.m.vyas | a.nijholt | d.k.j.heylen}@ewi.utwente.nl; gerrit@acm.org

ABSTRACT

This paper examines the role of visual information in a remote help-giving situation involving the collaborative physical task of designing a prototype remote control. We analyze a set of video recordings captured within an experimental setting. Our analysis shows that using gestures and relevant artefacts and by projecting activities on the camera, participants were able to discuss several *design-related* issues. The results indicate that with a limited camera view (mainly faces and shoulders), participants’ conversations were centered at the physical prototype that they were designing. The socially organized use of our experimental setting provides some key implications for designing future remote collaborative systems.

Keywords

remote help-giving, awareness, common-ground, collaborative physical task, design

INTRODUCTION

Computer-mediated communication (CMC) systems that involve *collaborative physical tasks*¹ should support coordination of participants’ speech as well as their actions [2, 3, 6, 7]. Kraut et al. [7] suggest that in such a CMC system, supporting ‘mutual awareness’ and establishing ‘common-ground’ between participants are the two important issues. Here, visual information becomes a major resource of communication and a support to verbal exchanges between the participants. The visual information about the object in question and other relevant information (e.g. gestures) not only help participants maintain and gain up-to-date understanding about the current situation but also allow participants to establish a common-ground during task performance.

In our project we focus on understanding the nature and the role of visual information as a resource for conversations in remote collaborative physical task. In the current phase we consider the aspect of ‘assisting’ or ‘help-giving’, in a task of co-designing a prototype remote control. Here, one of the participants uses different types of clay to design a prototype remote-control. We refer to him/her as Industrial Designer (ID). The second participant, at a remote location, provides assistance and guidance during this process without having direct access to the design material. We refer to him/her as User Interface Designer (UID). Here ID has knowledge about

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

product development, technology use and their integration, whereas UID can provide user-focused guidance. Hence, ID and UID have complementary expertise.

In our experimental setup (Figure 1), both participants were equipped with high resolution cameras with adequate support of audio-video technologies. The cameras could show participants’ heads and shoulders. Both participants could adjust their camera views, if needed. Both of them had the same documentation and specifications about the design of prototype remote control, but only ID had the design materials to develop a prototype.



Figure 1. Experimental setting of remote help-giving.

In this paper we report an analysis of 9 design sessions with different IDs and UIDs captured on videos – approximately 40–60 minutes each. The videos show both participants interacting with each other in real-time. From the analysis, we show that by using gestures and showing relevant artefacts and by projecting activities to the camera, both participants established mutual awareness and common-ground.

In the following, first, we describe the results of our analysis using several examples. Next, we discuss some issues discovered and describe future work.

RESULTS

We found three types of visual information utilized by the participants for establishing awareness and common-ground: 1) use of gestures, 2) use of artefacts, and 3) projecting activities

1. A ‘collaborative physical task’ involves two or more people working together on physical objects in the real-world. [3]

on camera. We will show how this visual information enabled adjustments in the design of the remote control prototype – allowing discussions of size, shape, interaction mechanisms and ways of using.

Use of Gestures

Participants used gestures to make each other aware of the situation as well as to provide instructions for specific actions. Their gestures were mainly *object-focused*, i.e. referring to the physical prototype. Both participants used head movements and facial expressions to convey agreement, confirmation or otherwise. This was a quick way to say ‘yes’ or ‘no’ to the other participant. We will describe some specific patterns that allowed more detailed communications.

Pointing to a Specific Part

On several occasions pointing to a particular part of the remote control was used to communicate ideas. For example, in Figure 2, this ID used pointing gestures to locate specific parts of the prototype and to describe position of buttons and screen of the prototype remote control. This kind of gesture was mainly used by IDs as they had direct access to the prototype. In order to make relevant design decisions, IDs needed to point to a specific portion of the prototype to discuss details.



Figure 2. ID (left) points to a part of the prototype to communicate with UID.

Describing a Specific Shape

Since UIDs did not have direct access to the prototype, UIDs frequently used gestures to communicate shapes and size of the prototype, to describe interaction mechanisms and to explain ways of using the prototype. From the two examples shown in Figure 3, (a) shows a UID explaining the size and shape of a button, and (b) shows a UID using a two-handed gesture to demonstrate a flap-like interaction mechanism for the prototype remote control. A fragment of the conversation from example (b) illustrates how participants were able to discuss different design possibilities through visual information.

ID : "The bottom of the remote control could have a slider"
 UID : "Fantastic. Yeh, that's my idea"
 UID : "I also like a flap window, like in the Motorola phones, you know" (UID poses as shown in Figure 3b)



Figure 3. UID (right) describes specific shapes using gestures to communicate with ID (right).

Animated Gestures

Some of the aspects related to the prototype remote control were not easily describable in words or through showing the prototype only. We observed that participants used animated gesture to explain their ideas clearly. Figure-4 shows an example when an ID is describing a ‘sliding’ behavior to confirm with UID’s suggested mechanism. Here is a short excerpt of their conversation.

UID : "Do you think you can make a sliding mechanism for the prototype as we had discussed in the last meeting?"
 ID : "You mean, like this..." (ID uses an animated gesture as shown in Figure 4 to demonstrate a sliding mechanism)
 UID : "Yeh, that way"



Figure 4. ID (left) uses animated gestures to make a design suggestion.

Use of Artefacts

We observed that in combination with speech different artefacts were used by the participants for aiding mutual awareness, for continuous coordination and for directing participants’ focus of attention. These artefacts included the prototype remote control but also other artefacts like paper-based drawings and some hybrid coordination techniques – mixing gestures with artefacts.

The Design Object

As the remote control prototype is the main source of the discussion, IDs have to continuously update UIDs by positioning it close to the camera, whenever needed. Here the temporality of the design object becomes very important. This temporality helps establishing a common understanding of the process. If the camera focused on the faces of the participants, the remote UID had to request to see the current state of the prototype by asking “*show me, how does it look now?*”, for example. Visual information related to the design object not only helped for establishing mutual awareness or common-ground, it also improved conversational efficiency. For example, when a UID could see what an ID had done, he/she would confirm or intervene appropriately.



Figure 5. Different stages (a, b & c) of the remote control projected by ID (left) to UID (right).

Related Materials

We also observed that participants used other materials like paper based sketches and drawing diagrams in order to communicate ideas to each other. An example is illustrated in Figure 6. Figures 6a and 6b point to different time-frames. In this particular case, using a sketch, remote a UID assists an ID throughout the design process. This can be seen in the figures where the UID works on her drawing while simultaneously explaining her drawings to the ID. Here, the development of the physical prototype of the remote control (accessible only by ID) and drawing sketches (used by UID) go hand in hand.

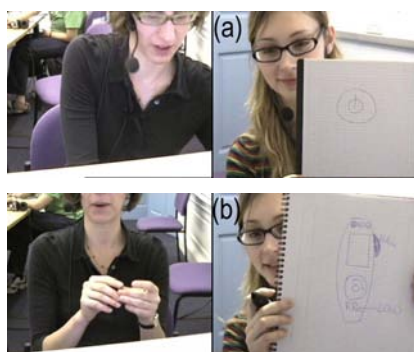


Figure 6. UID (right) continuously assists ID (left) using drawing sketch.

Hybrid Coordination

There were specific coordinative patterns where participants used a mix of gestures and artefacts in order to establish common-ground. We observed several instances of these types of hybrid coordinative patterns. As can be seen in Figure 7, a UID (right) mimics the shape of the prototype remote control to be able to explain a specific position that needs to be re-adjusted. Both ID and UID play a role here in establishing a common-ground. A fragment of their conversation illustrates the importance of this.

UID : "I think you could put the volume control there"
 ID : "Hum...?"
 UID : "Where your thumb is... "
 ID : "Here?"
 UID : "No..here"(UID poses as shown in Figure 7)

ID : "Here?"(ID poses as shown in Figure 7)
 UID : "Ya...ya"



Figure 7. An example of material common-ground established by participants.

Projecting Activities on Camera

We observed that projecting different activities towards the camera (i.e. showing actions in front of the camera) allowed participants to make each other aware of the start and progress of a particular design activity. This was done only by the IDs since they had direct access to the prototype. Since both participants had a limited view of each other, at specific times an ID projects available clay materials, adjusts the camera view to focus on specific parts and also adjusts the position of the prototype remote control to keep UID aware and up-to-date about the ongoing activities. By projecting activities on the camera the information is intentionally made commonly visible which in turn makes the production and understanding of references (made during conversations) easier.

As it can be seen in Figure 8, physical actions were projected so that the intended participant can see these actions and their meanings. Public visibility as a coordinative aspect has been echoed by many others [4, 8]. Especially, Robertson suggests that the public availability of different artefacts and embodied actions to be perceived by distributed participants in a cooperative process could enable their communicative functions.



Figure 8. Projecting actions on the camera.

Available Materials

We observed that in order to establish a common-ground, at the beginning of all design session ID shows all the materials available to him to UID. This enabled UID to better assist ID in the design process.

Adjusting Camera

Both ID and UID were able to adjust the focus of their own cameras as they were able to see their own view in addition to each other's views. As shown in Figure 9, an ID zooms in to the prototype to show details. This kind of activities occur either when requested by the UID or when they both finish an aspect of their specific phase of design activity. It was also seen that sometimes an ID forgot to adjust the focus of the camera, which did not provide sufficient information to UID.



Figure 9. ID (Left) adjusts camera to zoom in on the prototype.

DISCUSSION

In this paper, we examined how participants coordinated the design of a prototype remote control in an audio-video mediated environment. We have collected different patterns of establishing mutual awareness and of building common-ground between participants. Echoing others [6, 7], our results demonstrate that help-giving during remote collaborative physical tasks requires complex coordination between participants. Participants have to decide how and when to provide instructions and how to align these with their conversations.

We found that collaborative design activities were facilitated by three types of visual information: gestures, artefacts and projecting activities on camera. In table-1, we list design activities that our participants carried out using these categories of visual information. This list should not be seen as a complete taxonomy but it reflects the importance of visual information in a remote coordinative physical task. Importantly, we observed how participants integrate and align their activities using both behavior and speech.

Table 1. Different visual information and design-related activities they support.

Gestures	<ul style="list-style-type: none"> • Describing shape and size • Mimicking interaction mechanisms • Pointing and describing a position • Referring to actions required on a part of the object
Artefacts	<ul style="list-style-type: none"> • Transferring work-in-progress information • Showing shapes (using a drawing sheet) • Discussing planning mechanisms • Setting knowledge landmark for future actions
Projecting Activities on Camera	<ul style="list-style-type: none"> • Showing available materials • Status updates • Making information publicly visible • Directing co-worker's focus of attention

Why does Visual Information help in the domain of Cooperative Design?

Both participants had a different, geographically separated ecological setting. It has been shown that participants who share a common physical space can better coordinate each other's activities than when they collaborate from remote locations [1]. Our findings show that the three types of visual information that we identified help in building 'common spaces'. Participants have to rely on design object, paper drawings,

and creating common-ground through gestures. These provide a common frame of reference that supports awareness between remote ecologies and enables participants to align and integrate their collaborative activities.

The richness of gestures, artefacts and projected activities allows participants to effortlessly make sense of the co-worker's actions, as these are really mundane and participants do not have to 'decode' any abstract representations. As shown in [5], the intersubjective intelligibility of the common spaces, which are built within two separate ecologies, help in establishing an efficient coordinative environment.

Design is an inherently 'visual' domain. Our previous study [10] shows that visual information like sketches, physical models and prototypes developed within different fields of design (e.g. industrial & product design, architecture) help in coordinating design activities. We were able to confirm this visual character of design.

FUTURE WORK

Our overall research goal is to develop technologies to support remote cooperative design. The experimental setting that is used in our study provides indications of how visual information could be critical in supporting awareness and establishing common-ground amongst remotely located participants. We intend to apply more reliable ways of registering and interpreting these coordinative processes and to identify patterns. We also plan to expand our analysis to more than two participants, where we intend to have the system to perform real-time pattern analysis in order to support multiparty collaboration.

A following step will be to study design practices in real world. It has been evident from the past experiences of media spaces [5, 9] that because of the impoverished understanding of 'collaborative work', media space-like environments have not been very successful. Clearly, real world practices of designers are needed for understanding real-life coordination mechanisms.

ACKNOWLEDGMENTS

This work is supported by the European IST Programme Project AMIDA (FP6-0033812). This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

REFERENCES

- [1] Clark, H. H. and Brennan, S. E. Grounding in communication. In: Resnick, L. B., Levine, R. M. and Teasley, S. D. (Eds.). Perspectives on socially shared cognition, 1991, American Psychological Association. Pp. 127–149.
- [2] Fussell, S. R., Setlock, L. D., Parker, E. M. and Yang, J. Assessing the value of a cursor pointing device for remote collaboration on physical tasks. In: CHI '03 Extended Abstracts. ACM: NY, 2003. Pp. 788–789.
- [3] Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E. and Kramer, A. D. Gestures over video streams to support remote collaboration on physical tasks. Hum.-Comput. Interact, 19(3), 2004, 273–309.
- [4] Heath, C. and Luff, P. Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms. Computer Supported Cooperative Work, 1(1), 1992, 24–48.

- [5] Heath, C., Luff, P. and Sellen, A. Reconsidering the virtual workplace: flexible support for collaborative activity. In: Proceedings of ECSCW'95. Kluwer Academic Publishers, Norwell, MA, 1995. Pp. 83–99.
- [6] Kirk, D., Crabtree, A. and Rodden, T. Ways of the Hands. In: Proceedings of ECSCW'05. Kluwer Academic Publishers, 2005. Pp. 1–20.
- [7] Kraut, R., Fussell, S. and Siegel, J. Visual Information as a Conversational Resource in Collaborative Physical Tasks. *Hum.-Comput. Interact*, 18(1), 2003, 13–49.
- [8] Robertson, T. Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions. In: Proceedings of ECSCW'97, Kluwer Academic Publishers, 1997. Pp. 205–220.
- [9] Schmidt, K. The problem with 'awareness': Introductory remarks on 'Awareness in CSCW'. *Computer Supported Collaborative Work*. 11: Springer Netherlands, 2002, 285–298.
- [10] Vyas, D., Heylen, D. and Nijholt, A. Physicality and Cooperative Design. In: Proceedings of the 5th international Workshop on Machine Learning for Multimodal interaction. Popescu-Belis, A. and Stiefelwagen, R. (Eds.). *Lecture Notes in Computer Science*, Vol. 5237. Springer-Verlag, Berlin, Heidelberg. Pp. 325–337.

Session 4: Acting in Computerised Everyday Environments

A Descriptive Model of Contextual Activities for the Design of Domestic Situations

Pascal Salembier
ICD-Tech::CICO (FRE 2848
CNRS) University of Technology
of Troyes Troyes, France
Pascal.Salembier@utt.fr

Julie Dugdale
MAGMA-LIG
University of Grenoble
Grenoble, France
Julie.Dugdale@imag.fr

Myriam Fréjus,
Yvon Haradji
EDF R&D
Clamart, France
myriam.frejus, yvon.haradji@edf.fr

ABSTRACT

A challenging topic for cognitive ergonomics and cognitive engineering is the development of smart applications and devices which apply some “intelligence” of the situations, i.e. commonsense knowledge about the occupants of the household, their individual and collective routines, their expected patterns of behavior. Most people spend more time at home than in any other place, including work places, but few studies have been conducted on how new context-aware technologies could support people in their everyday life. Spaces that subtly reconfigure themselves according to their occupants and use can cause rejection or acceptance, depending on how intelligently they are designed. In this paper we describe a descriptive framework for contextual activities that aims at supporting collective thinking about the design of services for the domestic users.

Keywords

scenarios, modeling of situations, design, contextual technologies

ACM Classification Keywords

H.1.2 [Models and principles (e.g., HCI)] User/Machine Systems – *Human Factors*.

INTRODUCTION

Contextual technologies are said to help us manage different critical aspects of individual and collective situated activities in various social contexts. But meantime, it can be seen as a new source of complexity and a challenge designers have to face when introducing new services. The range of traditional issues have extended to new ones: for example what can we expect when moving information technology from the center of actors’ focal attention into the periphery and what are the likely consequences on the organization of cooperation between users and devices and between the occupants of households themselves. In the recent years many studies have been conducted in order to better understand the articulation between ambient technologies and the social organization of domestic activities. These studies show the variety of approaches currently used, from a mere description of domestic life through ethnographic studies which attempt to assess the impact and role of technologies on

domestic life, to more technology oriented views which tend to focus on computational, rather than social aspects of domestic life [1, 2].

The approach we have adopted in the In-Situ project is to build heuristic descriptive models of contextual activities in order to provide a basis to support thinking about the design of services for the domestic users. The idea is not to provide a computational cognitive model of the contextual mechanisms held by the members of the household when coordinating their individual and collective activities, but rather to define:

- An acceptable, semi-formal description of these activities with a special emphasis on particular dimensions.
- A descriptive model, that is a descriptive abstraction of actors’ activities in their environment, which refers to the analyst’s understanding of this behaviour.
- A transition towards the use of models for design, that is proactive models that can orientate and assess some design issues concerning the possible interest and impact of introducing contextual technologies in home settings.

The Section: “Background” describes the background of the framework and the notion of model-based evaluation and design at the center of the approach. Section: “General Approach” provides details of the model and focuses on describing the notions of viewpoints and cognitive equipment which are critical for the interpretation of context relevant information. Section: “Scenario Analysis Examples” describes how the model may be applied to the analysis of a domestic scenario and the type of results obtained. Section: “Discussion and Future Work” concludes the paper with a broader discussion of the implications of the model and gives our directions for future work.

BACKGROUND

The increasing use of mobile interactive systems and pervasive and ubiquitous services available anywhere, anytime leads researchers and designers to question what is interaction and cooperation. Ubiquitous and pervasive applications have contributed to put the notion of incidental interaction [3] at the forefront of research on interaction and cooperation in “intelligent” environments. For example: a young child walks around in a kitchen (a potentially hazardous area) and incidentally this information is made visible to her/his parents in another part of the house; the music being played in a room adapts automatically when people within starts chatting; the heating system adjusts the thermostat to the occupants’ current activities.

One critical design issue now concerns what contextual factors need to be incorporated into these systems and services. The general idea is that in order to enable meaningful interactions between the context-aware environment and its occupants, the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

systems has to be aware of the occupants' context of activities. But what value do context sensitive services create and what is the induced cost (in terms of expense, loss of control, privacy and so on)?

Amongst the different approaches that may help to open the design space, we propose a framework based on the definition and simulation of different models of context processing. This approach to model-based evaluation uses a model of how an actor or a group of actor would behave in a particular environment in order to generate feedbacks by simulation of the model (be it a pen and pencil simulation or a computerized one).

Like traditional model-based evaluation in HCI the approach presented here can be seen as a valuable supplement to more traditional evaluation (usability testing, user-based evaluation,...). But the main practical objective of this type of models is to inform the design process. It is intended to be used as a tool for helping us to conduct analysis of individual and collective situated activities in domestic settings. Additionally, the tool could help the design team to assess different artefacts or different implementations of context-aware systems for assisting the collective organization of activities in the household (including energy, comfort and safety management).

The framework is grounded in the analysis of data drawn from the empirical study of several scenarios concerning various aspects of domestic life. One of these scenarios, concerning the management of lighting, is used in this paper to illustrate the approach.

GENERAL APPROACH

Identifying Relevant Key Factors

By analysing both the literature on empirical studies in domestic situations and different real-life scenarios, a set of relevant dimensions for defining a framework for analysing and modelling contextual activities has been defined [4]. The following identified dimensions have been included so far in the descriptive model:

- *Routines* as a resource for efficiently organising individual and collective activities at a low cognitive cost.
- The role of *artefacts* in the domestic situation. The interpretation of an artefact's state at a given moment determines the local context of use.
- The role of the *organisation of domestic space* is a contextualised way of organising activities.
- *Implicit communication* between actors in the physical environment. These communications may or may not be related to the actions of the actors.
- The *awareness* that an actor has of others activities defines the context of activity for that actor.
- The *dynamics of actor engagement*. The actors may need to manage different concerns in parallel.
- The *evaluation of an actor's availability* is an important aspect in defining the context for the actor and for other actors.

This analysis has yielded important insights about the dimensions of domestic activities that are directly related to contextual thinking issues.

Defining Different Levels of "Contextual Ability"

Environment, Context, Situation

Contextual abilities are partly determined by the capabilities, or the cognitive equipment, that an actor has to interpret the situation. An actor uses perceptive and cognitive resources in order to recognise different events, to give them a meaning and then perform an action or not. These capabilities vary depending on the actor, for example an elderly person or a young child does not have the same perceptive and cognitive capabilities as a young adult. This means that the viewpoints of the actors will be different and hence the ability to perform an action will be directly affected.

In order to take into account these differences of equipments and to apply this idea of "levels of contextual capability" in the context of designing context sensitive systems, we drew inspiration from a generic classification made by Quéré [5], who, based on Dewey's seminal work, identified three complementary categories of "contexts": environment, context and situation.

- *Environment*

An environment can be defined as a relatively stable structure composed of a location, and in which different objects are present. For example, we can speak of the kitchen as an environment defined by more or less precise physical boundaries and by the artefacts disseminated over this physical space.

- *Context*

The context is the wholeness that enables meaning to be given to an event (a behaviour, a signal in the environment, etc.) and that enables the justification of meaningful actions. Broadly speaking, context can be seen as an "interpreted environment".

- *Situation*

A situation can be seen as an environment "ordered" by the experience through time and space of this environment. This "ordering" is made possible by configurations, that is walkthroughs in the environment, paced by actions involving available resources.

Let us consider the following example:

"X intends to purchase an object O. She goes out of her place, but realizes that it is raining; she then goes back home in order to get an umbrella. In the stairs she meets one of her neighbors; she chats with him for some minutes. She then goes upstairs to her flat but has forgotten the reason why she came back. She goes out and notices a traffic jam in the street; she also realizes that she is late and therefore decides to take the subway."

We can see that X evolves dynamically in different environments (her place, the stairs, the street, etc.) that are populated by objects, e.g. other people, and in which different events may happen (rain, meeting someone, traffic jam, etc.).

These different environments are interpreted in terms of background knowledge and practices and in relation with the actor's current course of action (previous and future). This contextual set enables X to give meaning to the events that happen in the environment, and to generate relevant actions and, if needed, to justify them. For example, if there is traffic in the street, it is likely that it will be crowded in the whole area, so it is not a good idea to take the car as X is already late due to what has just happened; that is why X finds it more appropriate to take the subway.

This ordered experience in time and space, constituted by different episodes (expected and unexpected) that take place in a succession of environments, can be viewed as a “situation” (Figure 1), that is, a sequence of different contexts whose meaning or motive is given by the successive engagements and commitments of the actor, by the dynamics of her/his activity. Without this engagement, the sequence of contexts is just a set of unarticulated “contextual snapshots”.

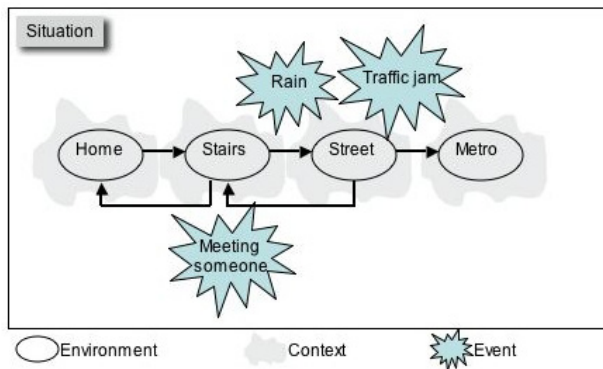


Figure 1. Environment, context and situation.

It is noticeable that a larger family of notions that stress the importance of the mutual constitutiveness of activity and context has emerged in the past years. For example the concept of “place” initially developed in the mid-nineties has recently gained interest in the HCI community. A place can be seen as a space invested with shared social meanings and expected behaviors; the space has to do with location even though the place has to do with activity. In a similar way, using the concept of situation is a mean (especially compared to the perspective stressed in the ethno-oriented studies) to re-introduce the actor’s point of view, that is the different manifestation of his/her cognition including his/her own private thinking. Concerning the underlying assumptions carried on when performing empirical analysis of activities in real settings, this kind of approach permits a switch in focus from approaches centered on the individual or on the collective toward approaches centered on the concept of situation [6].

From an analytical point of view, this three layers framework makes it possible to make an explicit conceptual distinction between the three, often confused, terms. Furthermore, it provides a basis for the integration of the aforementioned critical points identified in the literature and gives an analytical foundation for the examination of real world scenarios.

Technological Translation

In order to better fit to a transition from human to technological contextual capabilities we have interpreted the distinction between environment, context and situation as sensor, epistemic and historical equipment needed to design a context sensitive system.

▪ “Sensor” or “environmental” equipment

A human actor or a system can have an access to events in the world by means of different modalities (visual, auditory, olfactory, etc.). Unless he or she suffers a particular handicap, a human actor having the standard perceptive abilities would be able to capture a set of data in the environment in a predictable way. Similarly, some contextual data would be unavailable to an actor even with standard perceptive abilities: for example a human actor cannot detect a modification in the infrared spectra or a surge of power in the mains electricity supply.

▪ “Epistemic” equipment

Different kinds of knowledge may be required to correctly give meaning to a particular event. It may be general-purpose knowledge of the world in which the system or the agent moves. For example, by grouping sequences of actions into schemas we could plausibly infer an agent’s motives. It may be also specific knowledge about the local practices and routines i.e. the social context which gives meaning to and justifies the behaviours of the members of the group. For example, the usual organization of the household members’ activities during the preparation of the meal prescribes a mode of cooperative activity. This knowledge allows other human actors and smart systems to generate expectations about the actors’ possible actions.

▪ “Historical” equipment

The historical account of individual and collective experiences through time and space gives meaning to events in a particular environment. A specific previous episode of an agent can provide relevant clues to interpret the behaviour of this agent and to infer some aspects of his or her internal states. However, keeping a trace of all previous events is not enough; they must be organized in meaningful episodes.

Defining Different Points of View

Actors sharing the same physical environment may have different perceptions of the context. Similarly a context-aware device may have different levels of representation of the context of the occupants’ activities depending on the nature of its sensors, and the knowledge it can apply to interpret raw data acquired by sensors. To account for these different perceptions we introduced the notion of *point of view*, which describes the relevant context for an actor at any given moment. The idea is to compare the views of different actors at a given time and see to identify the consequences in terms of accuracy in the process of interpreting the situation. This may be used to inform the design process and, more precisely, to check to what extent a system equipped with specific contextual abilities (sensors and intelligent inferences) would be able to interpret a situation in a meaningful or at least useful way in order to act appropriately.

Three points of view were identified: *Actor*, *Analyst* and *System*.

The *Actor point of view* can be seen as a situated model of context when a person who is equipped with identified perceptive and cognitive abilities is engaged in a specific course of action. Different actors may interpret an action in different ways. Since the same context, as seen by an external observer, may vary for each singular actor it can be interesting to differentiate as many context viewpoints as actors present in the house. Most of the time the “context for the agent” remains an individual, situated experience of the world, which is only accessible by the agent him/herself. But the actor’s motive for performing an action may be publicly known if it has been made explicit, for example by a communicative act. However, very often there is no evidence for a motive or the internal states of an actor. The external observer therefore does not know the motive or has to infer it from perceptible manifest facts and background knowledge.

The *Analyst point of view* may take different faces. It can be the sum of all actors’ points of view (providing that the analyst has an access to them). When data on one or several actors are missing, the analyst must refer to available issues (manifest behaviour, location in the house,...). The *Analyst* point of view may also be seen as a hypothetical, idealistic viewpoint. We can ideally imagine an omniscient, ubiquitous observer who could have a non-restricted access to all the events and facts in the environment, including the motives of the different agents.

This kind of viewpoint is in some way close to, and may be seen as an enriched version of the “God’s view” used in simulation. The pragmatic interest of using such a viewpoint is that it provides a basis of reference to systematically compare the results of the application of different viewpoints to a theoretical optimum.

The *System point(s) of view* refers to the different levels of contextual equipment designers may consider at a given time of the design process. At a basic level (environmental equipment level) the system viewpoint describes the set of events that happen in the physical environment as raw sensor data. It may include physical events (e.g. door bell, microwave signal, phone ring), behavioural events (e.g. opening the refrigerator, entering the lounge). The environmental point of view concerns tracing the state of technological artefacts and the presence of identified people in the different part of the domestic space. The scope of environmental events that a system has access to depends on its perceptive equipment (sensors) and whether it can have access to the event at a time *t* or not.

A more sophisticated version of the system can add background knowledge (commonsense, local knowledge on the routines of the household,...) to the system equipment in order to interpret the behavior of the occupants or, broadly speaking, to build a picture of what is going on in the home at a particular moment. Meaningful contextual information can therefore be derived from the raw sensor data.

A step further, the designers may consider giving the ability to the context-sensitive system to keep a trace of the occupants’ commitments in order to build an historical representation of their activities. The underlying hypothesis being that this historical equipment may help the system to better interpret the motives of the human actors, and to better fit the occupants’ needs by automatically adapt its behaviour on context changes.

SCENARIO ANALYSIS EXAMPLES

Micro/Qualitative Analysis

The aim of this section is to describe how a simplified version of the framework may be used to analyse context and to show what additional information is provided by taking into account different levels of context.

The following excerpt of a real scenario comes from a large empirical study led in domestic situations. The chosen scenario concerns the implicit and explicit management of lighting by household occupants. For brevity we have extracted only a few minutes of the scenario:

The father (F) is engaged to sort some papers in the parents bedroom (18:43); the amount of light is too low, so he goes to the bathroom in order to turn on the light (18:44), then goes back to the parents bedroom (18:44:10). The mother (M) goes to the bathroom (18:47:00), takes something then goes to the bedroom and let the light on in the bathroom (18:48).

If we apply the viewpoint analytical framework, it is possible to describe the situation according to different points of view:

- The Actors’ point of view

According to the data drawn from verbal reports, the relevant contextual issues for the father can be summarized as follows:

- F is engaged in an activity that requires a comfortable level of light.
- The level of light in the bedroom is too low.

- Usual routine: turning on the light in the bathroom gives the extra light needed.

In order to understand correctly why F turned on the light in the bathroom while resting in the bedroom, one needs to know these different points or must be able to infer the third one rather than knowing (and recognizing) the routine. Understanding in real time the rational behind the decision to let the light on in the bathroom, or identifying a local routine may prevent a third part (member of the household or energy regulation device) to turn off the light. From a processing point of view the inferential process needs more resources than the routine identification process: some facts may require very specific situated knowledge about the household and its members.

If we consider now the mother’s point of view, the potential problem here is that the coexistence of contradictory contextual viewpoints might lead to a misunderstanding (the mother might consider for example that letting the light on in the bathroom is useless and energy-consuming) and then lead to an inadequate (from the father’s point of view) action (i.e. turning off the light). But the background knowledge associated to this context of activity (the mother knows that the father applies this routine when he works in the bedroom and understands the rational behind it) prevents her from doing so. This example put into evidence that it is not enough to share the same environmental context (being aware of the location of the actors and of the status of the artificial light in the rooms) with an actor in order to act in a sensible way. Indeed it is necessary to infer the actor’s underlying motives (which may require a complex chain of inferences based on the available facts in the environment and on general knowledge or common sense of the situation) or to identify the presence of a routine associated to a particular action or non-action. Here, sharing a common ground with the actor may prove very helpful: it enables to quickly recognize a situation in which a local routine is applied and gives relevant meaning to a collection of events in the physical environment.

- The Analyst’s point of view

In this example the analyst has an access to the points of view of F and M and can therefore give an informed account of the context and of the rational underlying the behaviour of the actors. Would the mother have switched off the light, he could have similarly given an explanation of the context by referring to the difference in the actors’ points of view and subsequently to the expectation breakdown between M and F.

- The System points of view

Let us consider different descriptions from the point of view of a sensor-based system (Table 1). The first one considers only the states of a pre-defined set of relevant artefacts (here the light in the different rooms).

Table 1. An example of sensor-based system point of view.

18:43:00	The father is in the bedroom
18:43:00	The light in the bedroom is on The light in the bathroom is off
18:44:00	The father is in the bathroom The father turns on the light The light in the bathroom is on
18:44:10	The father is in the bedroom
18:47:30	The mother is in the bathroom
18:48:00	The mother is in the bedroom

The sensor description listed above cannot enable a “simple” system to understand that the father intentionally switched on the light in the bathroom in order to work in the bedroom. A different level of description is needed to get a correct, meaningful picture of the situation. A “smart” device equipped with enough knowledge about daily events and individual and collective routines may correctly interpret the behavior of F and act in an appropriate way (in this example doing nothing...).

In the context of the application of this type of viewpoint for design purpose (assessment of different services associated to specific tools and interfaces, for example), it is possible to define a set of “context for the agent” viewpoints that cope with the cognitive equipment different versions of a system have been provided with (the actors’ internal states may be not considered, or the tacit rules that govern one aspect of the collective organization of the household may remain hidden).

Macro/Quantitative Analysis

The simple example discussed above put emphasis on the qualitative analysis of episodes of activities.

In order to better inform the process design, it may equally be useful to take into account some relevant criteria (energy savings, comfort, safety,...) in a more systematic way. To illustrate how the framework may be used this way, it was applied to a real-life scenario of domestic interaction concerning the collective management of lighting. The goal was to show how different contextual competences categories produce different results. The outputs in terms of energy consumption were compared to the reference scenario (real situation). These purely quantitative data are balanced by more qualitative considerations concerning the perceived comfort and the quality of the coordination between the actors.

The first criteria used to assess the effect of the two first levels of contextual ability (environmental and epistemic⁸) concerns the energy-saving management in the household during a particular period of time (from 6:30 pm to 8:30 am). More precisely we have identified the time during when the lights are

⁸ Historical capability was applied in a more complex scenario

left on in each room. The results from the reference scenario and the outputs from the « manual” simulation generated by the viewpoints 1 and 2 are shown in the Figure 2.

Results put into evidence that the first level of contextual capability and the associated actions induce a reduction of the time during which the light is switched on, and consequently some potential energy savings in most situations but the extreme case of managing the light in the bedrooms (Ctxt_sys-ENV-1). This later situation is in some way a bit artificial and must be imputed to the rather weak capabilities of the system (which in this version is only able to detect presence of actors in a place and the current binary status of the lights – on/off). It makes more sense here to consider that the light will be turned off in the bedroom by the occupants themselves before going to sleep (Ctxt_sys-ENV-2)). The gains are obvious in two types of cases: when a light has been left unintentionally on by a person (generally during a long duration when a light has been forgotten in a non visible area, and therefore remains on) or when a light has been left on intentionally during a short time; the rationale behind the decision being for example that the actor is expecting to come back quickly into the room, or that the light on may have a meaningful function even though there is nobody in the place.

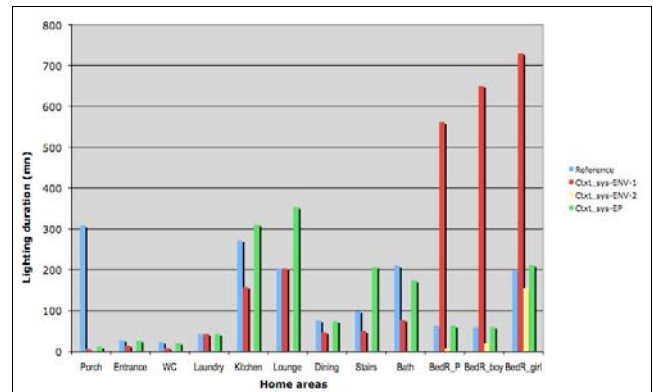


Figure 2. Results from the simulation of the light management scenario for two levels of context processing (environmental 1/2 and epistemic).

The second level of context processing adds to the “environmental” equipment, mainly based on simple sensors, a set of pre-defined general and local knowledge on the habits of the members of the household.

The results show that, with the energy savings criteria, in most of the cases the performance of the epistemic equipment is weaker than the performance of the environmental equipment. Additionally the results show that, in some cases, the performance of the system is weaker than what can be seen in the real scenario. In some cases it can be due to the fact that the system observes some local rules in a stricter manner. For example the light may be left on in a hazardous situation according to a usual tacit rule (for instance when young children are in the stairs with the light off). In some other cases this output of the manual simulation can be explained by the fact that the system applies a local routine or rule (a place is used in a continuous way during a period of time, and may therefore be left on). This kind of general principle may be directly associated with a « comfort » concern: during some particular moments associated with specific activities, the light will remain on in some a room in order to avoid repetitive

manipulations of light commands that may induce additional moves (when the commands are implanted in a non optimal manner for example) or unpleasant changes in ambient light.

DISCUSSION AND FUTURE WORK

Currently we have used the model in a manual mode to analyze the implications of different levels of context. Ultimately we envisage performing computer simulations of the scenarios. However, one problem is that the analysis of contextual issues goes well beyond what can be implemented in a simulation tool. It is unlikely that a computer based analysis tool would ever be able to replace totally a human analysis. Nevertheless, it may be interesting to think about how far a computer based simulation tool could go towards helping a human designer analyze the situation. One promising approach is to use a multi-agent based system and to draw upon the work in the domain of agent-based social simulation (ABSS). The objective is to simulate collective behavior in a multi-agent simulation with an explicit « physical » model of places, facilities, tools, and a socio-cognitive model including human abilities and general or local knowledge (for example, patterns of activities, routines and social rules that governs the organization of the household). One of the main advantages of ABSS is that it gives the opportunity to explore emergent sociocognitive phenomena, see different viewpoints and experiment with situations, which we might not be able to do in real life. ABSS is similar to an experimental methodology: simulation model can be set-up and executed many times, varying the conditions in which it runs and exploring the effects of different parameters [7]. ABSS may produce a set of data that may provide a heuristic basis and a design guidance for different actors involved in a design project. This approach has recently been demonstrated in different works focused on the notion of « participative simulation » which aims at exploring different organizational and technological issues⁹ with multi-agents platforms [8, 9].

For a framework to become a practical agenda in the design of context sensitive technologies it must articulate different criteria. In the example presented above we took into account two complementary dimensions: energy savings and comfort. But other criteria may be of critical importance when designing and assessing systems devoted to safety purpose such as the supervision of elderly occupants or young children. Here the problem of shared context [1, 10] between the occupants of the house and between the occupants and the smart systems become of crucial importance in order to avoid misunderstandings and breakdowns in coordination.

ACKNOWLEDGMENTS

This research has been funded by EDF (grant EDF/NEB-IRIT).

REFERENCES

- [1] Jackie Lee, C.-H., Bonanni, L., Espinosa, J., Lieberman, H. and Selker, T. 2006. Augmenting kitchen appliances with a shared context using knowledge about daily events. Proceedings of IUI'06 Conference, January 29 – February 1, Sydney, Australia, ACM.
- [2] Schmidt, A. 2005. Interactive Context-aware systems interacting with ambient intelligence. In: Riva, G., Valataro, F., Davide, F. and Alcaniz, M. (Eds.). *Ambient Intelligence*. Amsterdam: IOS Press. Pp. 159–178.
- [3] Dix, A. 2002. Managing the ecology of interaction. In: Proceedings of Tamodia 2002 – First International Workshop on Task Models and User Interface Design. Bucharest, Romania, 18–19 July.
- [4] Dugdale, J., Salembier, P., Fréjus, M. and Haradji, Y. 2005. An analysis of empirical studies of domestic activities. Descriptive framework and context modeling. Technical Report. EDF-IRIT. Clamart.
- [5] Quééré, L. 1997. La situation toujours négligée? *Réseaux* (85), 163–192.
- [6] Theureau, J. 2004. *Le cours d'action: Méthode élémentaire*. Toulouse: Octares (2nd Ed.).
- [7] Bellamine-Ben Saoud, N. T. Ben Mena, T. Dugdale, J., Pavard, B. and Ben Ahmed, M. 2006. Assessing large scale emergency rescue plans: an agent based Approach. Special Issue on Emergency Management Systems. *International Journal of Intelligent Control and Systems*, 11(4), Dec. 2006.
- [8] Guyot, P., Drogoul, A. and Lemaitre, C. 2005. Using emergence in participatory simulations to design multi-agent systems. In: AAMAS'05, July 25–29. Utrecht, Netherlands: ACM.
- [9] Ramanath, A. M. and Gilbert, N. 2004. The design of participatory agent-based social simulations. *Journal of Artificial Societies and Social Simulation*, 7(4).
- [10] Salembier, P., Kahn, J., Zorola-Villarreal, R. and Zouinar, M. 1997. Assessing methods: Cognitive modeling. WP6 RHEA Research Report. CEC, DG VII.

⁹ See for example the Simweb project (<http://www.simdigital.com>)

Analysis of Activity in Domestic Settings for the Design Ubiquitous Technologies

Germain Poizat
Université de Bourgogne
Campus Universitaire,
B.P. 27877
21078 DIJON Cedex
germain.poizat@u-
bourgogne.fr

Myriam Fréjus
EDF Recherche &
Développement
1 Avenue du Général
de Gaulle
92141 CLAMART Cedex
myriam.fréjus@edf.fr

Yvon Haradji
EDF Recherche &
Développement
1 Avenue du Général
de Gaulle
92141 CLAMART Cedex
yvon.haradji@edf.fr

ABSTRACT

The aim of this study was to analyzing and modeling domestic activities (i.e. what people do in the home). The ultimate goal is to help to design ubiquitous technologies which are context sensitive and which really fit with our needs.

Keywords

activity-oriented approach, activity-centred design, modeling domestic activities, home, ubiquitous computing, course of action

INTRODUCTION

The design of ambient and ubiquitous technologies for the domestic environment has become an active area of research and generating a growing body of industrial initiatives [e.g., 3, 4, 7] Most of those works address to the design of future smart homes. Various computer-based interactive systems are developed specifically to be used in domestic settings. However, the design of such systems requires a model of domestic context which call for a better understanding of the everyday nature of domestic activities as they emerge in situ. Different ethnographic studies contribute to the exploration of domestic practices, technological potential of Ubiquitous Computing, and way in which it might be integrated into domestic setting [e.g., 1, 5]. Our study aims to improve, through an activity-oriented approach, our understanding of collective activities in domestic settings. This deep understanding of householders' activity will be placed later at the centre of design process. To reconstruct the dynamics of domestic activity, this study was conducted in reference to the course of action empirical research program [6]. According to the course of action theory, (a) every activity is situated, meaning that it cannot be dissociated from the context in which it takes shape, and (b) the interactions between actors and environments are an asymmetric coupling.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

METHOD

Participants

Five households agreed to participate in the study. The composition of each household is as follows. Household 1: one couple with two children (3 yrs, 9 yrs); Household 2: one couple with three children (4 yrs, 14 yrs, 16 yrs); Household 3: one couple with three children (9 yrs, 19 yrs, 22 yrs); Household 4: one couple with one daughter (2yrs); Household 5: one couple with four children (13 yrs, 15 yrs, 16 yrs, 18 yrs).

Data Collection

Two types of data were gathered: (a) continuous video recordings of the participants' actions and (b) verbalizations during post interviews. Domestic activities were recorded with digital cameras six hours a day (continuously) during two week-days, and on weekends for each household. We looked for recording equipment that was not too intrusive (e.g., micro-cameras, tiny microphone, pocket digital recorders), did not require the presence of an observed during videotaping nor manipulation by the participant. Householders could also switch off temporarily the recording equipment with an ordinary interrupter. With their agreement, we recorded activities that take place in the following places: kitchen, lounge, office, bedroom, corridors, and hall. The verbalization data were gathered from collective self-confrontation interviews with the householders (i.e., adults, teenagers).

Data Processing

The data are processed to grasp the meaning each householder gives to his or her own activity. First, the householders' actions during the situation are described and self-confrontation data are transcribed verbatim. Second, Elementary Units of Meaning (EUMs) are identified. An EUM is a fraction of the activity that is meaningful to the actor in continuous flow of his or her activity. This discrete units, or EUMs, may be any physical actions, communication exchanges, interpretations, or feelings that can be shown, told, and commented one by one by each householder. Third, meaningful structures are identified for each householder by characterising the relations of sequencing and embedding between the EUMs. Fourth, the householders' EUMs and meaningful structures were synchronized on the basis of the video recordings and verbalization transcripts. Finally, all householders' EUMs and meaningful structures were compared at the same point in time in order to determine the coordination between them.

RESULTS & DISCUSSION

Our results showed different characteristics of collective activities in domestic settings with implications for the design of ubiquitous technologies. First, householders have many preoccupations, the status of which changes dynamically. Second, the domestic activity is a multi-scaling activity with macro and micro coordination between householders. Third, the domestic activity rest significantly on dyadic coordination and potential networks. Finally, our results pointed that the same task or behavior can have different meanings (e.g., close shutters) which may themselves be subject to negotiation between householders. The analysis of the domestic activity allowed identifying different design proposals to explore. It seems important to design a computer-based system with the ability to: (a) manage interruptions in relation with the task priority (from the point view of the actor) and how it may change dynamically according to the situation, (c) support the dynamics of action, reasoning, and interaction of the householders at different temporal scales, and (c) handle the ambiguous context [2].

REFERENCES

- [1] Crabtree, A., Rodden, T., Hemmings, T. and Benford, S. 2003. Finding a place for UbiComp in the home. In: Proceedings of the 5th International Conference on Ubiquitous Computing. Seattle, WA, October 12–15, 2003. UbiComp'03. Springer, Berlin. Pp. 208–226.
- [2] Dey, A., Mankoff, J., Abowd, G. and Carter, S. 2002. Distributed mediation of ambiguous context in aware environments. In: Proceedings of the ACM Symposium on User Interface Software and Technology. Paris, October 27–30, 2002. UIST'02. CHI Letters, 4. Pp. 121–130.
- [3] Edwards, W. K. and Grinter, R. E. 2001. At home with Ubiquitous Computing: Seven challenges. In: Proceedings of the 3rd International Conference on Ubiquitous Computing. Atlanta, GA, September/October 30–2, 2001. UbiComp'01. Springer-Verlag, London. Pp. 256–272.
- [4] Howard, S., Kjeldskov, J. and Skov, M. 2007. Pervasive computing in the domestic space. *Personal and Ubiquitous Computing*, 11, 329–333.
- [5] Rodden, T., Crabtree, A., Hemmings, T., Koleva, B., Humble, J., Akesson, K. and Hansson, P. 2004. Configuring the Ubiquitous Home. In: Proceedings of the 6th International Conference on Designing Cooperative Systems. Hyères, France, May 11–14, 2004. COOP'04. IOS Press, Amsterdam. Pp. 227–241.
- [6] Theureau, J. 2003. Course of action analysis & course of action centered design. In: Handbook of cognitive task design. Hollnagel, E. (Ed.). Lawrence Erlbaum, Mahwah, NJ. Pp. 55–81.
- [7] Zouinar, M., La Valle, N. and Relieu, M. 2005. Video observation and analysis of everyday life in homes. In: Proceedings of Interior Insights: Design, ethnography, and the home. London, November 24–25, 2005.

Food for Talk: Photo Frames to Support Social Connectedness for Elderly People in a Nursing Home

Margit Biemans
Novay
P.O. Box 589 Enschede
The Netherlands
Margit.Biemans@novay.nl

Betsy van Dijk
University of Twente
P.O. Box 217 Enschede
The Netherlands
bvdijk@ewi.utwente.nl

ABSTRACT

Social connectedness is crucial to someone's well-being. A case study is conducted to test whether the social connectedness of elderly people living in a nursing home and their family and friends can be improved through a photo frame. A SIM-based photo frame is used to keep the elderly people informed about the comings and goings of their loved ones. Eight elderly people living in a nursing home participated in this case study for 6–7 weeks. A content analysis of the photos revealed that the photos often were related to special events or holidays that happened in the past. Interviews indicated that the photos mainly served as food for talk, i.e. the photos initiated conversations between the elderly people mutually, with their family members and with the healthcare professionals. They all liked the photo frame and it didn't serve as a means to exchange news, but as a catalyst to talk – mainly – about the past.

Keywords

social connectedness, photo sharing, user evaluation

ACM Classification Keywords

H.5.1 [Information Systems]: Information interfaces and presentation – *multimedia information systems, evaluation/methodology*.

INTRODUCTION

Social connectedness is of crucial importance for someone's health and well-being (amongst others [1], [2]). Social connectedness affects one's happiness and contentment [3], and contributes to one's social and psychological well-being [4] and to general health or disease [5]. It also provides a buffer against stress, enhances coping skills [6], and leads to higher levels of life satisfaction and self-esteem [7]. Good social supportive networks have a positive effect on personal well-being and health [8]. Low connected people view their surroundings as negative, hostile, threatening, and unfriendly [9].

When people's physical or mental conditions degrade to the extent that they are forced to leave their home and have to

move to a nursing home, people start facing a new lifestyle, and face the threat of losing the bond with their immediate family and friends, which might lead to frustration and loneliness ([10], [11]). Currently most of the social contacts in nursing homes are mainly visits of family and friends and based on explicit communication (scheduled visits and phone calls). Hardly any (technological) means are available or used to share spontaneous moments or exchange affections. For these people it is challenging to continue maintaining their roles as parent, partner, or friend. Davies & Nolan [11] found that in such circumstances, well-being is underdeveloped at the moment; ICT can play a valuable role to improve the feelings of connectedness between these families, and thus have a positive effect on well-being.

We want to maintain and enhance social connectedness for people who stay in a nursing home, and for their immediate social network. Our aim is not to develop new technology, but to use available technology for a new purpose. Therefore, we provide the elderly people in the nursing home with a SIM-based photo frame to which family members can send photos, to keep their elderly relative informed of and involved with the small things in life.

Therefore, related work on social connectedness and digital photo sharing in nursing homes is described in the background session. Moreover, photo categorization is discussed as a way to analyze the photos for social connectedness. This results in the design of a case study. Eight elderly people used a digital photo frame for about 6–7 weeks. During that period, family members were able to send them photos. Interviews, questionnaires and a content analysis of the photos sent are used to evaluate the effect of photo sharing on social connectedness.

BACKGROUND

Being forced to move to a nursing home is a life event for both the elderly person involved and her family. The lives of people are brutally separated and people have to deal with this. They all face the threat of losing the bond with their immediate family, and experience frustration and loneliness. Kellett [12] identified four themes for families to pay attention to in maintaining a sense of attachment 1) engagement involvement –to create new ways of caring, 2) worth – ensuring their specialized knowledge of the elderly person is used for care, 3) concern – negotiate the boundaries between themselves and the staff, and 4) continuity – remaining involved and continuing to share a fruitful relation with the older person. The last theme aims at social connectedness, and is relevant for family and elderly persons to keep attached. In addition, the children of the elderly person in the nursing home provide an important link between the parent and the wider society [13], and especially grandchildren [10]. Other reasons for family members to improve the social connectedness are to enrich and enhance the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

life of their parent. Positive peer relations and positive family relations are additional aspects, contributing to thriving of the elderly people in nursing homes [14].

Photos and the Social Use of Cameras

The photographs of loved ones are typically symbolic of a personal relationship. They provide a constant reminder of the emotional feelings involved at that particular moment of time when the photograph was taken. Photo frames are typically used to display these photos and highlight the value and importance of these static images. Photo frames are considered to be part of furniture or decorative objects that blend in the home environment.

Recently, people are more and more using camera phones to capture images of life. In literature, taxonomies are available of why people capture such images. According to the taxonomy of Kindberg et al. [15] the intentions of people vary along two dimensions: *social versus individual* intentions and *affective versus functional* intentions. For studying social connectedness, the social part of this taxonomy is relevant. Of this part of the taxonomy, two aspects relate to a situation where people are remotely located, like in our context. The first is an affective category with photos intended for communication with absent family or friends. The second is a functional category with photos to support a task by sharing with remote family or friends (e.g. the photo of a choice of sandals is sent, and immediately after sending the selection of sandals is discussed by phone).

In a follow-up study Kindberg et al. [16] examined the images that belong to precisely these two categories. This study showed a variety of ways in which camera phones are used to connect people. They used timelines and the aspect of common ground to study and discuss the images in these categories. Examples of the category 'absent family or friends' were sent at the moment of capturing in almost 30% of the time, in order to extend an experience to absent friends (e.g. reaching the top of a mountain) or to illustrate a shared history (e.g. familiar food just made). The photos sent later typically involve telling a story. In the functional category 'remote task' about half of the photos were delivered at the time of capture, hence timely delivery seems to be more important here. Several of the photos in this category were used to convey news. Others were used for discussion later.

Photo Categorization

Research on the interpretation and inference of the meaning from photos is not exhaustive. We are particularly interested in the role of photo sharing in social connectedness, i.e. what can the sharing of photos tell us about social connectedness between the sender and receiver.

Nominal ways of categorizing photos are the common or default photo category lists that are used in literature and photo applications. These types of categories include lists such as vacations, special events, family and friends, art and fun [17] or the categories of Garau et al. [18] that are partly similar: travel, context (e.g. location, activity), portraits of people or animals, visual interests (e.g. landscapes, art), humor, media, events (mundane or special). A taxonomy for consumer photos which expands from the types of indoor and outdoor photos has also been created [19].

Lastly, by capturing and storing contextual data of the photos, one could categorize them based on contextual metadata. Naaman et al. [20] generated contextual metadata categories that are most useful for recalling and finding photographs. These include

location, time of day, light status, weather status and temperature, and events. They showed that the automatic collection of metadata is helpful in the organization of photo collections.

Towards the Research Question

Related work shows that social connectedness is recognized as an important aspect of the well-being of elderly people in nursing homes, particularly the social connectedness between the parent in the nursery home and her children and grandchildren. This connectedness not only relates to aspects of the life of the (grand)children, but also to keep in touch with the wider society. Therefore, this research aims to improve the social connectedness between the elderly parent in a nursery home and her children, and with the rest of the world. Therefore, the following research question is posed: "Does sharing of everyday things through photo frames have a positive effect on social connectedness between elderly people and their family?"

CASE STUDY DESIGN

Wireless Internet is not available in the nursery home, so the only possibility was to use SIM-based photo frames. The Vodafone™ 520 photo frame with integrated SIM card was selected. This enables users to send photos through Multimedia Messaging Service (MMS) as well as through a –personal password protected– Website. MMS provides opportunities for sending time related photos; sharing during the moment. The Website also provides opportunities to add text to photos.

Eight elderly people (one male, seven female) and their family members participated in this case-study. The elderly participants lived on one long stay department of a nursery home. Their ages varied between 50 and 89 (one person of 50 years old, the others on average 80 years old). They slept in rooms with up to three people, and during the day they all sat around tables in their communal living room. All eight people got a SIM-based photo frame, which were positioned on the table in front of their personal place. In this way, the elderly participants were not only able to see their own frame with photos, but also the frames of others.

During a regular visit, one family member of each elderly person received a Nokia 5000 mobile phone with a 1.3 MP camera and a 25 euro pre-paid SIM. Other family members could use their own mobile phone with camera or Internet to send photos. Some elderly people have nine children who participated in this study, while others have only one child involved.

As the elderly people spent most of the day in the communal living room, each photo frames was positioned on their personal place on the table. As a consequence, several photo frames stood on the same table, all positioned in a slightly different direction. The frames were positioned in the middle of the table, as the cords were rather short. Some of the elderly persons were able to touch and handle their own frame, i.e., delete, rotate and skip photos, others were too impaired to do this.

The case study started half December and ended at the second half of January (about 6 weeks). This is the Christmas and New Years period, a period of bank holidays and many events with family and friends, and also a time in which attention is paid to friendship, warmth and connectedness.

The effect of receiving photos on the feeling of social connectedness is measured in several ways, i.e., data triangulation [14]. The applied methods were structured interviews, quantitative analysis of the photos sent and a content analysis of these photos. The set-up and the results of the content analysis are described in a

separate section. The interviews were conducted at three moments; pre-trial, during the trial, and post-trial.

Ideally, the interviews should have taken place separately with the elderly people and their family members. However, some of the elderly people were not able to communicate by speech because they suffer from aphasia. Others were very emotional because of their cognitive decline. Therefore, the interviews were conducted by the elderly participants and a family member together.

The pre-trial interviews aimed at identifying a kind of baseline of the amount and ways of social contacts and the feelings of social connectedness of both the elderly persons and their family members.

Half way, the elderly person was visited and the health care professionals were asked how things are going, what their first opinions are about the photo frame, how they experience it, and whether they have any questions about using it. Contact with the family went through email. They could ask questions and help was provided to use the Website and the phone. When no photos were sent after one week, assistance was offered to the family.

During the post-trial interview, questions were posed to identify whether the use of the photo frame has effected or changed something in the quantitative and qualitative aspects of social connectedness of both the elderly persons and their relatives. Moreover, the post-trial interview also focused upon the use, usability and possible extensions of the photo frame.

RESULTS

In order to answer the research question: “Does sharing of everyday things through photo frames have a positive effect on social connectedness between elderly people and their family” the results are described in three sections. First a quantitative analysis is performed on the photos sent, subsequently the user experiences of the elderly and their family members are described, and finally the content of the photos sent is analyzed in relation to social connectedness.

The Photos Sent

For each of the elderly, the amount of photos sent over time is counted. Moreover, the way the photo was sent to the photo frame was also analyzed (MMS or Internet). From the Internet photos it was not possible to see who uploaded the photos (generic account). Photos sent by MMS could be traced by the phone number. Each family got one mobile phone with camera and a pre-paid SIM. Of course, other mobile phones could also be used to MMS photos to the frame. However, none of the family members sent a MMS from another phone number. Figure 1 provides a graphical overview of the amount of pictures sent per person, divided by MMS and Internet.

For each of the elderly participants, most photos were uploaded by Internet. The main reason for this was that family members found MMS difficult to use (no one used it with his own phone), they did not like the quality of the photos of the 1.3MP camera, and many of the photos sent were related to events that happened a while ago (for more details, see content analysis section). Remarkably, on 10% of the photos, the elderly person was available on the photo. For one person, this was even 50% of the photos.

Adding text was only possible by Internet uploads. In total, 20% of the photos contained text. This also differed a lot between the elderly persons (range 1–75%). Sometimes, text was added at the beginning of a series of photos, e.g. when

referring to a holiday “Scotland 1997” or a special event “Baptize of Henry”.

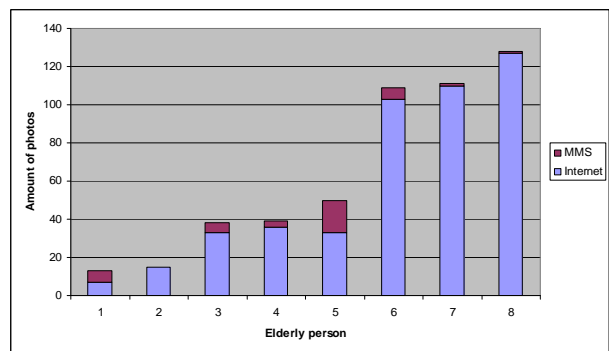


Figure 1. Overview of amount of photos sent per person.

We were also interested in the division of photos sent over time: do people send photos regularly, or all at once, or only in the beginning? When analyzing these aspects, it should be mentioned that not all family members started the trial on the same date. Some family members live far abroad and were not able to visit their mother or father earlier. Others encountered problems with MMS or uploading photos through the Internet and indicated that only after a week. Therefore, the starting date of each elderly participant differed. The first date someone sent pictures to the photo frame was taken as the start of the trial. In this way, some have only been involved for two or four weeks, and others for seven weeks. Figure 2 provides an overview of the amount of photos sent over time.

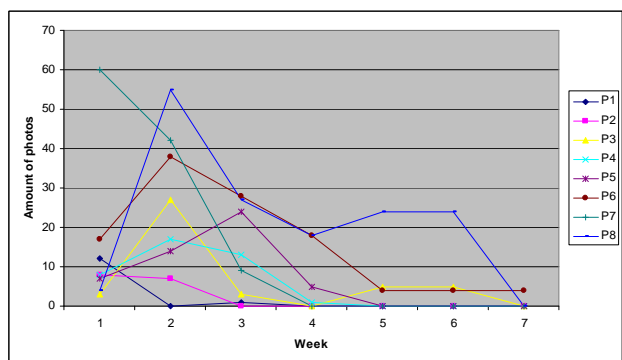


Figure 2. Amount of photos sent over time.

Figure 2 shows that sending or uploading photos takes place in batches. The peaks can be explained by family members uploading many “old” photos from previous holidays or special events via Internet. Often, many photos of the same holiday or Christmas diner were uploaded. Figure 2 also indicates that after a couple of weeks the amount of photos sent diminished. The freshness effect of the photo frame reduces, or many of the relevant photos are already sent. Another explanation is that four of the families started some weeks later with the trial, due to practical circumstances. As the counting of the weeks started at the moment the first photo was sent within one family, for this family, the trial took only 4–5 weeks instead of the 7 weeks on the X-axis of the figure.

User Experiences of Social Connectedness; Food for Talk

At the start of the trial, interviews were conducted with the elderly person and one family member together. Family members were enthusiastic of using photos to keeping in contact, especially because four of them have children or grandchildren living in a foreign country. Other family members – living in the vicinity – indicated to use the photos to keep the elderly person involved in their lives, and the lives of their (small) children. Two family members indicated that this might be a new medium to communicate with their mother who was not able to speak due to aphasia.

Also concerns were mentioned during these interviews. One elderly person was reluctant to use the photo frame because he does not like new technology. The family of another person indicated that they were afraid of the effort it would take; they are already busy with managing many aspects related to the elderly person. However, the elderly person was so enthusiastic that the family was convinced to try it.

After the trial, all the families indicated that the photos served as food for talk! The elderly participants were all very proud about the photos they received, regardless the numbers of photos. Sometimes they even got very emotional – in a positive way – about the photos. The photos not only served as food for talk between the elderly people and their family, but also between all the elderly people in the nursing home, between the elderly people and visiting family members of other elderly people (“*please look at my pictures?*”), and between the elderly people and the health care professionals. Moreover, “*it provides some diversion to my mother during the long days of sitting in a chair and reading books*”. The elderly people who are able to talk elaborated on their photos (to every one). The elderly people who are not able to speak pinpointed on the photo what they meant, and drew attention to it. The family members of the elderly people with aphasia experienced the photo frame as an enriched communication mean “*now, my mother sees things instead of only hearing*”. The elderly people were now able to see the story at the time things were happening, instead of only hearing the stories afterwards. And, it is really true “*A picture tells more than a 1000 words*”, “*now my mother has a new window to the world instead of watching the walls of the nursing home*”. In short, the elderly people really enjoyed the photos. The main effect of the photos was not as intended initially: to provide the elderly people with information of everyday life of their family members. Their main effect was that they served as food to talk!

The family members indicated that they had sent the photos with certain goals. Some families indicated to have sent recognizable photos for the elderly person, e.g. birthdays, Christmas diner, special event, etc. While other families indicated only to have sent photos of non-confronting events, i.e. no photos of events the elderly person cannot attend anymore, but photos of just normal everyday aspects like the children playing, the dog, etc. Others indicated to having sent photos to keep in touch during their business trips, and others just have sent photos the elderly person liked, and photos that made them smile!

Usability Aspects of the Photo Frame

During the trial, both the family members and the elderly participants encountered problems using the photo frame, the Website and the mobile phone with camera. The most mentioned item of the photo frame was its stand. The stand was unstable and the frame fell down many times. Especially when

the elderly person used or tried to use the buttons on the back of the frame to manage the photos, i.e., delete, rotate, skip, stand still, and text view. For some elderly persons, the buttons were too small or illegible, so they would prefer a remote control. Many people complained about the position of the photos on the frame, they were sometimes rotated, and turning them did not continue to have the desired effect during the next round the photos appeared on the screen. One family member indicated that he would like to have a USB port on the frame for uploading photos (SD port is available).

Many family members indicated that the Website did not work properly all the time. When uploading photos, many times an error message was shown, while nothing seemed to be wrong with the photos. Refreshing the page was a quick and dirty solution to solve this problem.

Most family members found the trial period too short, especially to start using the mobile phone with camera. They were not familiar with MMS, and wanted to read the manual before using it. However, they indicated that they intend to keep using it when they are abroad and phoning is very expensive, or when communication by phone is impossible because the elderly person is unable to speak.

Additional aspects mentioned as an added value for the photo frame were: a battery instead of the relatively short cord. Also a remote control was mentioned as a welcome extension for the current population. However, it is the question whether such a control is wanted in a normal living room. The most valued aspect was to make the photo frame an interactive device; a response could be given to a (new) photo. For example, sending a receipt, sending a thank you, sending a message like “nice photo”, or another response

CONTENT ANALYSIS OF PHOTOS

To analyze the photos that were sent a categorization was used that focuses on intentions of family and friends who took and sent photos to the person in the nursing home. This way aims to get categories that are meaningful for social connectedness. A similar categorization has been used before to analyze the photos sent by family and friends to people that temporarily (6–12 months) had to stay at the spinal cord lesion department of a rehabilitation centre [22]. The categorization is explained in more detail in [22]. The categorization used in this study is:

1. Message: “I tell you something with this photo or I will show you something new”
2. Greetings: “ I want to say hi to you”
3. Everyday life: “I want to keep you involved in the normal things and regular events in and around the house”
4. Special events: “I want to inform you about a special event”
5. Something funny or aesthetic: “I want to tell you that I think of you or I want to cheer you up”.

The first category contains photos that are meant for notification or discussion. Examples are photos of new things in and around the house. Photos in the second category show people greeting. A single portrait photo of a person or the first portrait photo in a sequence of photos sent within a short time-frame is categorized as a greeting. The photos in the third category are meant to keep the person in the nursing home informed about normal things in and around the house of their family. They typically contain photos of children, gardens and animals and photos of regular events (go to school/work, walk the dog). The photos of special events are meant to keep the person in the nursing home informed about and involved in special moments

(holidays, trips, parties). This category is similar to one used by Okabe [23]: photos of events that are considered noteworthy. According to Okabe [23] photos in this category are likely to become the topic of conversation among family or friends. The photos in the last category show something funny or aesthetic or a photo of something the elderly person likes. They are sent with affective intentions, to cheer the receiver up and at the same time to tell him or her “I think of you and make some effort for you”.

This categorization is used to analyze the photos sent to the eight elderly people living in the nursing home. The elderly participants and their relatives and friends were not asked to do the categorization themselves. That would have been impossible for most of the elderly persons and too much of a burden for their family, especially when living far away or abroad. Instead, the authors of this paper independently categorized the photos. This resulted in two categorizations of the 450 photos, with an inter-rater reliability score of 0.56 (Cohen’s Kappa). According to Landis and Koch [24], this is a moderate agreement.

We analyzed the differences between the ratings. The main differences were caused by lack of knowledge of the location and situation where the photo was taken (cf. [22]). In this study an extra difficulty was the time dimension. It appeared that many people sent photos from events (birthdays, holidays, and wedding anniversaries) in the past. Many of these differences could be resolved by using texts that were sent along with the photos and by using knowledge from interviews that were held with the elderly persons and their relatives who were involved in the sending of photos.

One difference between the ratings occurred repeatedly. It concerned photos of people in a boat. One rater thought this was a normal thing of everyday life; the other rater thought this was a special event. This difference probably reflects a difference between the lifestyle of the raters: one of them owns a boat and lives there quite often during the weekends, the other one hardly ever gets on a boat. Because the photos of the boat reoccurred on the photo frame of this person we decided that for the sender of the photo this was part of everyday life (category 3).

After the corrections the inter-rater reliability was 0.87; “almost perfect agreement” according to Landis and Koch [24]. The percentage figures presented below are from the corrected categorizations. The photos with remaining differences between raters were put in an extra category labeled “unknown”. The percentages of photos in the various categories (Figure 3) and the contents of each of the categories will be described in more detail below.

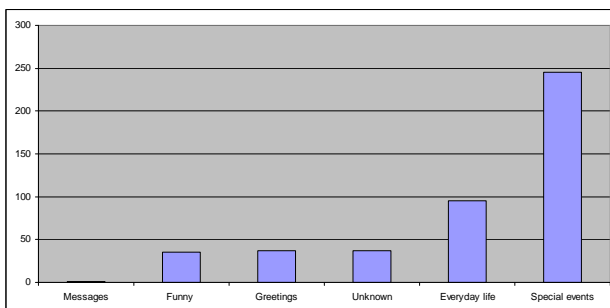


Figure 3. Overview of percentages of photos per category.

Messages: This category only contains one photo (out of 450; 0.2%): a happy new year wish written on a photo with display of fireworks. We anticipated that this category would be

smaller than for people who only left home temporarily but that it would only be one photo surprised us.

Greetings: The percentage of photos in the category greetings is 8.2. Often these are photos of close relatives of the elderly person: their children and grown-up grandchildren. Only occasionally their (late) husband was on the photo, though most of the women were widows. If other people were on the photos they were often accompanied by a child of the person that received the photo.

Everyday life: Only 21.1% of the photos were taken in the home environment of the family of the elderly or during regular events. Examples are photos of dogs, young children playing, people feeding their young children, the grandchildren or grand-grandchildren, relatives preparing and eating a meal. Also a small number of photos of regular events like people ice skating and a few photos with the elderly receiver on it together with relatives or friends belong to this category. Though second in size this category is surprisingly small considering it comprises all photos of small children and animals and people in their normal environment.

Special events: By far the biggest category. 54.4% of the photos are of special events. Most of these photos (65.7%) were taken many years ago and show holidays of the elderly people, a trip to Lourdes, wedding anniversaries, and Christmas dinners.

Funny or aesthetic photos: The category with funny or aesthetic photos contained 7.8% of the photos. In this category we find many photos of winter sceneries (it had been snowing and freezing during the period the photo frames were in the nursing home). Furthermore a few photos of nice old buildings, all without people on it.

Category unknown/undecided: In total 37 out of 450 (8.2%) of the photos were rated differently by the raters and the difference could not be resolved by the texts or the interviews. We analyzed these photos separately. In most of these cases it was not clear if the photo was old or recent. Furthermore, sometimes the location of the photo taken was not clear; old holiday photo, or recent photo of family living in a foreign country.

DISCUSSION AND CONCLUSION

Photos – sent by family members to a photo frame of elderly people in a nursing home – positively affect their feelings of social connectedness. Based upon literature (amongst others [13], [15], [16]), the postulation was made that especially sharing the small things of everyday life would contribute to social connectedness. However, the results of this case study revealed that sharing everyday things was not the most important effect of the photo frame. The frame was mainly used to send photos of special events that are meaningful to the elderly persons. These kinds of photos serve as food to talk between the elderly people, between the elderly people and family members and healthcare professionals. This is in line with Davies and Nolan [11] who studied the experiences of relatives of establishing a new role within a nursery home “contributing to community” through interacting with other residents, relatives and staff, taking part in social events and generally providing a link with the outside world. First, this discussion will focus upon various reasons for this.

Sharing the everyday small things of life is normally not part of the relation elderly people (around 80 years old) have with their children. These children are adults and have their own life and share these things with their direct social context (cf. [22]). In normal life, children might share special events (e.g. birthdays and holidays) with their parents. Moreover, the small

things of everyday life (e.g. typical family food, new shoes bought, growth of seeded flowers in the garden) might not be understood by the elderly persons, because they did not take part in these activities at home any more. Especially, as moving to a nursery home is often not a sudden moment, but the end of a (long) period of mental and/or physical degradation.

Two of the (younger) participants of our case study recently moved to the nursery home because of the disease Multiple Sclerosis. Their family members explicitly stated that they sent only “neutral” photos to their parents because they did not want to confront them with the typical small things in life that they miss at the moment, and probably have to miss for the rest of their lives. They do not want to evoke negative emotions with their photos.

One of the reasons to choose for the photo frame in this case study was its MMS possibility. Family members were able to send timely related photos by MMS. However, MMS was not used very much. Family members indicated that they had usability problems using it, many 40+ people do not have camera phones [22], and a mobile phone is normally a personal device. On the one hand, these issues might be the reason for not using MMS. On the other hand, family members indicated that they wanted to send “old” photos of special events or holidays, to evoke positive emotions by the elderly person. In the pre-digital photo age, other kinds of photos were taken.

Also the type of year might have influenced the results of the case study. The trial took place in December and January, a period of family events and typical weather circumstances. For the first time in 11 years, there was snow and ice. The focus of many photos was on these rather rare circumstances. Decades ago there was snow and ice every year, so it appeals to the “old winter feelings” of the elderly people. Therefore during the trial, less attention or time might have been spent to send normal photos.

The amount of photos sent differs a lot between the elderly participants of the case study (13–152). This difference did not affect the pride and joy of the participants to show and talk about their photos. The persons who had received many photos often received great batches of photos of one event (13–44 per event). The small amount of photos other persons received were not about a single event, but related to their (grand)children and pets. So, elderly persons who receive many photos, often receive more photos of the same kind of same event.

As the elderly people, their family members and the healthcare professionals were very positive about the use of the photo frame in the nursery home, the trial is extended with 5 months. In this way, all people involved can decide whether to introduce this as a new service, the elderly people and their family members can experience whether it is still useful in the long run. Moreover, it provides the researchers an opportunity to study the effects of photo sharing in a longer trial. In this way, the freshness effect of the photo frame, the time of year and the learning effect of MMS can be studied in more detail.

So, in general photo sharing has a positive effect on the social connectedness of elderly people in nursery homes and their family members. Most of the photos sent relate to special events that happened in the past and evoke positive emotions by the elderly persons. The photos sent serve as food to talk for the elderly persons, their family members and the health care professionals.

REFERENCES

- [1] Sadlo, M. 2005. Effects of Communication Mode on Connectedness and Subjective Well-Being. Thesis, Australian Center of Quality of life.
- [2] The Social Report. 2005. <http://www.socialreport.ms.govt.nz/2005/index.html>.
- [3] Spellerberg, A. 2001. Framework for the measurement of social capital in New Zealand. Research and Analytical Report 2001#14. Wellington, Statistics New Zealand.
- [4] Baumeister, R. and Leary, M. 1995. The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological Bulletin*, 117(3), 497–529.
- [5] Berger-Schmitt, R. and Noll, H. 2000. Conceptual Framework and Structure of a European System of Social Indicators. EuReporting Working Paper No. 9. Mannheim: ZUMA.
- [6] Cohen, S. and Wills, T. A. 1985. Stress, social support and the buffering hypothesis. *Psychological Bulletin*, 98, 310–357.
- [7] Takahashi, K., Tamura, J. and Tokoro, M. 1997. Patterns of Social Relationships and Psychological Well-being among the Elderly. *Journal of behavioral development*, 14, 417–430.
- [8] Henly, J., Danziger S. and Offer, S. 2005. The Contribution of Social Support to the Material Well-being of Welfare Recipients. *Journal of Marriage and Family*, 67.
- [9] Lee, R. M. and Robbins, S. B. 1998. The relationship between social connectedness and anxiety, self-esteem, and social identity. *Journal of counseling psychology*, 45, 338–345.
- [10] Sandberg, J., Lundh, U. and Nolan, M. 2002. Moving into a care home: the role of adult children in the placement process. *International Journal of Nursing Studies*, 39, 353–362.
- [11] Davies, S. and Nolan, M. 2006. Making it better: Self-perceived roles of family caregivers of older people living in care homes: A qualitative study. *International Journal of Nursing Studies*, 43, 281–291.
- [12] Kellett, U. 1996 Caring in nursing homes: the family perspective. *Geriatrics*, 14(4), 18–23.
- [13] Li, H., Stewart, B., Imle, M., Archbold, P. and Felver, L. 2000. Families and hospitalized elders: a typology of family care actions. *Research in Nursing and Health*, 23, 3–16.
- [14] Bergland, A. and Kirkevold, M. 2006. Thriving in nursing homes in Norway: Contributing aspects described by residents. *International Journal of Nursing Studies*, 43, 681–691.
- [15] Kindberg, T., Spasojevic, M., Fleck, R. and Sellen, A. 2005. The ubiquitous camera: An in-depth study of camera phone use. *IEEE Pervasive Computing*, 4(2), 42–50.
- [16] Kindberg, T., Spasojevic, M., Fleck, R. and Sellen, A. 2005. I saw this and thought of you: some social uses of camera phones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – Extended Abstracts. CHI '05: ACM Press, New York, NY, 1545–1548.
- [17] Van House, N., Davis, M., Takhteyev, Y., Good, N., Wilhelm, A. and Finn, M. 2004. From ‘What?’ to ‘Why?’: The Social Uses of Personal Photos. In: Proceedings of the CSCW'04.

- [18] Garau, M., Poisson, J., Lederer, S. and Beckmann, C. 2006. Speaking in Pictures: Visual Conversation Using Radar. In: Second Workshop on "Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology" (PICS), Ubicomp 2006.
- [19] Lim, J. and Jin, J. 2003. Using Dual Cascading Learning Frameworks for Image Indexing. In: Workshop on Visual Information Processing 2003, Conferences in Research and Practice in Information Technology, Vol. 36.

Disembedding Computers – Interfacing Ubiquitous Computers

William Edmondson
Advanced Interaction Group
School of Computer Science
University of Birmingham
Edgbaston, Birmingham, B15 2TT, UK
w.h.edmondson@cs.bham.ac.uk

Russell Beale
Advanced Interaction Group
School of Computer Science
University of Birmingham
Edgbaston, Birmingham, B15 2TT, UK
r.beale@cs.bham.ac.uk

ABSTRACT

In this paper we look at four different interpretations of the term ‘Ubiquitous Computing’ – many computers; people using them much of the time; embedded computers, and ‘invisible’ systems – and consider how the two more specialist interpretations are being undermined by the other two. We explain how the increased manifestation of computers in our environment alters the ways in which we should consider how to design ubiquitous systems. There are some specific implications for design of interfaces to artefacts containing embedded computers and these are discussed in the context of recent work on Projected Cognition.

Keywords

embedded computers, disembedded computers, ubiquitous computers

ACM Classification Keywords

H.5 Information Interfaces and Presentation (e.g., HCI); H.5.2 User Interfaces – *User Centred Design*; H.1.2 User/Machine Systems – *Human Factors*.

INTRODUCTION

The term “ubiquitous computing” is used frequently nowadays, but what do we actually understand by it? It is now too vague a term for sensible analysis of both artefacts and behaviours, since there are four different plausible interpretations with implications for understanding interaction behaviour. The first and most obvious is that computers are ubiquitous in the environment in the sense we have them at home, in offices, shops, libraries, cafés, surgeries, etc... *recognisable as computers*. The machines may have different/specialist operating systems and/or applications but they remain obviously computers and can serve, say, as a basis for dialogue between two or more people about material displayed on the screen (typically this can be (re)positioned to display to the person(s) other than the operator). The specific functionalities may be unfamiliar to non-operators, but the fact of functionality, and some coarse

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

detail, will be obvious to everyone. In addition, the items identified as computers probably share certain characteristics which make them so recognisable – they will have a USB port, say, and a keyboard/screen arrangement which is unremarkable (there may be differences in keyboard layout, but such differences only emphasize the presence of the core components). Furthermore, as fashions change in the design/configuration of the machines then so too people will find their willingness to identify an artefact as a computer will change or develop – CRT displays have given way to flat screens; desktops have given way to laptops. And more types of machine (notebooks, PDAs) have become available. Yet for all this – a ‘computer’ is readily recognized as such, with attendant assumptions concerning basic functionality. They are ubiquitous in the sense of ‘being everywhere’ and ‘everywhere’ recognisable for what they are.

The second interpretation of the term concerns the *use* of such machines, or devices resembling them. The impression or observation we seek to build on here is that people are ‘all the time’ and ‘everywhere’ using computers. To the casual observer it seems to be the case that many individuals are constantly wired up and/or wirelessly ‘connected’ to some resource or another (monopolising power sockets in cafés, airport lounges, etc., whilst ‘doing email’, or surfing the web, or indeed working). Here the sense of ubiquity is that of many people constantly using computers, in the colloquial sense (rather than, say, the specific scientific/mathematical sense). Indeed – with the distinction between mobile phones and other digital devices (iPhones, ‘Blackberry’ devices, various sorts of smart phones, iPods with wifi.....) becoming very blurred, the sense of being surrounded by ‘connected’ users is unavoidable.

For an individual, then, these two senses of ubiquitous computing are differentiated as i) encounters with many computers in many environments and ii) frequent use of personal computational devices/resources in any/all environments. These seem to be two aspects of the statement – ‘computers are everywhere’. Ubiquitous is not understood as ‘invisible’ in these senses or interpretations of the term.

However, there are two other, more ‘specialist’, senses of the term ‘Ubiquitous Computing’ which don’t mesh so readily with the foregoing observations, especially with regard to assumptions about visibility. Weiser’s work – in the 1990s – led to the term ‘ubiquitous computing’, and he promoted a specific concept – that computational devices would be everywhere, providing *invisible* computational support, or a computational environment [1]. Others have pointed out that this conception is incomplete because users’ interaction with the computational resources requires separate consideration [2]. For example, the notion of a system model, or more generally a cognitive model of the computationally enriched

environment, is not readily addressed through the core concept of invisible ubiquitous computing.

The fourth sense of Ubiquitous Computing is sometimes conflated with Weiser's more abstract notion, perhaps by way of furnishing evidence of the emergence of such an environment. Artefacts – typically 'white goods' found in the home (washing machines, cookers, fridges.....) – have computers embedded in them for functional reasons. It is not the case, of course, that refrigeration is a computational problem – the issue is merely that control of a domestic appliance can be made easier, and perhaps more informative for the owner, if a computer (of some sort) is embedded in the artefact to provide more information, the option of programmability, or whatever. For the purposes of discussion here this sense of 'ubiquitous computing' will be termed 'embedded' computers. There is clearly a sense in which the computational resource or function embedded in the artefact is invisible – one simply doesn't recognise that one's washing machine or sewing machine has a computer in it. Indeed, this focus on artefacts provides some support for Weiser's conception, and if one can extend the notion of artefact to include just about everything in a room (pen, paper, light-switch, chair....) and furthermore have these devices networked – then this sense of ubiquitous computing becomes Weiser's without effort.

DISEMBEDDING COMPUTERS

Since Weiser's original work the environment turned out to look different. If we look around us, we see an environment in which all these sorts of ubiquity exist – many computers, lots of people using them much of the time, with some embedded and invisible, and some of those networked together. But we do have all of these, not the seamless, invisible computational mesh once envisaged as the ubiquitous future. The first two types of experience of ubiquitous computing noted earlier moved the debate in a different direction. Our claim is that these developments have served to disembed the computer and thereby undermine the 'invisibility' of ubiquitous computing. How did this happen?

Firstly, it is possible that because people see so many computers around them they will expect to encounter them even more widely in their environment in the future (this was already discernable in the 1990s – see below). This has been encouraged by the spread of computer-like technologies such as touch/flat-screens for train information displays, ATM machines, etc. Indeed, computers as such are also becoming more variable in functionality, as DVD players and gaming machines, for example. Secondly, ubiquitous usage may well lead to expectations that artefacts will/should behave more like computers – the limited range of functionality offered by a machine with an invisible embedded computer may become irksome, leading to later generations of product becoming more complex. This is already discernable, arguably, in such things as washing machines (see below) and microwave cookers. Thirdly, complex artefacts with embedded computers are increasingly offering users an interaction experience reminiscent of computers – washing machines being driven through menu-based displays, and not via push-button interfaces. Fourthly, an important aspect of the interactional experience with many computer systems is 'tailorability' or 'personalization' and this too is surfacing in complex artefacts with embedded computers.

EXAMPLES

The outcome of the above influences on design is that in fact the embedded computer is becoming more often disembedded

and presented to the user as an interactional experience akin to using a computer. In one 'early' example (mid 1990s) of embedding a computational element – a sewing machine (see the Bernina 1080 illustrations: Figure 1 and Figure 2) – the fact of the computational element is visually prominent, despite the machine's obviously 'non-computer' appearance and controls. In a more recent device – a washing machine – the user's interactional experience is clearly informed by symbol conventions and selection behaviour found in the world of computers (see the washing machine (Miele W3622) control panel illustrations below, Figures 3, 4, 5, 6).

Bernina Sewing Machine

Figure 1 shows part of the control panel of the machine. The user's attention is drawn to the fact that the sewing machine has a computer inside it by the word 'computer' on the fascia. The control knobs behave as such things do, and the graphics or symbols relate to sewing, not to computing.

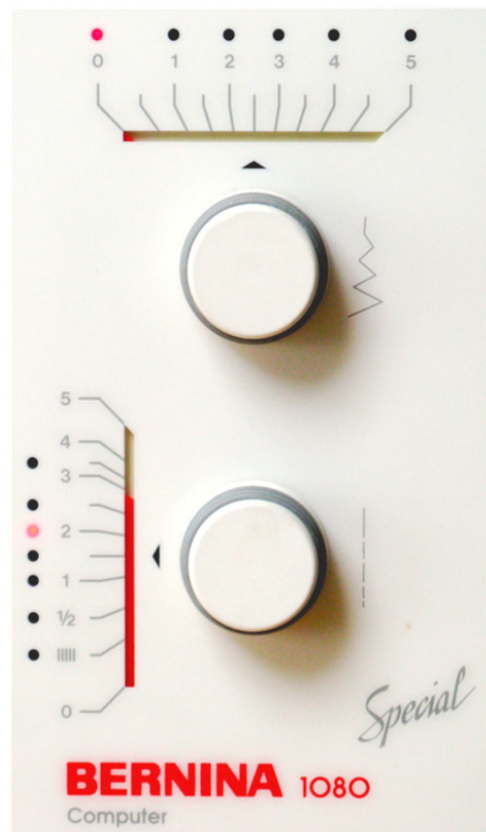


Figure 1. Part of sewing machine interface.

Figure 2 (below) illustrates another aspect of the interface – many buttons (the switch at the top reveals another 14 stitch types), lots of depiction, few dependencies, and some labels which reminds the user of the computer's involvement.

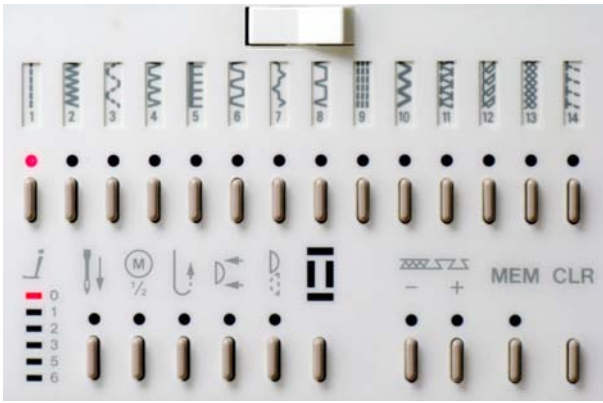


Figure 2. Part of sewing machine interface: many buttons – few dependencies.

Miele Washing Machine

The washing machine control panel is shown in four images (Figures 3–6). It has the customary rotary knob for selecting wash programme and some buttons for selecting other functions – along with a display giving textual information. One setting of the rotary control (Figure 5) actually is a pointer to a menu of possible settings shown elsewhere (Figure 6). Figure 3 illustrates the basic control knob. Figure 4 shows other controls. There is a small window on the panel, labelled ‘PC’, which is for optical communication with a PC used by a service engineer (for checking installed parameters, and adjusting them as well as updating programmes).



Figure 3. Washing machine – basic control knob.

Note the repeated use of switch position for temperature settings (30, 40, 60, 95 °C), only one of which is labeled °C.



Figure 4. Washing machine – remainder of control panel.

Figure 4 shows the remainder of the washing machine control panel for the wash selected in Figure 3. The highlighted spin speed shows this value to be variable (using the + and – buttons

below the display; in fact 1600 is the maximum speed of the machine, and also the default for the selected wash).



Figure 5. Washing machine – set of washes.

Figure 5 shows the selected ‘programme’ is merely a set of further programmes – shown as a menu in Figure 6.

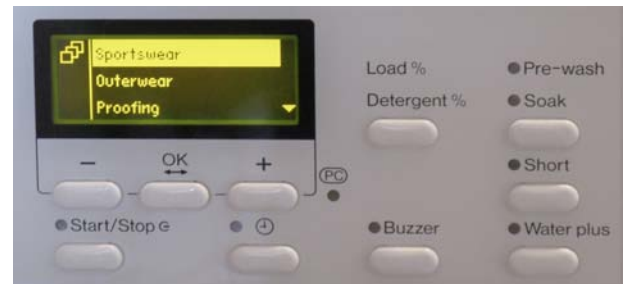


Figure 6. Washing machine control panel showing scrollable menu of further programmes.

INTERPRETATION

We noted above our view that the embedded computer is becoming more often disembedded and presented to the user as an interactional experience akin to using a computer. We offer the two examples above as illustrations of this theme over the last couple of decades. The sewing machine uses its computer in a variety of ways incomprehensible to most sempsters and sempstresses, so the front-panel mention of the computer is possibly as much for reasons of marketing (‘this sewing machine is *modern*’) as anything else – but that in fact makes the point drawn out earlier, that computers are known by users to be ubiquitous (and were, to product conscious consumers two decades ago).

The washing machine is clearly offering the user an interaction experience directly informed by widespread use of menus (indeed, the ubiquity of menu selection is as much the point as the ubiquity of the computer). This example, in our view, clearly brings out the computer as ‘disembedded’ and the device is ‘computer-like’ through its deployment of menus and its tailorability.

CONCLUSIONS

Our concern is that if this observation is substantiated (and the more products one recognizes with ‘disembedded’ computers the more obvious the observation) then designers need to work explicitly with the issues raised. It seems inevitable that domestic appliances will come to be seen as controllable and configurable in ways not previously considered important; in particular, in providing opportunities for users who wish to

project onto those devices their own means of achieving outcomes previously designed into the artefacts in very limited ways. The washing machine illustrated earlier quite clearly provides pre-specified wash cycles, but the adventurous user can do two things to 'take control' and use the machine with intentions which they project onto the device. For example, they may decide after experience of various cycles that they prefer more water, or that they wish to configure the machine for more gentle washing generally (this can be done through a 'settings' menu), or they actually prefer to wash their cottons on 'Express wash', or whatever (one should not forget that intentional 'mis-use' of a machine is nonetheless intentional use of that machine). In other work we have termed this support for individual user expectations, workflows and preferences 'Projected Cognition' [3, 4].

This is important, since specific implications emerge for designers following recognition of disembedded computers as prominent in the future of artefacts/machines:

- A) the functionality of the machine/artefact can no longer simply be assumed as obvious – task analysis and usability studies will become much more complex;
- B) many basic issues in HCI will need to be revisited (nature of display, special needs access/control issues, etc.);
- C) configurability/personalizability, and notions from Projected Cognition such as the expression of functional intention, will need to be understood as essential components of interaction with such machines;
- D) attention will need to be paid to issues of system maintenance in environments where, for example, apparently the same machine in two different locations turns out to be two significantly different machines.

Some of these issues are being addressed by some manufacturers already. Miele service engineers are equipped with laptops which interact with the computers in their appliances to determine the configuration/state of each machine they encounter – both

users and engineers can configure appliances so that settings are not obvious without a full system check and printout.

What appears much less obviously extendable from the status quo, and from pre-existing HCI work, is how to explore fluid and tailorable functional requirements for machines which previously had been considered essentially single-purpose. Washing machines, for example, are now so tailorable that it is a significant challenge to design for a user's appreciation of the scope of tailorability in something apparently so simple as washing clothes. Washing machines, having trained users to deal with few options, now offer great variability – whilst still offering a simple repertoire for those who want it. This style of requirements analysis and subsequent redesign is unfamiliar in this setting and will pose significant challenges as computers continue to become more disembedded. We consider this to be an urgent issue in ubiquitous computing, a complement to the challenge of engendering and supporting a system image for something thought to be desirably invisible.

REFERENCES

- [1] Weiser, M. 1993. Some Computer Science Issues in Ubiquitous Computing. *Communications of the ACM*, 36, 75–84.
- [2] Rehman, K., Stajano, F. and Colouris, G. 2002. Interfacing with the Invisible Computer. *Proceedings of NordiCHI*, 2002. Pp. 213–216.
- [3] Edmondson, W. H. and Beale, R. 2008a. Projected Cognition – extending Distributed Cognition for the study of human interaction with computers. *Interacting with Computers*, 20(1), 128–140.
- [4] Edmondson, W. H. and Beale, R. 2008b. Projected Cognition: capturing intent in descriptions of complex interaction. *Proceedings of 7th ACM Conference on Designing Interactive Systems*, 2008. Pp. 281–287.

Session 5: Supporting Activity by Virtual and Augmented Technology

Augmented Reality as Means for Creating Shared Understanding

Mirja Lievonen
Royal Holloway University of
London
Egham Surrey
TW20 0EX, UK
m.lievonen@rhul.ac.uk

Guido Kühn
Electronic Arts GmbH,
Cologne, Germany
gkuehn@googlemail.com

Duska Rosenberg
Royal Holloway University of
London
Egham Surrey
TW20 0EX, UK
d.rosenberg@rhul.ac.uk

Swen Walkowski
Accenture GmbH
Germany
swen.walkowski@gmail.com

Ralf Dörner
Wiesbaden University of
Applied Sciences
Kurt-Schumacher-Ring 18
D 65197 Wiesbaden,
Germany
doerner@informatik.fh-
wiesbaden.de

ABSTRACT

The motivation for the work presented in this paper comes primarily from user experience of video-conferencing (v-c) settings in real-life collaboration. The design issues in this context focus on making such settings interactive enough to support natural communication and collaboration. The initial assumption is that users in an interactive v-c setting should be able to navigate the remote space in order to establish clear reference by pointing to people and objects in it. Clear reference to parts of the context in which conversations take place (that is, deictic reference) is an important factor in effective communication. With this aim, we enhanced the videoconferencing system with the ability to visualize abstract representations of pointers and investigated pointing gesture as a tool for collaborative referring. We thus designed a prototype that combines the communicative function of pointing gesture with a hybrid representation of real video and virtual objects (pointers) that identify particular parts of it. A game controller was employed for pointing and Augmented Reality (AR) for visualizing the referent in live video stream. Usability tests were run on two versions of the prototype using five common types of joint task. Evidence based on data from video recording, questionnaire, and interview, shows effectiveness of the system in mediating the communicative function of pointing. Test users adapted to the context of interpretation quickly. Feedback was provided for enhancing visualization and pointing technique. The method successfully captured relevant visuo-gestural and linguistic aspects of communication to inform design.

Keywords

videoconferencing, communication, pointing, communicative function, deictic reference, shared understanding, common ground, remote collaboration, interface design

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *user-centered design*.

INTRODUCTION

Collaboration is a joint effort where participants, each in a particular role and location, focus on a common goal or topic. Conversation is a joint effort where participants take turns in contributing to it, and jointly carry responsibility for maintaining shared understanding in its course [4]. Communication and collaboration have therefore an inherently non-separable nature. Though there has been a lot of talk about location-independent way of working, there is certainly no context-independent way: when one component of a work situation changes, the whole situation changes accordingly.

Pointing as Problem in Videoconference

People are attuned to face-to-face communication that relies on the presence of a communicative partner. Therefore, it is not surprising that in technology-mediated settings, our intuitive expectations at times fail. For instance, people may find themselves pointing to the remote site view on the video screen, though immediately realizing it to be a failure since the participant on the other side cannot see what they are pointing to: the pattern of co-located pointing does not work. People have particularly to learn to create an artificial sense of social presence at the remote site.

The communicative function of pointing to a remote site object fails because ordinary videoconferencing systems fail to provide sufficient information about dispersed locations to see (& understand) what people at the far end are actually pointing to. The impact of such a failure is significant, because it is natural for people to use non-verbal channels in communication, in order to create and maintain shared understanding [12].

In face-to-face situations, gestural and verbal modes are naturally integrated. Turn taking, for instance, is much easier compared with videoconferencing [9] where participants have to rely on poorer visual means that give limited contextual information. Yet, videoconferencing has become a commonly used resource for multiple types of meetings and education. In professional contexts, it is a way to reach such expertise that is not locally available. Therefore, there is a great practical motivation

for the research in the gestural aspects of remote communication and collaboration, in order to enhance the cross-modal (both verbal and non-verbal) quality of communication and collaboration through videoconferencing.

In a collaborative research project¹⁰, we summoned multidisciplinary knowledge background in Social Sciences (Linguistics), Architecture, and Computer Science, setting off to investigate deictic references in video-conferencing. The assumption was that multidisciplinary background would help us address the problem of pointing in videoconferencing, explore ways to mitigate it, and enhance by design the cross-modal quality of video-mediated communication and collaboration.

Structure of Paper

In the following sections, we report findings from the IS-VIT project, where we developed a tool for pointing to objects located at the remote site in videoconferencing.

Section: “Pointing Gesture in Human Collaboration” describes the theoretical approach to deictic referencing which was based mainly on cognitive (usage-based) linguistics. The approach takes into account cross-modal aspects of communication. It provides us with design requirements described in Section: “Conceptual Model for Remote Pointing”.

In order to test the feasibility of our conceptual model in practice, we implemented two versions of the prototype for usability evaluation. The first testbed application¹¹ was developed for one-way pointing, the latter one had capability for two-way pointing. Both applications employed a game controller as a pointing device, and AR methodologies [1] to visualize on the screen the location of a referent. It provided the participants with a visual common ground for creating and maintaining shared understanding of what they tried to communicate to one another by pointing.

Section: “Usability Testing” covers usability evaluation method and procedure. We ran usability tests on both the testbed applications to explore user experience (N = 20 + 23) and gain preliminary insights for further development of the pointing system. The findings were based on data collected from video recordings, questionnaire, and feedback discussion.

Main results from our usability tests are reported in Section: “Findings”. We tried to establish whether the pointing system is viable and whether it can support effective communication in a videoconferencing setting: whether the users managed to reach shared understanding of their deictic references through mediated pointing; whether the test users adapted to unfamiliar spatial context of interpretation. The feedback gained from the interviews with the users provided ideas for improving usability of the system.

We conclude, in Section: “Discussion and Conclusions”, by providing three design implications for pointing systems and suggest topics for further research.

POINTING GESTURE IN HUMAN COLLABORATION

Role of Pointing

Deixis, the way of locating the referent of discourse, is typically communicated using different modes of communication, such as verbal utterances, pointing gestures, and gaze direction [11].

Deictic expressions (e.g. ‘that one’, ‘there’) rely completely on their context. Pointing and linguistic communication share the same social-cognitive and social-motivational infrastructures [17, 18]. Iconic gestures activate image-specific information about the concepts which they denote, and co-speech gestures modulate conceptualization [19]. Visual conduct, in concert with speech, is used to produce demonstrative reference; or, approaching from the opposite perspective: intelligibility of referential actions is grounded in the activities [10]. Deixis therefore bridges pragmatic and semantic aspects of discourse [13]: what is pointed at is what is referred to.

Deictic gestures make communication robust by disambiguating verbalization, more efficient by reducing a need for procedural utterances; they manage, to some degree, to enhance shared understanding even beyond language barriers. They can be synchronous to verbal expressions but often manage to convey a message, in concert with facial expressions and gaze direction, without any verbal utterances. In early life, they help us to express what we want by pointing towards the object of our interest, thereby appealing for others to get hold of it. And at old age, if our memory deteriorates, we can still manage quite a long way with a pointing gesture and demonstrative pronoun [13]. It is therefore no wonder that demonstratives go far back in the evolution of language and that children learn to use them early on [6]. In addition, demonstratives and interrogatives have similar pragmatic function of initiating a search for a specific referent [5]. Therefore, gestures and demonstratives can be regarded as very basic tools for human navigation in the world.

In face-to-face conversation, people rely on multiple modes of communication. They tend to pay more attention to the face and gesturing hand [16]. Cues, such as volume and tone of voice, gaze direction, facial expression, hand gestures, body position and orientation, regulate the flow of communication, facilitate turn taking, provide feedback, and convey subtle meanings [8, 11, 14]. Experiences from videoconferencing tell us that for instance turn-taking is not as fluent as in face-to-face situations.

A necessary component of deictic use is the ability to take perspective [7]. Thereby, two communicating perspectives necessitate a common ground for establishing a common reference in communication [3, 4]. Gaze direction and hand gestures are integral part of face-to-face communication. According to developmental psychology and usage based theory of language, joint attention plays a foundational role in communication [5, 18]. Joint attention implies a triadic organization: participants of the communication understand that each of them has a different point of view onto the focus of joint attention [5].

Attention manipulation comprises various communication modes. When people consciously direct others’ attention, they tend to take into account socio-cultural norms and conventions of context-appropriate behavior. That is the case for instance when listeners want to join in, or outsiders wish for to interrupt an ongoing communication.

¹⁰ IS-VIT (Interaction Space in Virtual IT Workplace) funded by the Academic Research Collaboration program between Germany (Wiesbaden University) and UK (RHUL).

Cognitive Perspective to Communication

In order to study human communication, in particular its pragmatic and semantic aspects, we have to presuppose reciprocity of perspectives with a common ground bridging them. Human perspective has multimodal nature because people rely on both verbal and nonverbal modes of communication. Communication studies have pointed out a requirement to include verbal, gestural and structural aspects of the communicative situation in the analysis [15]. Our approach therefore takes into account inter-, and intrapersonal, multimodal, cross-modal, and dynamic aspects of human communication and collaboration.

CONCEPTUAL MODEL FOR REMOTE POINTING

We developed a conceptual model for pointing application, using scenario techniques for eliciting ideas of use. Based on existing literature and partly on own prior findings from observations of interactions within groups, we selected the scenarios for observing pointing gesture in videoconferencing in the context of realistic communicative situations.

Our aim was to investigate whether it were possible to mediate the communicative function of pointing gesture in videoconferencing in a successful way, employing Augmented Reality techniques. Does such a pointing system enable shared understanding of deictic reference over distance?

The design instruction was to find a technical solution that provides an overlap of the visual (perceptual) fields of participants at both ends to provide common ground so that they can understand the meaning of pointing at a distance.

The conceptual model was based on Bühler's (1934) [2] views that the information a person articulates is always related to a specific point in (i) space, (ii) time and (iii) identity, i.e. the person's I-Here-Now-Origo at the time of articulation. The information referencing such Origo (deictic centre) is called a deictic reference, and all information available to such Origo is called the Information Space.

We interrelated communicating people's Origos and Information Spaces, taking into consideration that each communication mode has their particular requirements, (such as range and scope of sight of communicating partners, lighting conditions, no obstacle in the line of sight, range and scope of hearing, reach of hand).

The application designer came up with an idea of providing the participants at both ends with identical views of live video stream that would provide them with a visual common ground; in this context, visualizing the referent would allow all participants to see and grasp *who* is pointing, and *what* is pointed at – provided that the target object of pointing was within the camera view.

Assuming that the basic concept was viable, we implemented a testbed application in videoconferencing. We employed Nintendo's Wii game console controller for pointing. The system used infrared detectors for tracking the direction of the pointing relative to the video screen. The images shown in the videoconference were augmented with virtual imagery employing AR methodology in order to represent the location of the referent of pointing on the video screen. The application thus allowed the participants to navigate the remote space, using the AR pointers to focus on objects in it.

We implemented two versions of prototype: the first application was developed for one-way pointing. The latter version enabled

two-way pointing, i.e. both the local end and remote end participants were able to point to the opposite site. (In the following, we refer to the former as testbed A, the latter as testbed B)

USABILITY TESTING

The aim of the usability test was to discover in what ways and to what extent (if any) the augmented videoconferencing facility helps the participants to communicate and collaborate effectively.

We therefore designed a test situation where we could investigate the viability and usability of the pointing system. A pilot test and 3 series of tests were run on both testbed applications. Two additional tests were run on testbed A.

Testbed A

The Wii Remote controller was connected via Bluetooth to a PC running the Windows operating system. With GlovePIE (Glove Programmable Input Emulator), the incoming and outgoing signals have been mapped to keyboard events for easy accessibility in common interface programming libraries. We used the OSGART library, which provides the functionality of the ARToolkit within the OpenSceneGraph framework, to implement a prototypical video conferencing application. Visual representations for different pointers have been developed by writing 'Shader' programs with GLSL (GL Shader Language).

Several functionalities were mapped on WiiRemote, and organized into four different modes. Mode switch was operated by calling an 'On-Screen-Menu' via the remote controller.

The system provided, firstly, functionality for 'plain' pointing. In pointing mode, different representations were available for pointing to human vs. to nonhuman objects. A representation designed for pointing to people had the appearance of an object of pointing being in a limelight. An arrow –representation was designed for pointing to nonhuman objects, as well as for indicating directions, and a ring-option to indicate and zoom for the range of an area. It was also possible to turn the direction of the arrow, even make it spin around, to add emphasis on the referent, or to alert (or even amuse) the audience.

Apart from pointing mode, additional functionalities included an option to add short notes relative to the topic discussed, and a snapshot option: thereby the participants could collect a pictorial record (max. 4) of suggested alternatives, and retrieve them for show before concluding their common view. The snapshot option was furnished with a sound effect to provide shared awareness of each caption.

Finally, there was a mode for prerecorded information to be retrieved relative to remote site participants or objects. The participants were given a role name and affiliation, which was inserted into the system and could be retrieved by pointing at the marker worn by the respective participant. This was meant to serve the chair of a meeting in a situation where s/he might forget the name of a participant.

The functions were operated using the keys on WiiRemote. A few functions, such as snapshot and sticker notes, also required the user to rely on the keyboard of his/her PC. In order to 'mute' the pointing, the device had to be turned upside down, because this is an easy way to accomplish this 'muting' function as the user need not concentrate on pushing a specific button on the controller but interacts with it as a whole. That helped us to see how different ways of operating the controller worked in the test situation.

We made several improvements, based on our mutual discussions and user experience, before commencing the tests. By then, a rumbling effect had been included to indicate when the direction of the pointer was going out of the screen frame, a spinning arrow to accentuate or alert, feedback (sound) to inform about snapshot caption, and feedback to inform how many snapshots had been captured to memory.

Testbed B

Testbed B differed in four main respects from testbed A: Firstly, we omitted most functionalities and focused on plain pointing only. Another significant difference was that testbed B was furnished with automatic face detection: once the direction of pointing crossed the face on the screen, it automatically turned an arrow representation into a limelight one. Third difference was two-way pointing: testbed B allowed simultaneous pointing across the two sites. We could thereby pay attention to negotiation strategies and possible competitive situations. Fourth, people could point at both the local and remote site screen. We could then observe whether they point in a traditional way when indicating a local site object, or whether they primarily concentrate on the screen and communicate their pointing gestures using WiiRemote.

In testbed B, we used also color coding to distinguish between the two groups. The pointer representation was color coded according to the site which was pointing, and there was a similar color line around the respective side screen. If there were two arrows on the screen, you could know which of them represented the local site pointer, which one was from the opposite site.

All the options were controlled by WiiRemote. Two different functions were mapped on the cross-shaped key, using its up- and down arms a) for spinning the arrow clockwise and anti-clockwise around the indicated object in the virtual space, and b) for zooming the size of the virtual pointer representation. Depending on which screen the user wanted to point, s/he had to go for either A-button or B-button on the WiiRemote. The design instruction preferred the option of pressing a key while pointing, but the design after all required the user to press the button twice (both for turning the virtual pointer on and off).

Settings

Our test settings were influenced by Clark and Wilkes-Gibbs (1986) [3] when they studied referring as a collaborative process. Our settings consisted of a desk and two to three chairs for both communicating groups. There were two displays on the desk representing augmented video stream, one from the local site, the other from the remote site.

In the test setting for testbed A, the camera was positioned above the display representing the remote site view. Only one WiiRemote was available for pointing in this one-way pointing experiment. Our observation mainly focused in this case on the ways users interpreted the referent and context of pointing.

The two groups were separated from each other by a partition, so they could not see each other, but they had full audio presence. By this, we could disregard audio as a problem. In the settings for testbed A, the two computers were directly connected by a cable, so it was actually a simulation of videoconferencing. It was also a way to avoid video and sound delay caused by network latency problems in video- conferencing.

The two displays with the video stream, augmented with virtual objects, could be controlled by the local end participants in the

one-way pointing setting, and at both sites in the two-way pointing setting, in a synchronous event.



Figure 1. Test settings for the testbed A.



Figure 2. Test settings for testbed B.

The settings for testbed B (two-way pointing) were mainly similar to those for testbed A: two desks were separated by a partition, two displays on both desks, showing the augmented live video from the local and from the remote site.

In the setting for testbed B, we had – because of the constraints of infrared detectors – to locate the displays a bit further away from each other than in the settings for testbed A. We positioned the camera between the two displays as this helped us also to track the changes in gaze direction.

Tasks

The test battery consisted of such common joint tasks that involve pointing, such as

1. introducing people, referring to a person by 2nd person pronoun,
2. taking questions in a meeting (floor control, turn taking),
3. indicating an object by a hand gesture,
4. locating a destination/area on a map, and
5. instructing the way from place A to place B on a map.

We verbalized the tasks leaving plenty of freedom for the group to create their own interpretation of it, and thereby communicate and carry out a joint task. We also provided

relevant artifacts in such a way that the dispersed participants were compelled to assist each other over the video link in order to complete the task. For instance, there was a floor plan in one end, and the opposite end participants had to participate in room allocation.

Tests Sessions

Three series of tests were run on both applications in addition to pilot testing. On testbed A, two additional tests were run. The first test round was designed for a group of 2+2 participants (additional tests for 1+3). The usability tests on testbed B were for 3+3 participants. We had 43 test users in total. Though most of them were university students in their 20's, there was a wide range of test users in terms of age, nationality, profession and familiarity with videoconferencing.

The test groups had a chance to familiarize themselves with the application for a few minutes, after which they were asked to carry out five joint tasks. They were sitting on the opposite sides of a partition so that they could not see to each other's side, though they had practically collocated audio-presence.

The test tasks were estimated to last around half an hour but in many cases took longer. All the sessions were video-recorded, and feedback was collected through a short questionnaire and discussion after the test session.

Analysis of Data

The video recordings were analyzed using people-place-process framework. The designer of the test tasks and the designer of the application jointly watched all the recordings from the tests, identifying communicative situations, where people were pointing, for what purposes they did it, and what impact pointing had on communication if the pointing device was used. We looked for any indications revealing a success or a failure of the pointing system in supporting the group to carry out their joint task.

The focus was, on one hand, on pointing in the context of the utterances and actions, on the other hand on impacts on the audience: whether the communicating participants managed to understand 'mediated' pointing, and thereby, establish and maintain shared understanding of deictic reference during their communication.

1. What for / why was the person pointing in that particular context (identify an object, indicate a way, address a person, give the floor, take a question, etc.)?
2. How was the person pointing in that particular context (with a finger vs. with WiiRemote: using a local map vs. using a picture on the screen)?
3. What was the impact on the audience: did they seem to understand what the person pointing was trying to communicate?
4. What signs there were of shared understanding (e.g. gaze direction, nod, voicing, verbal confirmation)?

The analysis involved replay of critical parts over and over again in order to get a better understanding of relevant aspects of that particular situation, and saving captions for pictorial evidence.

We sought to identify the impact of 'mediated' pointing on the participants' behavior, for instance: what kind of behavioral indication did our data provide to support our assumption that they were able to infer communicative function of pointing and

thereby create and maintain shared understanding of the deictic references during the session? Were there nods, voices uttered for confirmation, gaze following the line of pointing on a map, turning to look at a person whom the chair was pointing when giving the floor, marking a correct point on a map, and so on?

In addition, we paid attention to activity and attitude in pointing: who were active, and whether they were competitive or collaborative (in testbed B setting), and techniques of pointing: whether they were pointing to a tacit local object with a finger or preferred pointing with the pointing device to the screen, and how they managed to handle the pointing device. We also looked for indications of any affect response in an unfamiliar context of interpretation.

The recordings from the tests on testbed B allowed us to deepen our analysis as we could meticulously analyze gaze direction in co-occurrence with discourse and pointing.

We built our findings on the evidence from video recordings as they provided us with a rich, detailed and spatiotemporally meticulous and analyzable source of data. In our experiment, we mainly wanted to find out whether the concept was viable at all, which did not require extensive testing. As we also wanted to learn user perceptions of the system, we had a questionnaire and a feedback discussion with the users straight after the test session. Data from questionnaires and comments in the feedback discussion provided preliminary indications of how the pointing system was perceived by novice users.

FINDINGS

We found out from the tests that it is possible to 'mediate' pointing gesture in an effective way. It was surprising how quickly the participants adapted to a novel context of interpretation: they did not seem to have any difficulty in inferring deictic reference from the information provided by our enhanced videoconferencing. Furthermore, they seemed to go for the least effort option which shows that the threshold for adopting mediated pointing was low: for instance in the settings where they could choose whether to point to the map at their local site or to its representation on the screen, many users preferred pointing to the screen (Figure 3).



Figure 3. Pointing to a floor plan.

Effectiveness

The findings from the tests support our initial assumption that the designed system would meet the requirement to mediate communicative function of pointing. The system was capable

of supporting the establishment and maintenance of shared understanding of deictic reference. The tests showed in principle that videoconferencing can be enhanced with pointing gestures, using visualization techniques in connection with tracking the direction of the pointing hand.

We got clear evidence that people identified the deictic reference without difficulty. (As the system had been introduced to them before the sessions they knew that each site had identical views of both sites.) In the test sessions on testbed B, they might for instance point to a remote site object asking:

What have you got over there?

Oh, that one, it is (the name of the object).

The video recordings show that people follow by gaze the path of the pointer representation on the screen, and at times nodding that they are following the communication.

The immediacy of the response indicated that creating shared understanding of the referent was no big effort; for instance, when a lady pointed to the floor plan located at the opposite site in a negotiation of room allocation, saying:

I want my room here!

There was an immediate protest from the opposite site:

That is my room!

Another type of clear evidence was caught from the floor control situation: if the chair was calling a remote site person to take the floor by pointing to his/her image on the screen and using the second person pronoun, the right person responded without hesitation. The person next to the person might turn, quite correctly, to look at the referred person.

Adaptation to the Context

The combination of functionalities mapped on WiiRemote was in the pilot test too heavy for any novice user to employ them parallel with other activities. Therefore, we modified the test design so that the application designer was playing the role of the chair in the meeting. He could then provide best practice for others. We turned the focus of our analysis on how the audience responded to pointing over a video link. We also assumed that the person sitting next to the chair might end up using WiiRemote – which was the case in the following tests.

The tests also showed that people adapted to the unfamiliar context of interpretation unexpectedly quickly. They got at times immersed to their tasks as if they were forgetting any difference. They even pointed to the remote site as if they were collocated, including the application designer himself – as can be seen in Picture 1. In one case, the person next to the chair, before she even noticed, pointed with a finger to the video from the remote site to show how the arrangement should be changed. When immediately realizing it, she burst out:

Oh gosh, I cannot show it!

The chair then passed the WiiRemote to her by saying:

Yes, you can – with this!

She started, however, to operate WiiRemote as if it were a cursor. This shows that she relied on a frame of interpretation familiar to her from using a mouse. Once the chair showed how to handle WiiRemote, she started to use it quite confidently. Later on, she just reached her hand to take WiiRemote from the chair's hand when she wanted to show something. Afterwards, when showing the video clip to her, she did not remember

having 'grabbed' the pointing device: she confirmed that she obviously did no more pay any particular attention to it; she seemed completely immersed and focused on the ongoing topic of a negotiation. This suggests the threshold to employing a mediated pointing device is hardly much higher than using a pointing stick.

We got some findings regarding the type and size of virtual pointer representation. We used a color limelight representation for indicating people at the remote site. When used, it had an activating effect: people interpreted it as a call to take the floor, as a request to act. From the user comments we learned that the possibility of being pointed to, kept them focused on communication, and thereby concentrated on the task. Some other comments revealed that it was annoying if someone was playing with the pointing device while others focused on the task, so pointing was in that case perceived as noise. (This happened only in the settings for testbed B.)

If the virtual pointer was persistently on the person's face in the video, s/he started to show signs of feeling uncomfortable. Yet, quite opposite perceptions were also mentioned: someone else found it pleasant to be in the focus of attention – as if feeling to be important. Visual appearance was a concern for some as they did not like the color (blue) of pointer representation on their face. An affect response was also captured when a (rather big) arrow approached a person's face in the video: the person might wince as if avoiding a real object falling onto him/her.

When the participants for the first time saw the system in action, we spotted facial expressions resembling curiosity and inquisitiveness familiar from the children's face when they try to figure out how something they do not yet understand, works, as if: what is the magic behind this? Even later on in the course of the test, if for instance the arrow was spinning in the video in order to accentuate a point in the ongoing discourse, there were exclamations such as:

Wau – How did you do that?!

The dispersed participants managed to carry out such cooperative tasks as for instance one of them remotely indicating where to locate a chair in the floor plan, and the other locally marking it for him/her. The person who was marking was checking in turns his/her local site screen for the referent of pointing relative to the position of his/her own hand, remote site screen for the person instructing, then drawing the mark on the floor plan behind him/her. The cooperation seemed pretty fluent.

The feedback from the user experience was positive in terms of perceiving the system as a potential improvement for videoconferencing. However, mapping too many functionalities on WiiRemote was too heavy for a novice to start using all of them after only a few minutes familiarization. A simpler combination was needed in order to make the system easy to use. (Even the application designer found the tests pretty tiring in his chairing role. Yet, speaking a foreign language may have added a toll to his cognitive load rather than mere operating the functionalities on WiiRemote.) Another weak point in the first design was the lack of mute key option for pointing.

In the second application again, automatic face detection of the system did not always work as expected. The design instruction suggested that the virtual pointer would have been activated only as long as the user kept pressing the key, because pointing turns into noise as soon as its communicative function has been completed. Yet, in the test sessions 'pointing noise' was partly due to nonchalance of social norms: some participants simply could not resist a temptation to explore their new 'toy'.



Figure 4. Flexible adaptation to negotiation situation.

Some of test users very quite fluent, others clumsier in handling WiiRemote: if people had to indicate the way from place A to place B on a map, some managed to do it fluently, using even expressive ways such as zigzag movement to show where to cross the road, or rotating movements to emphasize what they were saying. In the opposite cases, the movement of the virtual pointer looked pretty wobbly and restless.

Navigation Strategies

The users seemed to go for the least effort option which shows that the threshold for adopting mediated pointing was low: for instance in the settings where they could choose whether to point to the physical map on their site or to its representation on the screen, many users preferred to point to the representation on the video screen.

As most of the tested people were ‘young and techie’, they were at times more than active in their contribution. As a result, competitive situations emerged between the two teams when they negotiated room allocation. On the other hand, they managed also cooperative actions across the two sites well, as seen in Picture 4, where people are negotiating room allocation with remote site partners. At times there were, in the video image, both two virtual pointers and one finger on the map; at times people just grabbed the pointing device from the neighbor to show something, at other times again they made it available for one another. They were cooperative to the point that they bent to give space for pointing to the area that was in shadow for the camera, as can be seen in Picture 4.

The tests on testbed B provided rich data of the role of gaze direction: we could observe from the camera angle how keen a person’s gaze was going back and forth between the screens monitoring the overall situation and searching moment by moment for the most relevant object and focus of attention. Obviously the layout of the screens and the camera provided the communicating participants a successful visual illusion of social presence.

DISCUSSION AND CONCLUSIONS

Our findings from the usability tests of two videoconferencing pointing systems show that Augmented Reality enhances usability of videoconferencing.

Methodological Conclusions

The findings are compatible with the results of the cognitive science research on the ways people create the common ground, where ‘audience design’ has been identified as a critical factor in joint action to establish mutual understanding.

The assumption underlying the people-place-process framework applied in this experiment was compatible with the usage-based approach to language and situated approach to communication; it was particularly useful in taking into account cross-modal aspects of human communication and collaboration. The project yielded evidence for the significance of such framework in the study of mediated communication relative to the design of collaborative work environments.

The method applied in the project was viable for obtaining scientific results about fundamentals of human communication, where technology is a research resource.

Design Implications

In order to make dispersed collaboration more natural and effective, workplace design has to provide

- a visual common ground for dispersed groups to create shared understanding of who is pointing and what is pointed at,
- pointer representations used for pointing at people have to be designed taking into account that people tend to be concerned about their visual appearances in a video image, and
- the code of usage (‘protocol’) should follow the norms of appropriate behavior in co-located situations.

Further Research

Both the findings and the limitations of the project give rise to further research. One limitation of our project was using only PCs. The following step would be to explore pointing mediated on large scale videoconferencing system. Another question is

how people discern the deictic centre and referent of pointing, if more than two people are pointing at the same time.

In our project the users had full audio presence and the delay in video was hardly noticeable. The impact of transmission delay on communicating pointing gestures is therefore a question for further research. Also a question for a multidisciplinary team arises regarding spatial reasoning when people adapt to an unfamiliar technology-enhanced context of interpretation.

ACKNOWLEDGEMENTS

The ARC-project “Interaction Space of the Virtual IT Workplace”, carried out between Centre for Research in Information Computing and Communication (iCOM) at the Royal Holloway University of London, and Department of Design, Computer Science, and Media at Wiesbaden University of Applied Sciences, in 2006–2008, was funded by DAAD and British Council. A grant from Emil Aaltonen Foundation supported writing this paper.

REFERENCES

- [1] Billinghurst, M. and Kato, H. 2002. Collaborative Augmented Reality. *Communications of the ACM*, 45(7), 64–70.
- [2] Bühler, K. 1934/1990. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer.
- [3] Clark, H. H. and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition*, 22, 1–39.
- [4] Clark, H. H. 1996. *Using language*. Cambridge: Cambridge University Press.
- [5] Diessel, H. 2003. The relationship between demonstratives and interrogatives. *Studies in Language*, 27(3), 635–655. DOI = 10.1075/sl.27.3.06die.
- [6] Diessel, H. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17(4), 463–489. DOI = 10.1515/COG.2006.015.
- [7] Forrester, M. A. 1996. *Psychology of language*. London: Sage.
- [8] Fullwood, C. and Doherty-Sneddon, G. 2006. Effect of gazing at the camera during a video link on recall. *Applied Ergonomics*, 37, 167–175. doi:10.1016/j.apergo.2005.05.003.
- [9] Gemmell, J., Toyama, K., Zitnick, C. L., Kang, T. and Seitz, S. 2000. Gaze Awareness for Video-Conferencing: A Software Approach, *IEEE MultiMedia*, 7(4), 26–35, Oct.-Dec. 2000. doi:10.1109/93.895152.
- [10] Hindmarsh, J. and Heath, C. 2000. Embodied reference: A study of deixis in workplace interaction. *Journal of Pragmatism*, 32, 1855–1878.
- [11] Kendon, A. *Conducting Interaction: Patterns of Behaviour in Focused Encounters*. Cambridge: Cambridge University Press, 1990.
- [12] Kraut, R. E., Fussell, S. R., Brennan, S. E. and Siegel, J. 2002. Understanding the effects of proximity on collaboration. Hinds, P. J. and Kiesler, S. (Eds.). *Distributed work*. MIT Press. Pp. 137–162.
- [13] March, E.G., Pattison, P. and Wales, R. 2009. The role of cognition in context-dependent language use: Evidence from Alzheimer’s disease’. *Journal of Neurolinguistics*, 22(1), 18–36. DOI:10.1016/j.jneuroling.2008.05.002.
- [14] Prince, G., Demiris, Y., Marom, Y., Kozima, H. and Balkenius, C. Eds. 2002. *Proceedings of the Second International Workshop on Epigenetic Robotics, Modeling Cognitive Development in Robotic Systems*. Lund: Lund University Cognitive Studies, 94.
- [15] Roth, W.-M. 2004. Perceptual gestalts in workplace communication. *Journal of Pragmatics*, 36, 1037–1069. doi:10.1016/j.pragma.2003.11.005
- [16] Tan, R. and Davis, I. W. 2004. Differential video coding of face and gesture events in presentation videos, *Computer Vision and Image Understanding*, 96(2), 200–215.
- [17] Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. 3rd edition. Harvard University Press.
- [18] Tomasello, M. 2008. *Origins of Human Communication*. MIT Press.
- [19] Wu, Y.C. and Coulson, S. 2007. How iconic gestures enhance communication: An ERP study. *Brain and Language*, 101, 234–245. doi:10.1016/j.bandl.2006.12.003.

Field Observations of Therapists Conducting Virtual Reality Exposure Treatment for the Fear of Flying

Willem-Paul Brinkman, Guntur Sandino, Charles van der Mast

Man-Machine Interaction Group

Delft University of Technology

Mekelweg 4, 2628 CD Delft, The Netherlands

w.p.brinkman@tudelft.nl, mgsandino@gmail.com, c.a.p.g.vandermast@tudelft.nl

ABSTRACT

Recent research suggests Virtual Reality Exposure Therapy (VRET) for the treatment of fear of flying as an important reliable technique for this phobia. This paper focuses on the role of the therapist during an exposure session. Six therapists were observed in 14 sessions with 11 different patients. Results show that in 93% of the observed sessions, therapists started with a similar flight pattern. Furthermore, a total of 20 errors were observed where therapists initiated inappropriate sound recordings such as pilot or purser announcements. Findings suggest that the system might be improved by providing the therapist with automatic flying scenarios.

Keywords

virtual reality, exposure treatment, task analysis, field observations, user interface

ACM Classification Keywords

H.5.2. [Information Interfaces and Presentation (e.g. HCI)] Multimedia Information Systems – *artificial, augmented and virtual realities* – *Evaluations/methodology*. User Interfaces – *Ergonomics, Graphical user interfaces (GUI)*.

INTRODUCTION

In the industrial world flying has become an accepted mode of transportation. People fly to meet business partners, to attend conferences, to have holidays, and to meet friends and family. For some people however, flying comes with an undesirable amount of anxiety. Even so much that they avoid flying altogether or endure it with intense anxiety or distress. The fear of flying is categorised as a situational type of specific phobias in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) [1]. In its diagnostic criteria the manual also states that sufferers recognise that their fear is excessive or unreasonable, and it interferes significantly with their professional or social life. Reports on the fear of flight affecting the general US population vary, with estimations of 13.2% [4]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

and 3% [17], and only 0.4% in a survey among young woman in Dresden, Germany [3]. This survey also found that on average their responders developed this fear at an age of 15 years old, which lasted around 6 years.

Exposure in vivo, i.e. exposure to the real life situation, is regarded as the golden standard in the treatments of phobia and an extensive amount of research has been conducted in this area [6]. During this treatment, therapist and patient first develop a hierarchy of feared situations, and the goals a patient wants to achieve. The exposure starts with a situation less feared and is gradually increased to more anxiety arousing situations with prolonged periods of exposure until anxiety becomes extinct and habituation takes place. Besides its effectiveness, the treatment also has a number of drawbacks. First of all, therapists are not always in full control of the real situation. Also, arranging the exposure, e.g. flying as a passenger on a plane, can be time demanding, logistically difficult to set up and expensive especially as multiple exposure sessions are needed. Furthermore, the thought of being exposed to the situation they fear and normally avoid is so uncomfortable for some patients that they are unwilling to undergo treatment. Exposure in Virtual Reality (VR) is therefore seen as an alternative that might overcome these drawbacks, especially as recent meta-studies [8; 13; 14] indicate that exposure in VR is as effective as exposure in vivo. VR exposure in the treatment of fear of flying is now seen as an important, reliable technique to be used in the treatment of this phobia [5]. Besides its effectiveness, patients are more willing to be exposed in VR than in vivo. In a survey [7] among patients 76% preferred in VR exposure over in vivo exposure and refusal rate dropped from 27% to 3%.

Instead of focussing on the effectiveness of the treatment, this paper reports on how therapists conduct the Virtual Reality Exposure Therapy (VRET) in the treatment of fear of flying. A field observation is presented, analysing the interaction between therapists and VRET system, but also with the patient during an exposure session in VR. Before the field study is presented, the next section will give a brief introduction into the set up of the VRET system and the task of the therapist and the patient. The paper concludes with a number of design implications that are drawn from the observations.

BACKGROUND

The Dutch clinic where the therapists were observed used a VRET system that was developed by Delft University of Technology in collaboration with the department of Clinical Psychology at the University of Amsterdam. Besides the flight simulation, the system also includes worlds for the treatment of acrophobia (fear of heights), and claustrophobia. Figure 1 shows the communication between the patient, the therapist and the VRET system. The functional architecture of the Delft VRET system [18] was based on a task analysis of the therapist and the

patient established by interviews and observations in an university setting [15]. As therapist and patient have different task goals, the system also needs to support them differently. The patients' main goal is to get rid of their fear. To achieve this they follow the instructions of the therapist, however, they might occasionally try to avoid the feared situation to get rid of their fear only for the short term. Furthermore, they have to understand the treatment by asking questions about it. For exposure in VR to work, the patients need to have a feeling of being there (in the VR world), i.e. a feeling of presence. The type of display technology and locomotion techniques used in VRET systems can affect this feeling and patients' anxiety level [10; 16]. Still, increase in presence does not automatically also lead to treatment improvement [10]. Presence is not a key factor for therapists' task goal, which is to cure the patient. During the exposure session they monitor the patient's fear level, which is often done by asking patients to rate their anxiety on Subjective Unit of Discomfort (SUD) scale [20]. Based on this information therapists need to control the exposure and answer questions about the treatment patients might have.

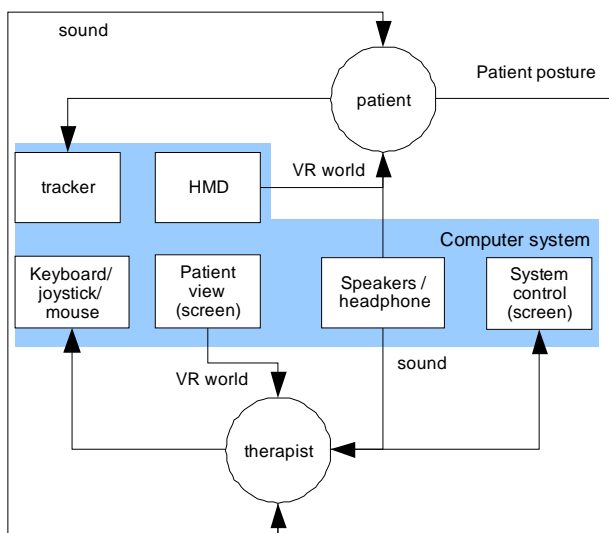


Figure 1. Communication between therapist, patient, and parts of the VRET system, adapted from Schuemie [15].

Therapists interact with the system using keyboard, joystick and mouse. Furthermore, they look at two screens: one displaying what the patient is seeing, the other screen (Figure 3) showing functions to control the system, such as patient information, flight plan, but also sound control and patient VR view. During the session patients wear a Head Mounted Display (HMD) with a six degrees of freedom tracking system. Furthermore, the patient sits in an actual airplane chair, which vibrates during the session to simulate the movement and trembling of the airplane. The vibration will increase especially during take-off, turbulence and landing. The chair is positioned next to a part of the airplane cabin. The therapists are positioned behind a table

facing the patient, with in front of them a monitor that shows what the patient is seeing and another monitor that shows the therapist console to control the VR simulation (Figure 2).

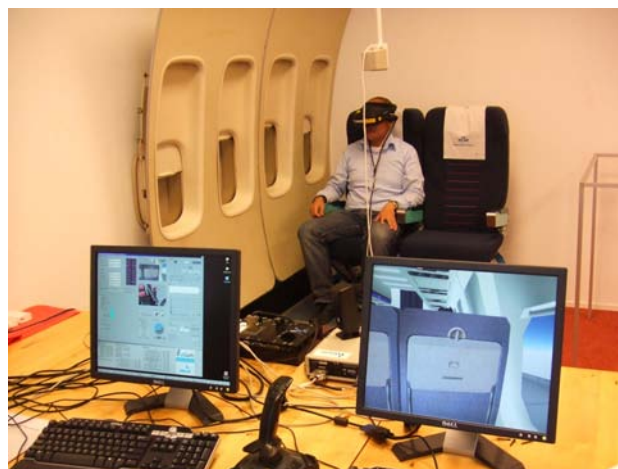


Figure 2. Set-up of VRET system in the treatment of fear of flying.

The design of the therapist user interface (Figure 3) was the result of a number of design iterations including usability evaluations [9]. Its main widgets are: Session information control (A) to enter session and patient information; Flight plan control (B) to set destination, time of day etc of the flight; Simulation control (C) to start or stop the simulation; Flight control (D) to set the stage of flight; Map control (E) to select the patient's seat; Patient view (M) to monitor what the patient is seeing in the VR world; Free view (N) to monitor the patient projected in the VR world; Cabin control (F) to set cabin light, seat belt light, and open and close window shutters; Roll control (G) to tilt the airplane; Flight view (L) to see the current stage of the flight; Note/SUD score (K) to enter comments and to record SUD scores; Time (J) to set the timer of the SUD alarm; System status (I) to monitor network connection; and Sound control (H) to play sound recordings such as purser or pilot announcements, or bad weather recordings. The therapists interact with these widgets by using a mouse and a keyboard.

METHOD

In 2006 the VRET system was installed at a Dutch clinic. Two years later, however, news arrived that some therapists were uncomfortable using the system as it had malfunctioned on some occasions. The system was repaired, and to build therapists' confidence again a researcher would be present in a number of sessions as technical assistant repairing the system on the spot if needed. It was soon realised that the researcher was in a unique position to make field observations of the interaction between on one side the therapist and on the other side the VRET system and the patient.

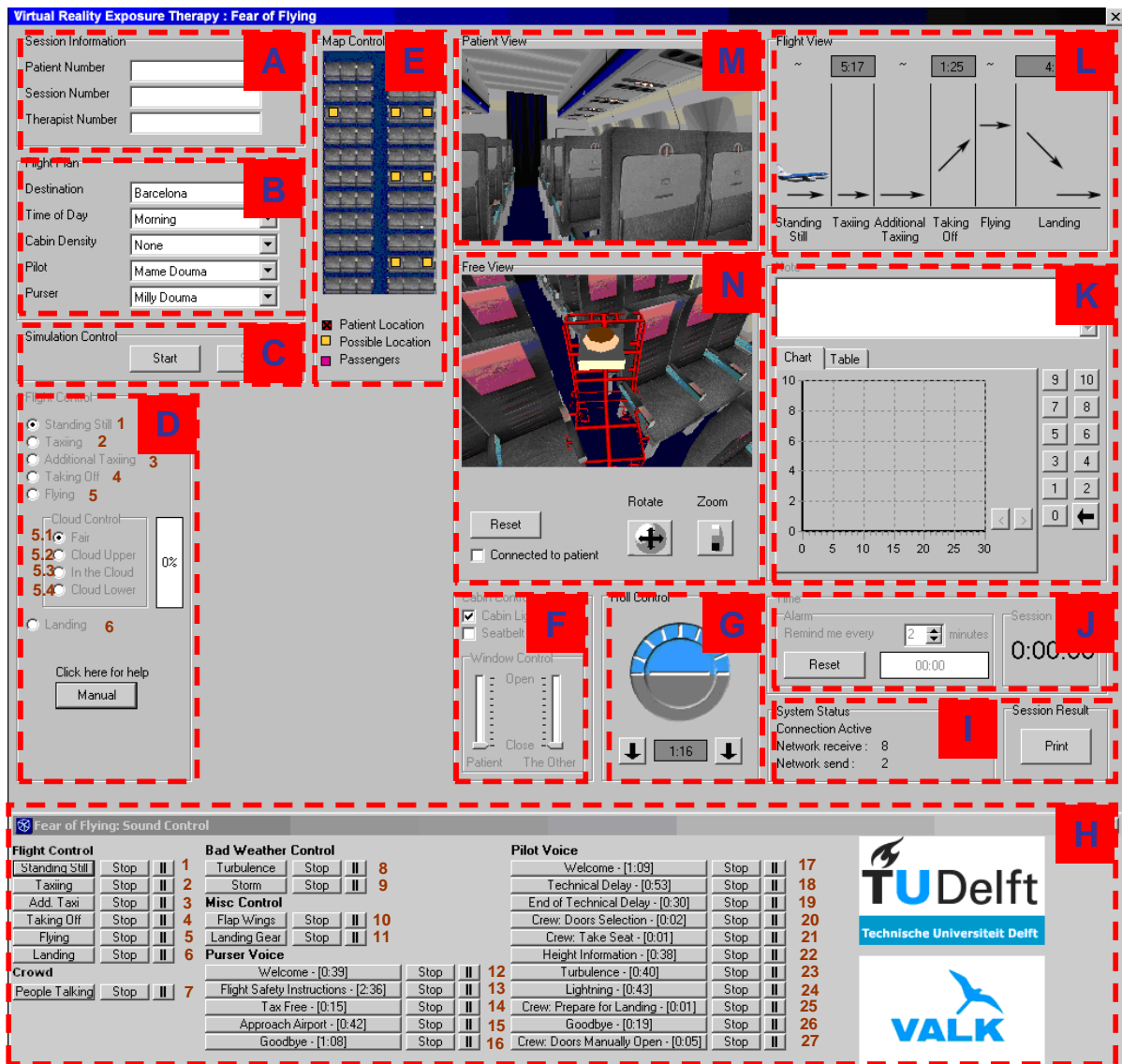


Figure 3. Screen coding scheme of the therapist console.

Participants

Six therapists working in the clinic participated in the field observations. One of the therapists was also a pilot. The clinic is specialised in the treatment of aviation related anxiety. The clinic not only treats fear of flying in passengers, but it also helps cockpit and cabin crew for all types of mental health problems. Both patients new to a VR exposure and patients with prior VR exposure experience were included in the observations.

Procedure

During the session the observer sat beside the therapist at the table with the two screens of the therapist console (Figure 2). During the session the observer made recordings of his observations, and when needed asked the therapist for clarifications after the session once the patient had left.

Material

All recordings were made with pen and paper and to ensure patient's privacy even further, no identifiable references to patient identity were recorded. To facilitate the event sampling, a coding scheme (Figure 3) was created which uniquely identified the interaction elements of the user interface. Each interaction element received a letter, extended with a number in some cases to identify specific buttons. The coding scheme allowed the observer to quickly make a record of any observed interaction in his log. Besides the interaction events, the phase of the flight was recorded, the length of the phase, and the comments made by the patient or the therapist, include requests for a SUD score.

RESULTS

Prior to the exposure, patients had an intake interview. Here the therapist also trained patients in a number of relaxation exercises which they could use during the exposure session. Similar to other reports [15; 19] in which the VRET system

was used, at the start of the exposure session, the therapist introduced the patient to the VRET system, explained the use of the HMD and how to adjust the helmet for their head size, eye strength and position of their eyes. After this the VR world was started and calibrated with a joystick. A flight plan was then selected e.g. destination: Paris, time day: morning, Cabin density: moderate, pilot: Mame Douma, and purser: Milly Douma. Patients were often located at a seat next to a window where they could see the airplane wing. The window shutters were opened, after this the flight simulator was started with the plane in the standing still stage of the flight. During the simulation the therapists initiated the different stages of the flight (taxiing, taking off, flying, landing) and played various sound recordings such as announcements from the purser or the pilot, but also sound from the airplane (flapping wings, retracting landing gear), or caused by the weather condition (turbulence or storm). During the session therapists monitored the patients and their anxiety (e.g. tense muscles, crying), thereby with intervals asking for SUD scores. In some sessions patients were also asked to do their relaxation exercises during the exposure.

Data Preparation

In total 23 VR sessions were observed. However, only 14 sessions were included in the analyses as four VR sessions were with acrophobia worlds, two VR sessions involved a VR simulation of an airport and not an airplane, and in three sessions the recording was incomplete. The 14 sessions included 11 different patients and six therapists. On average each session took 25 minutes. Some patients had two sessions immediately after each other. However, they always had at least a small break in between the sessions to avoid simulation sickness.

Figure 4 shows a part of the log recording. At 11:39 two buttons were pressed, H8 (turbulence) and H23 (pilot announcement of turbulence). One minute later at 11:40, the therapist asked a SUD score, which the patient replied with a score of eight. The patient also commented that he/she was very curious, and was looking around because he/she liked to know what was going on. At the same time the therapist moved the plane above the clouds (D5.4). Two minutes later, at 11:42, the therapist changed to a more or less cloud free weather condition (D5.1, flying fair) stage and gave the patient an exercise to relax the muscles and asked the patient to pay attention to his/her breathing. This and all the other written logs were coded and entered into spreadsheets for further analysis.

....
11:39	H8, H23	
11:40		SUD: 8, P: very curious, looks a lot around, wants to know what is going on
11:40	D5.4	
11:42	D5.1	T: 'muscles are a bit tense; contract them a bit more and then let go to relax. Notice your breathing'
....

Figure 4. Short part of the field-recording log (P: patient, T: Therapist).

Event Sampling Results

Table 1 gives an overview of the mean number of events observed per therapist. Notable is the relative high level of

interaction with the VRET system. On average therapists made 45 (SD = 8.7) mouse clicks. Looking at the interaction therapists directly had with a patient either by asking a SUD score ($M = 7.6, SD = 2.4$) or making a comment ($M = 1.1, SD = 1.2$), this was significantly ($t(5) = 13.8, p < 0.001$) lower than their interaction frequency with the VRET system. Although a high interaction frequency with the patient during an exposure might be undesirable as it might affect their feeling of presence, a high interaction frequency with the VRET system seems undesirable as well. This was also confirmed in the discussions with the therapists after the exposure sessions. They indicated that the system was at times demanding too much of their attention, and blame this on the design of the user interface, with its 'extensive number of buttons' as they put it. Asking for a SUD score with an average interval of 3.6 minutes was significantly below the often reported [2; 19] five minutes. However, the use of a two minutes interval [11], or a three minutes interval [15] have also been reported. As Figure 3 shows, the alarm is set to go off every two minutes, and none of the therapists seems to have changed this setting as the mean interaction frequency with the time control (J) was zero (Table 2). When the alarm was triggered the background of the screen flashed a number of seconds. However, the therapists were not aware of this. Most of them thought that this was simply a hardware malfunction of the screen. Furthermore, in a usability evaluation [9] conducted in 2002, participants also mention not to like the SUD reminder.

Table 1. Frequency of events and session time (average session results).

	Therapist						Mean
	A	B	C	D	E	F	
Session(s) observed	1	1	2	3	3	4	
SUD asked	7	5	6	8.3	7	12	7.6
Patient's comments	8	3	2.5	1.3	0.7	3	3.1
Therapist's comments	0	3	0	1	0.7	2	1.1
Perform exercises	0	2	0	1	0.3	1.3	0.8
Repeated phases	1	2	0.5	1.3	3.3	3.3	1.9
Mouse click	42	37	42.5	42.3	42.3	62	44.7
Voice announcements	10	8	10	9.3	9.3	12.5	9.9
Session length (min)	30	24	27.5	21.7	22.3	25.8	25.2
SUD interval (min)	4.3	4.8	4.6	2.6	3.2	2.1	3.6

Most of the interaction with the system involved playing sound recordings (Table 2). Followed by the interaction with the flight control, which is used to set the phase of the flight and allow the plane to fly below, in or above the clouds. Some elements were rarely used or only by a few therapists. For example, only one therapist used the roll control. This therapist was also a pilot, and probably had more experience in using more advanced options of the simulator, or had a more in-depth understanding of the aircraft's behaviour. Furthermore, this therapist, with his 62 mouse clicks, had an interaction frequency far above the average of 45 mouse clicks.

None of the therapists use the print option (I) as also no printer was attached to the system. This seems unfortunately as this function was previous rated as very useful [9]. None of the

therapists used the note taking facility. The therapists avoided using a keyboard during the exposure as the typing sound might distract the patient. Furthermore, as the system was stand-alone without a printer, therapist had also no access to the computer notes afterwards in their office. Instead therapists wrote their comments on the patient’s paper form. No interaction with the Free View panel was recorded. Although Schuemie’s guidelines [15] recommend that therapists should be offered this view, it might be more useful in VR settings where the patient actually moves through a virtual world for example in the treatment of acrophobia where patients walk up to an edge of a roof terrace [15].

Table 2. Frequency of therapist interaction with VRET system.

Screen element	Mean	SD
A- session info	1.0	0.1
B- flight plan	1.0	0.1
C- simulation control	1.0	0.1
D- flight control	9.2	1.1
E- map control	1.0	0.1
F- cabin control	1.2	0.4
G- roll control	0.3	0.6
H- sound control	22.7	5.6
H- flight control	5.1	0.9
H- crowd	1.1	1.2
H- bad weather	2.0	0.7
H- misc control	4.7	2.6
H- purser voice	4.1	0.7
H- pilot voice	5.8	1.0
I- system status	0.0	0.0
J- time	0.0	0.0
K (SUD)	7.6	2.4
K (notes)	0.0	0.0
N- free view	0.0	0.0
<i>Total</i>	44.7	8.7

State Sampling Results

During the observation a record was kept of the stages (phase) of the flight: standing still (S), taxiing (T), additional taxiing (A), taking off (O), flying (F), flying fair (F1), flying below clouds (F2), flying in clouds (F3), flying above clouds (F4), and landing (L). Examining Table 3 quickly shows a consistent starting pattern of standing still, taxiing, taking off, flying, and landing. If no distinction is made between taxiing and additional taxiing and in the different flying phases, 93% of the observations had a similar begin pattern of STOF (Table 4). For longer patterns less similarity was found, with two observations that were extended with an additional standing still (STOFSL) phase or with a taxiing and taking off phase (STOFLTO). Interesting is that only in two observations the

therapist went from a landing phase to a stand still phase. Apparently, the landing was often regarded as the last phase, ignoring the fact the plane has to come to a complete standstill before, for example, the doors could be opened. However, this idea might not have been reinforced by the design of the system as in the flight control panel (D) the landing phase was at the bottom of the list (Figure 3).

Table 3. Sequence of flight phases.

Therapist	Sequence of phases
A	S T A O F1 F2 F3 F4 F1 L
B	S T O F1 F2 F3 F4 F1 F4 L
C	S T A O F1 F4 F3 F2 F1 L S T A O F1 F2 F3 F1 L
D	S T O F1 F2 F3 F1 L S S T O F1 F2 F3 F4 F1 L S T O F1 F2 F3 F1 L
E	S T O F1 F2 F3 F4 L S T O F1 F2 F3 F4 F1 L S T O F1 F2 F3 O F1 F2 F3 T O F1 F2 F3 L
F	S T A O F1 F2 F3 L T O O S T O F1 F2 F3 F4 F1 L T O F1 F2 F3 F4 F1 L S T O F1 F2 F3 F4 L S T O F1 F2 F3 F4 F1 L S

Table 4. Similarity in flight phase patterns.

Start pattern	Match
STOF	100%
STOFL	93%
STOFSL	14%
STOFLTO	14%

Figure 5 shows a transition diagram of the phases in a flight. Again pattern STOF (L) can be seen as the dominate path therapists followed in the sessions. The diagram also shows only a small number of variations in the phase transitions, for example, after flying (F) taking off (O) again, or going back to taxiing (T) and to taking off (O) again. This was observed in the last session of therapist E. In the previous session, the patient had shown a high level of anxiety during take offs. By exposing the patient multiple times to this stage of the flight the therapist aimed at habituation of the fear situation resulting in a lower level of anxiety.

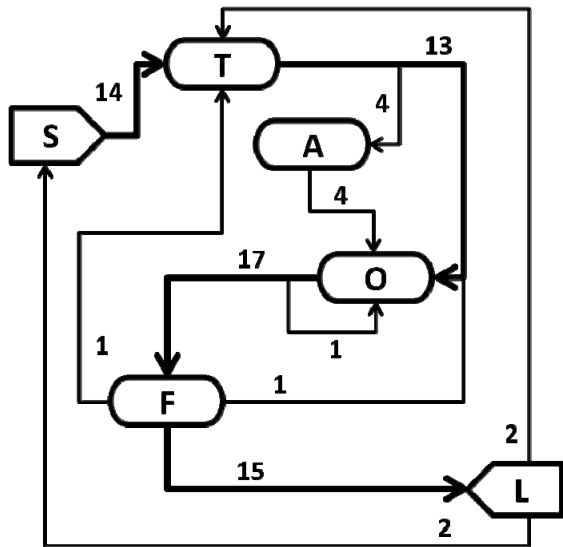


Figure 5. Transition diagram of phases and the frequency of the phase transition observed (S – standing still, T – taxiing, A – additional taxiing, O – taking off, F – flying, and L – landing).

The patterns of various flying phases were also analysed. As Figure 6 shows exposure often included the transitions from flying fair, to flying below the clouds, to flying in the clouds, to flying above the clouds, and finally to going back to flying fair. This was often followed by a landing phase.

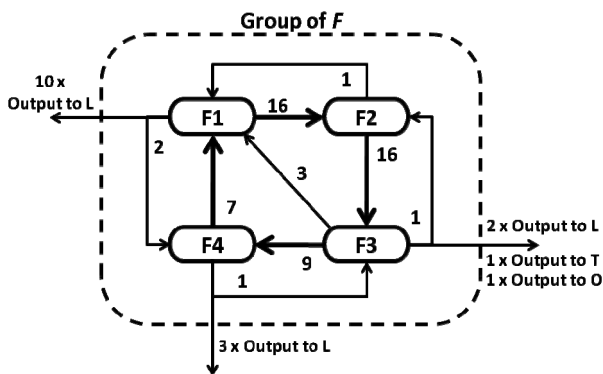


Figure 6. Transition diagram of flying phases and the frequency of the phase transition observed (F1 - flying fair, F2 - flying below clouds, F3 - flying in clouds, and F4 - flying above clouds).

Examining the therapists’ interaction with the VRET system per phase (Table 5), the flying phase had the highest level of interaction ($M = 15.7$) and made up the largest part of the exposure with an average of 9.2 minutes. Still, looking at the average interaction frequency per minute across the phases, this was below two per minute ($M = 1.7, SD = 0.4$).

Table 5. Frequency of interaction events with VRET system, phase time, and interaction per minute averaged over sessions.

Phase	Freq. Interaction	Time (min)	Interaction per min
Standing still	4.5	2.2	2.0
Taxiing	7.1	4.1	1.7
Add. Taxiing	0.8	0.8	1.0
Taking Off	5.6	2.7	2.1
Flying	15.7	9.2	1.7
Flying fair	6.8	3.4	2.0
Flying below clouds	2.5	1.3	1.9
Flying in clouds	5.2	3.2	1.6
Flying above clouds	1.2	1.3	0.9
Landing	10.1	5.3	1.9

Errors

In one of the updates of the system, a sound control panel had been added to the therapist user interface as a patch to extend the simulation with more sound recordings (e.g. flight safety instructions, and people talking at the background). To reduce redundancy the sound panel in the original user interface was hidden with a grey panel (Figure 3, right side of element D). However, the original user interface was designed with error prevention in mind. The system only allowed therapists to select sound recording that were appropriate for the current stage of the flight. With the new sound panel therapists could play sound recordings at any moment. Table 6 shows that during the 14 sessions, therapist played 20 inappropriate sound recordings. For example, on six occasions, they played the pilot announcement asking the crew to open the doors while the plane has not come to a complete standstill yet, or on two another occasions the pilot welcome announcement was played while the plane was taxiing. In reality, however, pilots are often occupied during taxiing for example communicating with the tower, and therefore will make such announcements before taxiing. Furthermore, in his welcome announcement the pilot also mentioned that the luggage was being loaded on board. This example clearly illustrates that there might be several reasons why therapist make these errors. First, they might not be aware of the content of the announcement. Second, they might not have an accurate mental model of a flight. Third, they might have an accurate mental model, however, they might have thought the flight to be in another phase, in other words a mode error [12]. Fourth, therapists might have problems with fitting a sound recording into the timeslot of the phase thereby overshooting the phase or by anticipating on this, playing the sound recording too early. Interesting in this context are the observations of the therapist who was also a pilot. Four errors were also observed in his sessions, for example giving height information (H22) while taking off. This makes it less likely that an inaccurate mental model of a flight can simply explain all errors. Still in all of this, it is important to consider that there were no indications that any of these errors had a negative effect on the treatment.

Table 6. Errors made by playing sound recordings.

Phase	Voice Announcement / Aircraft sound	Freq
Standing still	<ul style="list-style-type: none"> ▪ Purser- Flight safety instruction (H13) too early; is normally issued during taxiing. 	1
Taxiing	<ul style="list-style-type: none"> ▪ Pilot- Welcome (H17) too late; pilot too busy during taxiing to make announcements, also the pilot mentioned in his welcome that the luggage is being loaded. 	2
Taking off	<ul style="list-style-type: none"> ▪ Purser- Welcome (H12) too late; during take off purser is sitting down and will not make announcements. 	1
	<ul style="list-style-type: none"> ▪ Pilot- Crew: door selection (H20) too late as doors should have been closed before take off 	1
	<ul style="list-style-type: none"> ▪ Pilot- Crew: take seat (H21) too late as the crew should already been sitting 	2
	<ul style="list-style-type: none"> ▪ Pilot- height information (H22) too early; plane is still climbing 	1
Flying	<ul style="list-style-type: none"> ▪ Landing gear sound (H11) too early/late; should be retracted while climbing after take off, or extended just before landing 	1
	<ul style="list-style-type: none"> ▪ Pilot- Crew: take seat (H21) too early; should be issued just before starting the landing 	2
	<ul style="list-style-type: none"> ▪ Pilot- Crew: prepare for landing (H25) too early; should be issued just before the landing 	1
	<ul style="list-style-type: none"> ▪ Landing gear sound (H11) too early; should only be extended just before landing 	1
Landing	<ul style="list-style-type: none"> ▪ Purser- tax free (H14) too late; should be announced while flying 	1
	<ul style="list-style-type: none"> ▪ Pilot- Crew: door manually open (H27) too early; should be announced after complete stand still 	6
<i>total</i>		20

DISCUSSION AND FINAL REMARKS

Two main conclusions can be drawn from these observations. First, therapists perform dual tasks, interacting with the patient and with the VRET system. Reducing their task load might be possible by reducing the need for frequent interaction with the VRET system, as this is currently relatively high compared to the interaction frequency with the patient. Secondly, the current set up allows therapists to make unnecessary errors. For the exposure, therapists now need knowledge about the treatment and the patient, but also about flying and about running the VR simulation. Especially the need for the last two should be minimized allowing the therapists to focus on their main task and that is to cure the patient.

The observations also have a number of design implications. (1) Because of the consistency in the sessions it might be possible to develop a treatment oriented instead of the simulation oriented user interface for the therapist, taking the sequence of flight phases as a starting point. For example, in each phase, inappropriate simulation elements could be hidden to avoid errors. (2) To reduce system interaction frequency, to extent the variation in the flights, and to improve the realism of the experience, it might also be possible to provide therapists with several automated flight simulations scenarios (for example good or bad weather flight, short or long taxiing). In these scenarios the simulation runs automatically, applying the appropriate flying routines, but still allows therapists to control when to move to the next phase, or change to another scenario altogether if required because of the patient’s response. Furthermore, the system should also support therapists if they like to deviate from the standard flight sequence. For example, expose patients to multiple take offs if needed.

Based on the reported field observations, the therapist user interface is now being redesigned. Besides the automatic flight scenarios the redesigned user interface now also includes better support for notes taking, whereby therapist can select predefined comment flags that are placed on a single timeline overview of the flight. The automation might reduce part of therapist’s task load. Therefore, preliminary work has also started whether a therapist can simultaneously give multiple patients an exposure in VR. Still, with all these new research directions it will be important to keep in mind the lessons learned from these field observations about the dual task therapists are performing and that the system should be designed to avoid errors.

ACKNOWLEDGMENTS

We would like to thank the therapists of the VALK Foundation for their participation in the field observations.

REFERENCES

- [1] APA, Diagnostic and statistical manual of mental disorders. American Psychiatric Association, Washington, DC, 2000.
- [2] Banos, R. M., Botella, C., Perpina, C., Alcaniz, M., Lozano, J. A., Osma, J. and Gallardo, M. Virtual reality treatment of flying phobia. *IEEE Trans Inf Technol Biomed*, (3)6, 2002, 206–212.
- [3] Becker, E. S., Rinck, M., Turke, V., Kause, P., Goodwin, R., Neumer, S. and Margraf, J. Epidemiology of specific phobia subtypes: findings from the Dresden Mental Health Study. *Eur Psychiatry*, (2)22, 2007, 69–74.
- [4] Curtis, G. C., Magee, W. J., Eaton, W. W., Wittchen, H. U. and Kessler, R. C. Specific fears and phobias. *Epidemiology and classification*. *Br J Psychiatry*, 173, 1998, 212–217.
- [5] Costa, R. T. da, Sardinha, A. and Nardi, A. E. Virtual reality exposure in the treatment of fear of flying. *Aviat Space Environ Med*, (9)79, 2008, 899–903.
- [6] Emmelkamp, P. M. G. Behaviour therapy in adults. *Bergin and Garfield’s Handbook of psychotherapy and behavior change*. Lambert, M. (Ed.). Wiley, New York, NY, 2004. Pp. 393–446.
- [7] Garcia-Palacios, A., Botella, C., Hoffman, H. and Fabregat, S. Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients

- with specific phobias. *Cyberpsychol Behav*, (5)10, 2007, 722–724.
- [8] Gregg, L. and Tarrier, N. Virtual reality in mental health: a review of the literature. *Soc Psychiatry Psychiatr Epidemiol*, (5)42, 2007, 343–354.
- [9] Gunawan, L. T., Mast, C. van der, Neerincx, M. A., Emmelkamp, P. and Krijn, M. Usability of therapist's user interface in virtual reality exposure therapy for fear of flying. 10th Annual Scientific Conference on Web Technology, New Media, Communications and Telematics Theory, Methods, Tools and Applications. Schreurs, J. and Moreau, R. (Eds.). Eurosis, Hasselt, Belgium, 2004. Pp. 125–132.
- [10] Krijn, M., Emmelkamp, P. M., Biemond, R., de Wilde de Ligny, C., Schuemie, M. J. and van der Mast, C. A. Treatment of acrophobia in virtual reality: the role of immersion and presence. *Behav Res Ther* (2)42, 2004, 229–239.
- [11] Muhlberger, A., Herrmann, M. J., Wiedemann, G. C., Ellgring, H. and Pauli, P. Repeated exposure of flight phobics to flights in virtual reality. *Behav Res Ther*, (9)39, 2001, 1033–1050.
- [12] Norman, D. A. Categorization of action slips. *Psychological review*, (1)88, 1981, 1–15.
- [13] Parsons, T. D. and Rizzo, A. A. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *J Behav Ther Exp Psychiatry*, (3)39, 2008, 250–261.
- [14] Powers, M. B. and Emmelkamp, P. M. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *J Anxiety Disord*, (3)22, 2008, 561–569.
- [15] Schuemie, M. J. Human-computer interaction and presence in virtual reality exposure therapy. Delft University of Technology, 2003.
- [16] Schuemie, M. J., Abel, B., Mast, C. A. P. G. van der, Krijn, M. and Emmelkamp, P. M. G. The effect of locomotion technique on presence, fear and usability in a virtual environment. EUROMEDIA '2005: 11th Annual Euromedia Conference 2005. Pp. 129–135.
- [17] Stinson, F. S., Dawson, D. A., Patricia Chou, S., Smith, S., Goldstein, R. B., June Ruan, W. and Grant, B. F. The epidemiology of DSM-IV specific phobia in the USA: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychol Med*, (7)37, 2007, 1047–1059.
- [18] Mast, C. A. van der, Technological challenges in the Delft virtual reality exposure therapy system. *International journal on disability and human development*, (3)5, 2006, 205–210.
- [19] Wiersma, J., Greeven, A., Berretty, E., Krijn, E. and Emmelkamp, P. De effectiviteit van Virtual Reality Exposure Therapy voor hoogtevrees in de klinische praktijk. *Gedragstherapie*, 41, 2008, 253–259.
- [20] Wolpe, J. *Psychotherapy by reciprocal inhibition*. Stanford University Press, Stanford, 1958.

User Expectations for Mobile Mixed Reality Services: An Initial User Study

Thomas Olsson¹, Pirita Ihamäki¹, Else Lagerstam¹, Leena Ventä-Olkkonen²,
Kaisa Väänänen-Vainio-Mattila¹

¹Tampere University of Technology, Unit of
Human-Centered Technology
Korkeakoulunkatu 6
P.O. Box 589, 33101 Tampere, Finland
{thomas.olsson, pirita.ihamaki, else.lagerstam,
kaisa.vaananen-vainio-mattila}@tut.fi

²Nokia Research Center, Oulu
Yrtypellontie 6
90230 Oulu
leena.venta-olkkonen@nokia.com

ABSTRACT

Mixed reality, i.e. the integration and merging of physical and digital worlds, has become an integral part of the ubicomp research agenda. Often, however, in development of first technology concepts and prototypes, the expectations of potential users are not considered, and the development easily becomes technology-driven. To understand the expectations and needs of potential users of future mobile mixed reality (MMR) services, we conducted altogether five focus group sessions with varying user groups. We investigated the early impressions and expectations of MMR as a technology by evaluating various usage scenarios. Based on this initial study, we found relevance issues (what information to receive, how and when) and the reliability of MMR information to be the most salient elements that were anticipated to affect the overall user experience. In mobile and busy situations the MMR information content has to be something that is very important or useful for the user, especially if receiving the information or interacting with it draws the user's attention away from the tasks executed in the real world.

Keywords

mobile mixed reality, augmented reality, context awareness, mobile services, user expectations, user studies, scenarios

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)] User Interfaces – *user-centered design*; H.5.1 [Multimedia Information Systems] – *Artificial, augmented, and virtual realities*; H.1.2 [User/Machine Systems] – *Human factors*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

INTRODUCTION

The concept of *mixed reality* refers to the integration and merging of the real and virtual worlds where physical and virtual objects complement and interact with each other [3]. Broadly defined, mixed reality is understood to cover the extensive continuum between the two opposite, discrete ends of reality and virtuality [2, 20]. In practice, it is often implemented as augmented reality (AR), where the real world is augmented with digital (virtual) information, and augmented virtuality (AV), where a virtual world is augmented with real-world information. With AR, perception of the user's environment can be enhanced, enriched and be made more transparent to the surrounding data (e.g. information, advertising and resources related to places, objects and situations). In this paper, we focus especially in *mobile mixed reality* (MMR), particularly with means of augmented reality on mobile devices.

A central use platform of mixed reality is the mobile domain which expands the mixed reality services to cover a diverse set of use cases and scenarios in the mobile environments. Due to the rapid development of sensor and communication technology, mobile devices are becoming more and more aware of their environment, user's context and information resources near-by. Hence, mobile devices have become a fruitful platform for creating and interacting with mixed reality objects and services. A predominant interaction paradigm in the mobile domain is the magic lens [2] where the user may browse the world via the camera view, and access additional information aligned on top of the view. Also data glasses and other head-mounted displays can be used to create a highly immersive experience of the mixed reality environment. Although AR is very visual by nature, mobile devices can enrich the augmentation of real world also with auditory and haptic information (e.g. informing about information affordances in the environment by haptic cues [12]).

For the last decade in HCI, there has been a prominent interest in studying the user experience (UX) of various products and services. It is regarded as a subjective and holistic concept including both instrumental (e.g. utility and usability) and non-instrumental (e.g. joy, appeal, aesthetics) elements that result from the use of a product [10]. The experience evolves over time as user's previous experiences affect the overall experience of the product [10, 15]. If provided at the right time and in the right place, MMR information can assumedly offer the user rich, useful, delightful and positively surprising experiences. However, it is still to be studied how does the holistic, temporally-evolving and subjective user experience of

MMR build up. In mobile contexts the use situations vary from safety critical, urgent and demanding tasks to free-form leisure-oriented activities. Hence, it is critical to understand in which situations the user may be offered extra information, and in which situations the user must not be interrupted.

So far the research on mixed reality has mostly focused on development of enabling technologies: various types of displays, techniques using eye-tracking, auditory, haptic and gesture-based interaction, as well as algorithms and tools for modeling virtual objects and mapping them on top of the real world view. Nonetheless, the user experience and acceptance perspective of mixed reality services has been studied very little. Such research has mostly focused on usability issues of individual demonstrators. The user-centered design approach is based on understanding the requirements of the user, context, and tasks, but the needs and expectations of the users have not been studied as a starting point of mixed reality. Furthermore, the application areas of the existing MMR applications have been mostly working environment (e.g. remote controlling systems in process automation or in design of 3D objects). The leisure and everyday mobile contexts have not been studied extensively. Being immensely rich by nature, the mobile context provides a basis for very diverse set of use cases where MMR technologies could be utilized.

The context of our study is the DIEM (Devices and Interoperability Ecosystem) project, which aims at building new kind of smart environments that comprise of ecosystems of digital devices [6]. One of the key development areas of DIEM is MMR. The aim of our study was to understand the initial expectations of potential users of future MMR services: what content there could be, and in which kind of contexts it could be successfully used. Thus the study will help in projecting and assessing the early prospects of MMR application areas.

RELATED RESEARCH

We first present a few studies relevant of MMR services and applications, and secondly introduce the theoretical background of user experience (UX).

Mixed Reality Systems

We present related research on applications and system prototypes that utilize design ideas related to MMR. These systems emphasize characteristics of AR (augmented reality), not AV (augmented virtuality). Many of these applications are some kind of games or guide applications, but there are also other ideas of systems that can be useful in everyday life. Also other relevant concepts, e.g. related to context awareness and sharing information are presented.

Ludford et al. [19] have developed a location-based reminder (LBR) system called PlaceMail. The system runs on a GPS-equipped mobile phone. People use it to create and receive personal task reminders. PlaceMail utilizes the phone's location-sensing GPS to deliver the message when the user is near the place. As people use LBRs, they generate local place data, like lists of places they go for everyday tasks (or place bookmarks) and reminder messages related to the bookmarked locations.

Gleue and Dähne [8] present the ARCHEOGUIDE (Augmented Reality-based Cultural Heritage On-site GUIDE) project, which provides cultural heritage sites with archeological information. With a small mobile computer and a display unit visitors are able to experience the real site while appreciating visualizations of the virtual reconstructions integrated seamlessly into the

natural field of view. The mobile device tracks the user's position on the site. The ARCHEOGUIDE system is able to compute the current view of the reconstructed objects by determining the viewing direction.

Herbst et al. [13] present a mobile outdoor mixed reality game for exploring the history of a city in the spatial and the temporal dimension. The story of the game called Time Warp is based on the legend of the Heintzelmannchen of Cologne. The legend tells that one day these Heintzelmannchen disappeared and the goal of the game is to bring them back. The game aims to fuse real and virtual elements to create the illusion that users are present in the City of Cologne during different time periods.

Flintham et al. [7] describes two mobile games in which online participants collaborated with mobile participants on the city streets. The first one, Can You see Me Now? (CYSMN), was designed to be a fast-paced game in which up to twenty online players were chased across a map of the city, by three runners who were moving through the actual city streets. The main goal of CYSMN was to engage and excite the online players by giving them a sense of the runners' experience of the city, and of how their online actions could affect events on the streets. In the second game, bystander, a local player takes a journey through the city on the trail of a mysterious person whose name and picture they have quickly been shown. An online performer collaborates as a partner with them and guides them in the search. Between them the two participants travel through the city streets at the same time and across an online map in search of the mysterious target person.

Brown et al. [5] present the co-visiting system which allows three people to visit the Interpretation Centre simultaneously, one physically and two digitally. The physical visitor is in the Interpretation Centre itself with special mobile equipments including location system. The virtual reality visitor uses 3D display with avatars representing the other visitors. The web visitor for one uses a standard web browser displaying several Java applets, one of which is a variant of the physical visitor's map. The idea is to support looking at exhibits as a group of physical visitors, virtual reality avatar visitors and web visitors. These three kinds of visitors share an exhibition space represented via different user interfaces.

Theoretical Background of User Experience

User experience (UX) is often understood to cover more or less all the aspects that affect or result from end-users' interaction with a technological product or service. Hassenzahl and Tractinsky [10] define UX as "a consequence of a user's internal state (e.g. predispositions, expectations, needs, motivation, mood), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.)" Furthermore, a definition by Kankainen [15] describes the temporally evolving nature of user experience: "[...] result of a motivated action in a certain context. User's previous experiences and expectations influence the present experience; this present experience leads to more experiences and modified expectations". Hassenzahl [11] describes user experience to further involve aspects such as joy, hedonism and ludic values, to complement the traditional utilitarian and pragmatic view present in traditional HCI research and usability. These have been complemented by introducing also social and cultural factors as elements that affect the overall experience [1, 4].

Studying user expectations of novel technology gives an approximation of user experience before practical example applications exist and users will have actual experiences with them. Similar studies have been successfully conducted in related literature, and provided useful results for further development of the technology, see e.g. [12]. It is vital to identify what the users expect the user experience to be like with such technology, what needs could they fulfill with the help of it, and what kind of requirements do they have for interaction with the technology. Secondly, to provide early feedback of technology being developed and to facilitate the user acceptance process, it is important to involve potential users of the future services early in the user-centered development process [14].

OUR USER STUDY

This section further specifies the goals of the study, and presents how we applied focus groups as a research method, as well as the background of participants and what kind of scenarios were used as stimuli.

Study Objectives

For developing successful and acceptable mixed reality services, it is vital to understand the potential users' subjective needs with regards to that kind of services, as well as the expectations towards user experience of them. We set our focus on *location-based mobile* augmented reality information and services. We aimed at 1) identifying and innovating potential use cases for location-based MMR services, 2) finding out what kind of AR information users would value and need, and 3) inquiring the users' needs and expectations (e.g. potential benefits and drawbacks) for MMR services. Hence, the study setting was twofold: evaluating existing (technology-driven) concepts, and identifying needs for new ones. As this was our initial study to deal with user expectations of MMR services, the research was highly explorative and by nature.

Study Methodology

There are many definitions of a focus group in the literature, but aspects like *organized discussion* [16,17], *collective activity* [22], *social event* [9] and *interaction* describe the elements that focus groups have as a form of social research in the HCI field. We chose focus group as our research method as it is excellent for the generation and evaluation of early ideas and facilitating rich discussion. This was a suitable approach as the holistic picture of MMR requires clarification and diverse aspects have to be considered from the users' perspective. Furthermore, one benefit of focus groups in studying early user expectations is their flexibility to adjust topics between sessions.

Participants

We conducted five focus group sessions with different types of user groups: 1) active travelers or tourists, 2) senior high school students, 3) technology oriented people, 4) wellness-oriented people, and 5) people with ecologically sustainable values. Most participants represented early adopters, as they were required to be at least somewhat interested in new technical products, and thus be potential first users for the MMR services. Incorporating various user groups in the study was intended to bring diversity in the user expectations and ideas. With such a small amount of users and representative groups we did not aim at drawing any conclusions of the differences between certain user groups. Participated users ages varied from 18 to 59, and 13 of the participants were male and 10 female.

Table 1. Background questions about participants' technical orientation and communication habits (1 "I strongly disagree" – 5 "I strongly agree") (N = 23).

Question	Mean	St. dev.
"I find technology useful in my everyday life."	4.4	0.7
"I am usually one of the firsts among my friends to get new technology."	2.7	1.0
"I like to help my friends and relatives to use technical devices."	3.9	1.2
"I like to edit information in Wikipedia or similar."	2.1	1.4
"I like to share information about me for example in Facebook or Irc-gallery."	3.3	1.3
"I would find it useful that my friends knew my location and what I'm doing."	2.7	1.2
"I'm worried about my personal information spreading in the web and getting in the wrong hands."	3.2	1.3

Focus Group Setup

Each focus group session had two parts. First, a rather free-form discussion about potential use cases was initiated with a general introduction to the theme of augmented reality in mobile domain. Next, discussion continued around certain predetermined scenarios that involved various potential MMR use cases. These were presented as textual narratives enriched with visual illustrations. Finally, to gather more structured data for consolidating the explorative and subjective discussion data, the participants filled in an expectations survey that consisted of statements about certain assumedly important elements of expectations. The survey was completed in the end of the session. Altogether, the gathered data was qualitative by nature.

Each session lasted 1–1,5 hours and had 4–6 participants. All sessions were recorded and the recordings transcribed. The explorative data was analyzed with a thematic analysis approach – objectively and systematically identifying common themes and meanings from the data and further categorizing them at a higher abstraction level [23].

Narrative Scenarios as Stimulus Material

We had prepared two textual scenarios describing potential use cases and contexts for MMR services: 1) a normal day in a familiar urban context, and 2) as a tourist in an unfamiliar context. Both scenarios were 600–700 words in length, and were divided in four sections to be more understandably presented to the participants. Each section was presented sequentially as a text on a paper sheet, and the context of the section was illustrated with the help of a data projector. Focus groups 1 and 4 discussed the tourism scenario and groups 2, 3 and 5 the "normal day" scenario. Figure 1 exemplifies one section of the "normal day" scenario (translated from Finnish), and Figure 2 presents an illustration of one of the sections.

"...Juhani steps into a bus and touches a digital info spot with his phone. This way he pays for his journey and also gets the bus company's BusNews-bulletin and services provided in that bus. These services include music lists created by other travelers and different kinds of mobile games. Juhani activates an info sticker on the backside of the seat in front of him and this opens the BusNews bulletin contents list on his cell phone screen. The icons and texts on the sticker become downright alive when Juhani browses the information content in the sticker with his cell phone..."

Figure 1. An excerpt of the "Normal day" scenario.



Figure 2. An illustration of one section of the Tourism scenario.

Table 1 exemplifies various use cases presented in the two scenarios. Due to paper length limitations, not all use cases are listed. The scenarios involved also other than purely MMR-specific aspects, such as context awareness and automatic information retrieval, which made the use cases more credible in practice. The scenarios were used merely as stimulus for discussion. The use cases or scenarios as whole were not directly evaluated in a structured way by the participants. Therefore, also no comparisons were made.

Table 2. Examples of use cases that especially bring out MMR aspects in the two scenarios.

“Normal day” - scenario	Tourism scenario
Pointing the mobile device to the digital info screen and studying other bus options.	User could browse information signs, e.g. history of attractions and info shared by other visitors.
Activating an info sticker on the backside of the seat in front of him and this opens the BusNews bulletin on his cell screen.	Getting additional information about interesting targets just by pointing the camera viewfinder on them.
Pointing the mobile to the interactive wall of his office and looking at his calendar.	A virtual tour guide recognizing photo objects and suggests shooting from a different angle.
Scanning medication products to the mobile device display. The device compares the medicines and gives tips.	Downloading the menu of the restaurant from an information sign and seeing highlights on the specialties of the place.
Activating movie posters with a mobile device to see virtual movie trailers.	Projecting pictures of how a historical place has looked for example 10 000 years ago.
The mobile device notifies about a virtual graffiti message left by a friend at a wall of a near-by house.	Afterwards seeing statistics on spent calories and the ecological footprint.
User draws a virtual graffiti by drawing in the air with his mobile device.	Setting the mobile device to an adventure mode, in which it automatically notifies about the places user has labeled beforehand.

RESULTS

First, we report the results of the focus group discussion that followed the MMR introduction and scenarios, and second the results of the summarizing expectations questionnaire.

Focus Group Discussion Results

In general, participants found MMR to be an interesting and versatile approach providing new ways of interacting with the information in the near-by environment. MMR services were seen to have the ability to provide richer, more relevant and even surprising information about everyday environments. Furthermore, MMR information that would not be otherwise acquired would be useful especially in unfamiliar contexts (e.g. while travelling). Overall, information for practical benefit was more desired than purely fun and leisure related information. The discussion focused more or less on augmenting the real world with *visual* information.

The most interesting MMR information with regard to everyday life was seen to be information concerning weather forecasts, public transportation (e.g. schedules and routes of various transportation methods) and location specific information of services (e.g. suitable restaurants within radius of 500 meters). Receiving augmented information automatically for example about nearby attractions would be useful not only for tourist but for locals too. Also real-time locations of various objects as augmented information was considered very intriguing, e.g. finding out when the next bus is coming and where the bus is at the moment. Furthermore, the aspect of social guidance was seen salient in future MMR service. The services could be used for pointing out like-minded people or members of one’s communities in the surroundings and inform about the information they recommend. This aspect of social awareness and recommendations came up in the discussion regarding both everyday and unfamiliar contexts.

“[...] I could just quickly browse what services there are within a few blocks” – Wellness oriented focus group

Regarding unfamiliar contexts and situations the participants considered that MMR services could help and support in cultural issues (e.g. translating short text or providing information about unfamiliar technical devices). Also, participants brought up the idea that they could acquire information regarding their own interests and tagged locations, such as attractions and history of the environment. However, they did not want to acquire too much information about the traveling resort before actually travelling. They saw that they might lose the joy of finding and exploring, and thus the main purpose of the vacation would be disturbed. Instead, the user could select certain targets and objects as points of interest that would be explored when in situ. Using MMR services in the nature was also brought up. The service could show where people have walked and provide navigational cues, extra information about certain sights and plants. In time, MMR services could replace the current information signboards.

As mentioned, the ideas for augmented information were mostly visual by nature. One participant’s idea was that MMR could serve as a tool in creating and visualizing 3D models on top of the real world information. Furthermore, the participants envisioned use cases for fitting clothes virtually with augmented garments. This was considered useful especially for physically challenged people. MMR could also serve as a tool for self-expression: augmenting one’s appearance with virtual hairstyles, makeup, or other adornment. Also, the idea of virtual house decoration and furnishing received much support (e.g. choosing colors or placing furniture in one’s apartment). Finally, most users expected the services to be based on maps (both indoors and outdoors), as the interaction with digital maps has become so prevailing in current mobile services.

Personalization and Information Relevance

Regarding user expectations, one of the key findings was the prominent need for personalizing the services. With this we not only mean personalizing the user interface but also personalizing the service features to be able to offer more relevant and meaningful information and services for the user. The physical devices did not have a big role in the discussion of personalization: the appearance of the device and user interface was not emphasized. Instead, the information content and interaction with it were considered more essential.

“When starting to use any device, no matter whose it is, it would become personalized easily with your own information”
– Technology oriented focus group

Personalization was seen as a tool to help finding interesting and relevant services and people and automatically receiving information that one would not have been actively searching for. The discussion repeatedly emphasized the importance of being able to limit the amount and type of information that is acquired with the means of MMR. If personalization is not possible, information flood, dangerous situations, or frustration of use might take place. Especially in case of advertisements, the need for personalization was seen to become even greater.

The relevance of the information was found to be highly dependent on the situation and tasks the user is pursuing. For example, during spare time and in unfamiliar environments one might like to receive lots of various types of information even automatically, whereas in repetitive situations or in hurry getting extra and irrelevant information would disturb and interrupt important tasks (e.g. commercials popping up automatically as the user goes to a place where s/he often visits). Hence, it was seen important to be able to set personal modes and filters (e.g. “shopping”, “meeting friends”) by which the amount and type of information would be determined. Still, several users pointed out that the user should have the possibility to get all the available data if wanted.

“I’m willing to receive advertisements while in shopping mode” – Ecology oriented focus group

In determining the relevance, also the recommendations from people like oneself were considered useful. The MMR service could be aware of what kind of users and events the user likes and in what kind of communities he/she belongs to, as well as what other like-minded unfamiliar people one considers interesting (e.g. celebrities, people with similar backgrounds as the user). The amount and the modality of presentation of the information provided by the service should follow the preferences of the user.

All in all, the concept of context awareness could be identified from several parts of the discussions. The service was seen to be aware of the user’s momentary needs, so that it could adapt and provide information depending on the user’s current context. However, also the problems related to this (e.g. the complexity of context as a holistic construct) were identified in the discussions. The context awareness could be personified as a personal agent, as one participant suggested, that predicts what the user wants and automatically offers interesting and relevant information. For example, another participant suggested that the service should recommend certain adverts for the user, which she/he might be interested in, based on his/her earlier behavior. The activities the user performs should then update and further specify the model of the user. Participants hoped that the service would learn the user’s types of interest and needs, e.g. based on location information.

Reliability and Sociality of MMR Information

There was rather much discussion about reliability and credibility of information in MMR services. Probably the main reason for this was the fact that the information content was seen possible to be created by anyone – as in Internet. The participants tended to trust the information provided by public and official institutions and well-known vendors more than that created by other users.

In general, it was considered useful that users would be able to create content and metadata (e.g. descriptions and tags) themselves. Other users’ comments especially on restaurants and other specific services were regarded as very useful and relevant information. Yet, users should be able to filter the other users’ comments, so that the user would not get comments from totally strange or dissimilar people. Moreover, all publicly available information was required to be up-to-date to be useful and trustworthy. Otherwise one would not base decisions (e.g. related to purchases) on such information.

“If a something has gone extremely well, of course you want to recommend it to others, and vice versa.” – Travelers’ FG

Active updating was seen to raise the trustworthiness of the information. In addition, comments and feedback created by other service users are interesting and reliable only if there are enough commentators. Also, the comment is more trustable the more the user knows about the commentator. People would trust the comments of MMR services as much they now trust comments on Internet forums, where mostly numeric measures (e.g. amount of positive comments) can be used to determine the trustworthiness of information if the other users are unknown. One single comment would not be necessarily trustable, but for example several negative comments would probably influence decision on buying.

Privacy Concerns: User Status and Location

Besides leaving comments, the participants were willing to share information about their current and future statuses. Users would describe what they are doing and what they are going to do and share the information with other users as well as with the service for it to be more aware of the user’s context. In the latter sense, the future status information was considered more useful. We interpret that the reason for such extrovert needs to be the current micro-blogging culture in Web2.0 services, such as in Facebook and Twitter.

Sharing people’s locations divided opinions, as expected. On one hand this kind of services would increase social awareness, which was seen as positive thing, but on the other hand the participants would want to retain their privacy. Sharing location data would be especially useful for finding and meeting friends without arranging separate meetings at particular place and time. Another positive viewpoint on sharing location information was feeling of safety of one’s family or other significant people (e.g. parents knowing their children’s locations). Being aware of one’s own location would increase feeling of safety especially in foreign locations.

“It would be nice to see that there are couple of friends near-by and a nice special offer on beer”.
– Wellness oriented focus group.

All in all, most participants were willing to share some limited amount of location information as long as the accuracy of the information is controlled by the user. Participants wanted to share their location only with selected people and for selected time duration. User should be able to choose who can see her/his movements and where she/he appears to be for other

users. It was also suggested that location information need not always be specific – in some cases only accuracy level of the city district is enough. Also, the location could be shared anonymously (i.e. not revealing the name of the friend who is nearby) or revealed not before the user is in the vicinity of the other user.

Interaction with MMR Services and Devices

Due to the complexity and variety of mobile contexts, interaction with the MMR devices is a challenging topic. When hands are occupied by another task, traditional mobile interacting with keyboard or touch screen becomes impossible. The user’s attention is largely reserved to coping in the physical environment: for example while waiting for public transportation, people tend to engage only in such multitasking that does not hinder them from noticing their transportation to come. Therefore, also the interaction technique and paradigm depend, amongst other things, on the user’s current cognitive and physical capacity.

Participants brought out that when actively searching for information in new contexts, they would prefer browsing the augmented information smoothly and continuously by using data glasses instead of hand-held devices. On the other hand, when not actively acquiring MMR information or interacting with the devices, the user requires cues of the information affordances in the environment. Participants thought that visual cues are often not enough, but also audible or haptic cues would be needed, especially with regard to the most important information. However, when actively using and interacting with the service, visual cues would be preferred over other modalities. The actual interaction and browsing of the information content were regarded so intensive and mentally loading that the visual modality would be preferred in the interaction after the user has been informed of the existing affordance.

What comes to viewing the augmented environment through a mobile device camera view, users said that it would serve this purpose rather well. In addition to the magic lense paradigm, the view could be utilized as a magnifying tool of 3D modeled information and as a window to alternative points of views (e.g. viewing the same environment as it was in antecedent eras).

Overall, pointing objects with a camera view to receive an overview of additional information was considered as a highly intuitive way to get a holistic picture of the environment and its affordances. However, continuous pointing to and interaction with a certain physical object must not require continuous pointing towards the object. Additionally, participants brought out that it must be challenging to determine which object the user wants to interact with in the camera view. Instead of constantly pointing to an object, the user could, for example, take a photo of the view and continue interacting with that. Furthermore, in tasks requiring high accuracy, pointing could be done with one’s fingers or a glove.

“One shouldn’t have to point with the device in his hand for example ten minutes continuously pressing some button”.

– Technology oriented focus group.

Participants suggested also other means of interaction, such as small movement of wrist, to be considered as ways of interaction as well. Also, certain contexts let people perform actions that are significant only at that specific situation. Hence, awareness of the context could be used in determining the user’s mode and which input techniques are enabled in each context. The most futuristic discussions envisioned the device to be able to read input information from the user’s eye movement, or finally read it directly from the mind.

One of the most salient requirements was found to be the need to tag and bookmark the browsed information. This would both increase relevance of the service and make it easier to find and utilize the information again later on. Furthermore, the augmented information does not necessarily need to be virtual information but augmentation could also be used in highlighting the most relevant and critical things in the physical environment (e.g. approaching familiar people or dangers in the environment). One participant mentioned also that the augmentation could be hiding unwanted information with augmented graphical effects. Hence, MMR services could help the user to both observe the most important parts and hide the irritating or unnecessary of both the digital and real-world environment.

One of the most useful potential features of MMR services was seen to be the ability to truly create joint services: one would be able to interact with the augmented information and use other information services via them. For example, the scenario example where the user directly selects and buys a movie ticket from an MMR advertisement was much appreciated.

Expectations Questionnaire Results

Here, we briefly present the results of the expectations questionnaires. Overall, the results based on focus group discussion supported and consolidated the results of the expectations questionnaire. The idea of mobile services with augmented reality or virtuality elements was regarded as intriguing, but at the same time the participants had doubts about the interaction with services, as well as privacy and information validity issues. MMR features were seen to bring added value, social aspects and liveliness to mobile services and increase one’s understanding of the surroundings. The questions and descriptive statistics are presented in *Table 3*.

Table 3. Expectation questionnaire concerning MMR services and devices (1: “I strongly disagree” – 5: “I strongly agree”), N = 23 (translated from Finnish).

I believe that mixing reality and virtual reality in mobile services...	Mean	St. dev.
... would bring added value to mobile services.	4.4	0.6
...would increase my knowledge concerning my surrounding environment.	4.0	0.6
...would ease controllability of mobile devices.	3.0	0.9
...would help me reach my goals via mobile services.	3.7	1.1
...would bring liveliness and continuously more content to mobile services.	4.2	0.7
...would stimulate me, i.e. would evoke memories and emotions, cheer me up and develop me.	3.7	0.9
... would demand much of resources concerning information processing, for instance attention and learning abilities.	3.8	1.0
... would increase attraction of mobile services..	4.0	0.7
... I could trust the validity of information I have received.	2.8	1.0
I believe that if services mentioned in this focus group interview would exist I would use them often.	3.5	1.0

DISCUSSION

Methodological Discussion

Focus groups offer a rich setting for innovation as the input and comments from other people can instantly create new thinking among the participants. However, because of its highly social nature a focus group session has some deficiencies when the aim is to elicit users' personal opinions and expectations. Considering issues of which the participants might not have a strong insight or opinion in advance, the discussion easily becomes directed towards a general consensus. Thus, individual opinions and needs that differ from the majority might not be disclosed at all. Regarding this study, it is challenging to conclude whether this affected our setup and results or not.

In addition, it might be challenging for the participants to communicate or even identify one's needs and expectations of such a futuristic topic. With such immature technology and little concrete solutions provided, it was challenging for the participants to picture the actual interaction and, for example, affordance perception issues that are often present when considering mixed reality environments. To minimize this effect of bewilderment from a futuristic technology we kept the contexts, tasks and users in the scenarios as present-day as possible, and only the technology and interaction with it were envisioned. Still, retrospectively thinking, we see that the sessions would have benefited from more concrete, visually described and detailed examples as stimulus materials.

Overall, with such an abstract and conversational research approach to a futuristic technology, the results are not strictly MMR specific – part of the results and ideas could be applied to almost any mobile and context-aware technologies. With such approach we could not investigate users' thoughts regarding, for example, detailed interaction with the future MMR services operational, for example, how other interaction modalities than the visual modality could be utilized. We see that the visually focused stimulus material and use cases affected the discussion so that no ideas were presented regarding, for example, auditory augmentation of the real world.

Because of the various backgrounds of the participants there was rather big diversity in participants' technical knowledge. Based on our experiences of various groups, we regard that with such a study setting it required the participants to be somewhat technologically oriented in order to understand the concept of mixed reality and to be able to innovate around it.

Design Implications

All in all, the focus group sessions provided us the "first contact" with users in regard to MMR service requirements. The understanding from this study serves as a basis in designing the features of early service concepts and prototypes, as well as the user experience provided by them. We received an extensive amount of expectations and design ideas, for example, with regard to relevance and personalization. The results related to information sharing and privacy complement our earlier studies (see e.g. [18, 21]). Most of the ideas were something to ease the everyday life, like location information and interpreting different languages in different surroundings. Next, based on the results we propose few implications for designing MMR services and discuss the validity and reliability of the results.

In regard to determining the relevance of information, the service could utilize a personal agent, predicting what the user wants and automatically offering interesting and relevant information. This agent could improve and update the model of user while

s/he is using the services. The user should be able to add some kinds of bookmarks and notes to the browsed MMR information. This would serve as a feedback tool for being able to further specify the relevance more specifically. Also, the recommendations from people like oneself were considered useful in determining the relevance. Therefore, the service should leverage other users' recommendations and other knowledge of the users' current social environment and the communities they belong to. Relevance also determines the amount of information provided in various phases of the use of the service. For example, in unfamiliar contexts the services should not provide too much information in advance so that one loses the experience of adventuring and exploring.

To support users' trust for the service and the reliability of its information content, the information should always be up-to-date, and one should see when and by whom the information has been modified. The reliability of the provided information can be an issue if the data is created by other users. For example in pharmacy users might want to get personal service from real personnel instead of trusting the augmented information.

The participants regarded augmentation of the real world not only as putting objects on top of the real world view but also as a tool to utilize the information in a more flexible way. The camera view of the device could also be used as a magnifying glass or seeing overviews of an area in the real world. Although the general opinion was that MMR services should not hide or distort the information from the real world, also contradictory use cases were discussed. The users might benefit from information diminishing instead of augmentation in situations where the real-world information is something that people might not want to see or hear (e.g. unethical issues, advertisements, content with adult only elements). One participants' solution for this was to hide information in the real world by augmenting blur effects on top of the real world view.

All in all, the presented user expectations are based on a limited number of use cases and types of interaction and thus, the results should not be generalized to represent entire user segments. Still, they provide important information for identifying the potential in the UX of MMR services and what are the most salient elements that affect how the UX builds up. The first expectations can provide a basis and starting point for successful development of MMR technologies with good user experience.

CONCLUSIONS AND FUTURE RESEARCH

As the main contribution of the paper we present new insight regarding users' expectations for mobile mixed reality services. We were able to elicit general level requirements and potential users' needs related to information content, type of interaction and general user experience issues in various MMR environments. We found out that the relevance and reliability of information content are central issues in determining how disruptive or fluent the interactions with MMR services are considered. Relevance affects the overall UX by influencing the sense of utility of the service, as well as how entertaining or stimulating they are for each user. It is based on several contextual aspects: the user's internal state, general-level user needs and the social environment the user acts in. The concept of personalization (e.g. user-set preferences) was also discussed largely as it was seen as a solution to determine what information might interest the user.

Reliability of information was a momentous aspect in both utility and leisure specific use cases. However, regarding

public information, the users tend to trust information created by trustworthy authorities, whereas regarding leisure and product information, the users mostly relied on comments and opinions of social network and other people similar as oneself. Furthermore, privacy issues, intuitive interaction, experiencing the real world itself, and the social nature of MMR services were also seen as important elements in MMR user experience.

Despite the challenging nature of studying expectations of non-existing services and applications, we regard this study to be an important move towards aspiring to understand the requirements for user experience when starting to develop MMR applications. The results serve as a basis for further exploring of the various factors that will affect the felt user experience of the technology prototypes and pilot services.

In future research, we will continue gathering requirements and expectations in real-context settings to gain more innovative and rich user data and use cases. The laboratory-like context in this study was not regarded to provide enough of stimulus for the users. Future research will also further specify issues such as personalization, filtering information, context awareness and proactive activity of the services. The study also pointed out a need for further research on map-based interaction and navigation in the user interface. Finally, the future target groups could be more limited to get more in-depth information regarding certain use cases or environments. We assume that the needs and expectations of users who do not represent “innovators” or “early adopters” would be beneficial to study, e.g. with more ethnographical research approaches.

ACKNOWLEDGMENTS

We thank all the participants of the study. The research was supported by the Finnish Funding Agency for Technology and Innovation (TEKES) under the project DIEM (grant 40223/07).

REFERENCES

- [1] Arhippainen, L. and Tähti, M. 2003. Empirical Evaluation of User Experience in Two Adaptive Mobile Application Prototypes. In: Proceedings of MUM 2003, Norrköping, Sweden. Pp. 27–34.
- [2] Azuma, R. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 355–385.
- [3] Bach, C. and Scapin, D. L. Obstacles and Perspectives for Evaluating Mixed Reality Systems Usability. *Acte du Workshop MIXER, IUI-CADUI*, 2004.
- [4] Battarbee, K. and Koskinen, I. 2005. Co-experience: user experience as interaction. *CoDesign*, 1(1), 5–18.
- [5] Brown, B., MacColl, I., Chalmers, M., Galani, A., Randell, C. and Steed, A. 2003. Lessons from The Lighthouse: Collaboration A Shared Mixed Reality System. *Letters CHI 2003*, 5(1), 577–584.
- [6] DIEM, 2009. <http://www.diem.fi>.
- [7] Flintham, M. et al. 2003. Where On-Line Meets On-The-Streets: Experiences with Mobile Mixed Reality Games. *Letters CHI 2003*, 5(1), 569–576.
- [8] Gleue, T. and Dähne, P. 2001. Design and Implementation of a Mobile Device for outdoor Augmented Reality in the ARCHEOGUIDE Project. In: Proc. of the 2001 Conference on Virtual Reality, Archeology, and Cultural Heritage, Glyfada, Greece, Nov. 28–30, 2001. ACM. Pp. 161–168.
- [9] Goss J. D. and Leinbach, T. R. 1996. Focus groups as alternative research practice. *Area*, 28(2), 115–230.
- [10] Hassenzahl, M. and Tractinsky, N. 2006. User Experience – A Research Agenda. *Behaviour and Information Technology*, 25(2), 91–97.
- [11] Hassenzahl, M. 2004. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4), 319–349.
- [12] Heikkinen, J., Olsson, T. and Väänänen-Vainio-Mattila, K. Expectations for User Experience in Haptic Communication with Mobile Devices. In: Proc. of MobileHCI 2009, Bonn, Germany, Sept. 15–18, 2009. ACM Press.
- [13] Herbst, I., Braun, A.-K., McCall, R. and Broll, W. TimeWarp: Interactive Time Travel with a Mobile Mixed Reality Game. In: Proc. of MobileHCI 2008, Amsterdam, The Netherlands, Sep. 2–5, 2008. ACM Press. Pp. 235–244.
- [14] ISO 13407:1999. 1999. Human-centered design processes for interactive systems. International standard of International Standardization Organization, 1999.
- [15] Kankainen, A. 2003. UCPCD: User-Centered Product Concept Design. In: Proceedings of the 2003 Conference on Designing for User Experiences, ACM Press. Pp. 1–13.
- [16] Kitzinger, J. 1994. The methodology of focus groups: the importance of interaction between research participants. *Sociology of Health*, 16(1), 103–210.
- [17] Kitzinger, J. 1995. Introducing focus group. *British Medical Journal*, 311, 299–302.
- [18] Lehikoinen, J., Olsson, T. and Toivola, H. Privacy Regulation in Online Social Interaction. In: Proceedings of IADIS 2008, Amsterdam, The Netherlands, 25–27 July 2008.
- [19] Ludford, P., Frankowski, D., Reily, K., Wilms, K. and Terveen, L. 2006. Because I Carry My Cell Phone Anyway: Effective Everyday Task Management. In: Proceedings of CHI 2006, Montreal, Canada, April 22–27, 2006. ACM Press. Pp. 889–898.
- [20] Milgram, P. and Kishino, F. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems*, Vol. E77-D, No. 12, 1994. Pp. 1321–1329
- [21] Olsson, T., Soronen, H. and Väänänen-Vainio-Mattila, K. User Needs and Design Guidelines for Mobile Services for Sharing Digital Life Memories. In: Proceedings of MobileHCI 2008, Amsterdam, The Netherlands, Sep. 2–5, 2008. ACM Press. Pp. 273–282.
- [22] Powell, R. A. and Single, H. M. 1996. Focus groups. *International Journal of Quality in Health care*, 8(5), 499–504.
- [23] Rourke, L., Anderson, T., Garrison, D. R. and Archer W. 2000. Methodological Issues in the Content Analysis of Computer Conference Transcripts. *International Journal of Artificial Intelligence in Education*, 12(2000), 8–22.

Session 6: Cognitive Analyses of Control Activity

The Orchestra: A Conceptual Model for Function Allocation and Scenario-based Engineering in Multi-agent Safety-critical Systems

Guy A. Boy

Florida Institute for Human and Machine Cognition
40 South Alcaniz Street
Pensacola, Florida 32502, U.S.A.
gboy@ihmc.us

Florida Institute Technology
150 West University Boulevard
Melbourne, Florida 32901-6975, U.S.A.
gboy@fit.edu

ABSTRACT

Function allocation in safety-critical systems is not a new research topic. However, there is a need for unifying what is separately done in engineering, human factors and organization science. In aviation for example, functions have been allocated on the flight deck and on the ground control stations, but very little has been done in between flight decks and ground stations. As we go forward, a multitude of machine agents grows and intertwined information flows become more important to understand information management by both individuals and organizations. This paper proposes a new approach to function allocation among human and machine agents based on the orchestra metaphor, where agents are considered as musicians. This kind of model requires a common frame of reference (the music theory), contracts (scores) must be appropriately initially and formally coordinated (the role of the composer), real-time coordination (the role of the conductor), and specific abilities to perform according to contracts (role and proficiency of the musicians). Contracts are seen as scenarios or storyboards, with an additional responsibility dimension. More generally, authority has become the central driving force of the model, where authority encapsulates both control and accountability. The understanding, which we commonly have of an orchestra, is extended by including authority trading that is based on negotiation among agents, as musicians. Examples are taken from the air traffic management domain currently under investigation worldwide. The Orchestra model is compared to previous related theoretical models of socio-technical systems.

Keywords

function allocation, scenario-based engineering, human and machine agents, coordination, safety-critical systems, orchestra model

ACM Classification Keywords

H.1.2 [User/Machine Systems] Human Factors – *function allocation, human-centered design*; J.4. [Social and Behavioral Sciences] Sociology – *organization sciences*; J.7. [Computer in Other Systems] Command and Control, Process Control.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

INTRODUCTION

Technology is now almost always equipped with layers of software that enable machines to interact like humans do, at least in very limited contexts. We commonly talk about human and machine agents (HMA). Human-computer interaction was traditionally thought as a person facing a computer as a one-to-one relation. Today, there are HMA societies in the sense of Minsky (1985), i.e., an agent being a society of agents. Human modeling, often commonly thought as information processing (Newell and Simon, 1972; Wickens, 1992), progressively migrated towards multi-agent organization modeling (Hutchins, 1995; Boy, 1998). The cognitive function concept emerged as a useful representation to support such modeling of both individuals (i.e., an individual as a organized structure of cognitive functions) and organizations (i.e., a set of cognitive functions distributed among a set of agents). The multi-agent approach is a fundamental approach to modeling contemporary human-machine systems. This paper proposes a new multi-function conceptual model that encapsulates basic organizational and cognitive processes that support the identification of emergent cognitive functions in HMA societies. It is based on previous work on cognitive function analysis (Boy, 1998) and function allocation work (Grote et al., 2000). This model was motivated by the introduction of new concepts of operations in air traffic management (ATM), such as task delegation from ground controllers to flight crews. This kind of transfer of authority will inevitably induce the emergence of new cognitive functions among the various ATM agents whether they are humans or automation. More specifically, the current hierarchical model of air traffic control (ATC), where authority is centralized on the ground, is evolving toward a distributed model of authorities that need to be coordinated among agents that are also evolving. This paper proposes the Orchestra model that suits well this kind of evolution where agents require a common frame of reference (a music theory analog), contracts (such as scores), and coordination (i.e., the role of the conductor). In addition, dynamic negotiation needs to be taken into account. Consequently, the paper proposes a methodology for the rationalization of cognitive functions during the life cycle of a multi-agent safety critical system.

FUNCTION ALLOCATION

Paul M. Fitts edited a famous report on human engineering for an effective air-navigation and traffic-control system in 1951, where he and his colleagues drafted possible roles of the human operator in future air-traffic control and navigation systems. They developed principles and criteria to design and assess the division of responsibility between human operators

and machines, as well as among human operators themselves. They anticipated issues in decision-making, the nature of information, the form that information may take (i.e., encoding), the rate of flow of information, its storage, perturbation, redundancy, and related research problems. They mostly focused on visual and voice communication problems. Among other things, this report provided what is now known as the Fitts's list of where humans appear to surpass machines and conversely. This preliminary work led to several lists of strengths and weaknesses of human operators and automated machines (Chapanis, 1965; Swain and Guttman, 1980; Sheridan, 1987). They were called MABA MABA, i.e., "Men Are Better At – Machines Are Better At". This was an easy but very limited way to provide guidelines for automation design (Parasuraman, Sheridan and Wickens, 2000). Later on, Hollnagel and Woods (2005) based their approach on the fact that joint cognitive systems (humans and machines) are dynamic and therefore complex, and need to cope with this kind of complexity at both individual and organizational levels. This approach is descriptive and requires operational developments.

Function allocation cannot be only addressed from a static point of view; it can also be highly dynamic. It can be dynamic because underlying processes are dynamic; it would be better to talk about real-time function adaptation, even if this is often referred to as dynamic function allocation (Corso and Maloney, 1996; Hildebrandt and Harrison, 2003). It can also be dynamic because cognition is distributed (Hutchins, 1995; Wright, Fields and Harrison, 2000).

It is interesting to notice that the next generation of ATM systems will have to be designed taking into account principles and criteria for both static and dynamic function allocation. What drastically changes today is the magnitude of the air capacity, i.e., the number of aircraft is tremendously more important than in 1951. Consequently, the conceptual model shifts from a single agent approach to a multi-agent approach. It is no longer possible to analyze each agent in the system individually because the interrelations are far more important than before. Technology is information intensive and organizational setups need to be revisited. Furthermore, agents are no longer only human operators, but also automation in the form of various kinds of software agents dedicated to specific tasks. For that matter, function allocation cannot be thought as an a priori process, but as an evolutionary process. The separability of human-automation sub-systems has become a real issue. The overall ATM system is becoming like a multi-agent biological entity where complexity is as much in the links between agents as in agents themselves. This is why function allocation among a set of interconnected agents is a difficult problem.

In complex dynamic human-machine systems, it is crucial to know which agent does what and when. In addition, each agent should have the capacity to execute the task he/she/it has to perform, i.e., an appropriate cognitive function should be allocated to this agent. Each function allocation has a cost that should be carefully understood and eventually measured. Finally, each function allocation induces a level of confidence and trust in the agent (Campbell et al., 1997). When the agent is a human, this is characterized in terms of level of training and experience. When the agent is an automated system, trust can be characterized by several metrics such as reliability, flexibility and cognitive stability, i.e., the ability to recover from human errors or system failure (Boy, 2007).

The cognitive function paradigm was chosen to support the multi-agent approach of the next generation of ATM systems in the French national PAUSA project (Boy et al., 2008). A cognitive function is defined by three kinds of attributes that

are its role in the organization where it is involved, its context of use, and its resources that are required to implement it. A cognitive function could be defined recursively by a network of cognitive functions that may be distributed among various agents in the organization, and across various meaningful contexts of use, whether nominal or off-nominal. Anytime a new cognitive function is defined or moved from an agent to another, it is crucial to look for new cognitive functions that emerge from the various interactions in the related agent network. Technology-centered automation is often defined as the transfer of a cognitive function from a human agent to a machine agent. When this process is extended to the search for emergent cognitive functions, it can be called human-centered automation (HCA). For that matter, HCA is an incremental process where design, test, practice and discovery of the emergent cognitive functions are intertwined. This approach to automation is strongly based on operational expertise, development of scenarios, human-in-the-loop simulations (HITLS) and formative evaluations.

SCENARIO-BASED ENGINEERING

Scenario-based design (SBD) is not new. SBD changes the focus of design work from defining system operations, i.e., functional specifications, to describing how people will use a system to accomplish work tasks and other activities (Carroll, 1995, 2009). SBD elaborates a traditional principle in human factors and ergonomics, i.e., human attributes and requirements should be taken into account in design and development. SBD consists in describing usage situations as design objects. It starts with the involvement of appropriate domain experts. In our case, pilots and ATC personnel are the domain experts. During the early phases of design, envisioned scenarios are developed from expert knowledge and knowhow. Such scenarios are usually built as extensions of current observed scenarios in the real world. They may be designed as analogs of similar configurations and chronologies observed in other domains. Scenarios are constantly readapted to support human-centered design appropriately.

In the context of function allocation for multi-agent safety-critical systems, such as the construction of commercial airplanes, the need for scenarios takes another dimension. Even if it is clear that scenarios are needed from the early beginning of the design process, they are also needed along the whole life cycle of the product. They are not only storyboard guides in human-centered design and development; they also provide an excellent framework for the rationalization of evaluation-test results, which in turn are used for re-engineering the product and improve its safety, usefulness and usability. For that matter, we move from SBD to scenario-based engineering as a support to function allocation during the whole life cycle of a socio-technical system.

A distinction is made between declarative and procedural scenarios (Boy et al., 2008; Straussberger et al., 2008). First, we deliberately choose the multi-agent model to support the development of scenarios. Scenarios are thought of in the same way as movie scenarios. Declarative scenarios describe the necessary objects and agents involved in the final product. These objects and agents are presented in the form of structure and function. Such descriptions necessarily lead to the way objects and agents interact among each other, and consequently to application use cases. Procedural scenarios describe chronologies of events and interactions among objects and agents. Such descriptions are stories and episodes that lead to appropriate definitions of such objects and agents. Declarative and procedural scenarios may be initially developed by different groups of people in isolation. These two types of

scenarios are developed concurrently to improve completeness of both objects/agents and their possible interactions. They are incrementally merged into synthesized generic scenarios.

Technical systems should never be looked at in isolation, but always as part of a bigger socio-technical system, which includes humans operating the system as well as formal and informal structures and processes within which they work. This is why scenarios are so important because they support the rationalization of the meaningful interactions in the socio-technical system. There are situations that are very difficult and even impossible to predict before they actually happen. These situations are usually called surprises. For example, the 2002 mid-air collision accident at Überlingen, Switzerland, which has been extensively analyzed (Weyer, 2006), has shown the effect of the introduction of the Traffic alert and Collision Avoidance Systems (TCAS) as a deconstruction of order or even a regime change, which may be a gradual shift from central control to decentralized self-organization. Some accidents such as the Überlingen one highlight such evolution, and sometimes revolution, in the overall socio-technical system where coordination has become one of the major issues. This is why we deliberately choose a multi-agent approach, instead of a single-agent approach (e.g., the pilot facing a screen), to express function allocation. To do this, we need to develop a common frame of reference, task delegation, and information flows among agents. The evolution of ATM is seen as a shift from army to orchestra from recent experience-based investigations (Boy and Grote, 2009; Boy, 2009). This metaphor and its possible extensions support very well the ongoing evolution of the airspace multi-agent system. In this way, we expect that we will be moving from the design-and-surprise approach to a principled approach of design based on scenario-based engineering toward the development of more robust and resilient socio-technical systems. We argue that such a principled approach to design could have avoided the Überlingen accident, by recognizing that a connection between TCAS and ground-based STCA (Short-Term Conflict Alert) systems could have facilitated the controller's situation awareness.

During the last decades, human factors researchers tended to blame the fact that engineering waited surprises to correct ergonomics of products, i.e., structures and functions of products. There will always be surprises unfortunately. The main issue is to try to anticipate them as much as possible. It is difficult to imagine other ways than constantly developing deeper knowledge and knowhow from positive and negative experience using the product. Our technological society is developing very fast, tremendously faster than before. We do not take enough time to analyze our mistakes and generate syntheses, in the form of "golden rules" for example. Scenarios are good tools to pose questions such as Who, What, When, Where, Why, How and How much (5W2H): Who and What are the agents and objects and relationships among them along relevant dimensions such as time (chronology), functional and structural dependencies, topological organizations and so on; Why do they exist in terms of role, context and resources (i.e., cognitive functions); When and Where they are useful, active or potentially related to one another; How do they work or how can people work with them or use them; How much load is involved in user's activity, in terms of workload, appropriate cost of any kind and so on. This is why maturity has become a field of research that is far from being mastered and "mature".

Scenario-based engineering should then look for maturity. The issue of maturity has been analyzed before (Boy, 2005). We know that we must focus on product maturity and practice maturity, i.e., what the product is for and how it is really used.

Product maturity is strongly based on the quality of high-level requirements and on their constant adjustments to the real world during the whole life cycle of the product. Of course, if high-level requirements are not right or strong enough in the first place, chances are that designers and engineers will have to re-engineer the product many times in the future, and sometimes get rid of the product unfortunately. This is why starting right is the best advice that we could give to a design team. But what does it mean to start right? It means starting with the appropriate scenarios. In addition, it means that the product should be thought as a global entity and not a juxtaposition of pieces that will eventually be assembled later. This is another reason why human-in-the-loop simulations are crucial as early as possible during the design/development process. Finally, teamwork must be cohesive from the beginning of design to operations and obsolescence of the product. Following up on this analysis, there is a need for a conceptual model that could support function allocation, scenario-based engineering and maturity reaching; this model is presented in the next section.

THE ORCHESTRA MODEL

The Orchestra model requires the definition of the authority concept. It was designed over the years (Boy, 1991) and finally refined during a study carried out from 2006 to 2008 on authority sharing in the aeronautical system, the already mentioned PAUSA project. Authority is defined from two main perspectives, i.e., control in the engineering sense (i.e., who is in charge and competent for a given task and function), and accountability in the legal sense (i.e., we are always accountable to someone else, and accountability includes responsibility).

Results were based on the definition of a scenario-based approach that supports the design of such HMA systems. We used the distinction between declarative and procedural scenarios. The main problem was to obtain meaningful and generic scenarios that would be the source of emergent cognitive functions during further HITLS, which range from very simple paper and pencil narrative simulations to the use of interconnected very sophisticated realistic simulators. Both involve domain experts.

In the beginning of a project of this kind, strong expertise and experience from operational practitioners is required to develop useful scenarios. In ATM, these practitioners are both pilots and ATC controllers (ATCOs), and also aerospace designers and certifiers. It is of course also important to motivate and carefully filter their inputs through creative designs in a reflexive way. However, testing will always remain the mandatory (long) step on which design and development processes will have to comply.

No simulation can be purposefully and efficiently carried out without a conceptual model. In this scenario-based engineering approach to function allocation, the Orchestra model is an alternative to the traditional army-type model that supports a hierarchical decomposition of functions. Four categories of entities must be defined.

- First, the music theory that supports the various information-flows and provides a common frame of reference for all agents in the environment.
- Second, the scores that agents are required to use in order to support their assigned functions during operations. Composers typically develop scores and articulate them among each other. These composers still remain to be identified correctly in the ATM case.

- Third, conductors who provide the operational timing patterns, and consequently will be responsible for the effective information flows, i.e., the overall symphony performance to take the orchestra metaphor.
- Fourth, musicians themselves who are required not only to perform what their scores say, but also to articulate their own plays with the others’.

In an HMA organization such as an orchestra, agents are interrelated with respect to three kinds of interaction models (Boy, 2002). These models are distinguished with respect to the level of knowledge each agent has of the others in the organization.

- (1) When agents do not know each other, the best way to interact safely, efficiently and comfortably is to be supervised. *Supervision* is the first interaction model. None of the supervised agents has the authority to decide what to do; a supervisor does it for them.
- (2) *Mediation* is the second interaction model. Agents have a common frame of reference (CFR) through which they are able to interact. They still do not know each other deeply, but they know that they can interact between each other through the CFR. In addition to the CFR, there are mediating agents who facilitate interactions. In WYSIWYG user interfaces in addition to desktop metaphors, there are mouse-sensitive help lines that pops-up on demand for example. In this model, the authority is distributed among the agents.
- (3) The third interaction model is *cooperation by mutual understanding*. This is what people usually do when they interact with each other. This model assumes that agents are able to construct a mental model of the others in order to perform better in future interactions. Very simple instances of such a model have been developed and used so far on computers. For example, some pieces of software are able to learn user’s habits and are able to incrementally provide smart options or suggestions. This is the case of Microsoft Word that is able to learn user’s specific lexicon from frequent uses of words. Web browsers remember frequent URL, etc. In this model, authority is traded between the agents. In human-human interaction via machine agents, related technology should provide appropriate situation awareness means to enable sustainable and symbiotic communication.

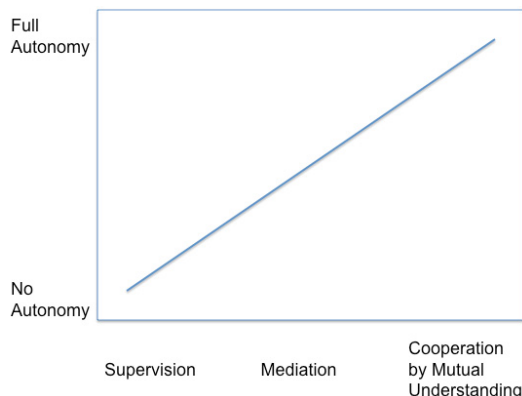


Figure 1. Interaction models from no-autonomy to full-autonomy of agents.

To summarize, there is a continuum from the supervision model of interaction where authority follows a top-down army-type model, to the mediation model of interaction where

authority follows a transversal orchestra-type model, to the cooperation by mutual understanding model of interaction where authority follows a more-chaotic trade model (Figure 1). These interaction models are very useful to support the way cognitive functions are implemented in complex software not only from a human-computer interaction point of view, but also from an internal subsystem-to-subsystem point of view. In particular, they also provide an articulated way to validate large object-oriented software.

AN AERONAUTICAL APPLICATION

Authority sharing is one of the major themes of the next generation of air traffic management (ATM) system, flight deck automation in particular. The fact that we will have more aircraft in the sky (i.e., air traffic capacity increase), and we want to enhance safety, requires deepest research on the way various functions are being reallocated among the various agents. We need to better identify pilots’ information requirements and communication needs to perform tasks currently managed by air traffic control (ATC), which will greatly increase the needs for pilot’s awareness of the surrounding airspace, (human and system) failure identification and recovery, and unexpected-event handling in this dynamic and complex multi-agent infrastructure.

Therefore, we need to co-design and co-adapt both technological and organizational support. Avionics software is now highly sophisticated, enabling many machines to be considered as agents, i.e., having cognitive functions as humans have. Human and machine agents are more interconnected in the air space than before, and their inter-relationships are often crucial to understand, master and support. This evolving ATM multi-agent world is highly situated and context identification is a primary concern. In particular, flight deck automation will have to be designed taking into account that pilots will gain autonomy thus changing the coordination requirements.

Consequently, function allocation needs to be addressed during the whole life cycle of all ATM systems. Cognitive function analysis is typically used to support the analysis, design and evaluation of such function allocation. More specifically, cognitive processes, such as authority sharing, distribution, delegation and trading, must be addressed. While there are human cognitive functions that can be predicted during design, there are some that will only emerge from use. This is why scenarios should be extensively developed and HITLS carried out.

We are currently working on the difficult problem of spacing and merging (S&M) in dense traffic to improve the sequencing of arrival flows through a new allocation of spacing tasks between air and ground. Today, ATCOs solely manage aircraft S&M in busy airspaces. They control both the sequencing decisions and manage the merging routes, airspeeds and altitudes, guiding each aircraft. Controllers are aided by today’s tools, which range from simple Letters of Agreement (LOA) and standard navigation aids, to more advanced systems like today’s GPS approaches and integrated Flight Management Systems (FMS). The new Required Navigation Performance (RNP) procedures are the latest improvement down the traditional path of providing the pilot with standard procedures and a more accurate way to follow them. While this approach is an important one, it alone will not solve the future problems of airspace congestion because it addresses only execution and does not address the major issue, which is coordination. Today, ATC is a centralized army-type decision point, i.e., all decisions must pass through this point and be distributed in a serial manner to all pilots within the managed airspace. This is a clear

bottleneck that is highly dependent on the skill of the controller to analyze the situation, make decisions, and then communicate the required information to each aircraft as necessary.

Pilots under instrument meteorological conditions (IMC) have traditionally been “flying blind” with respect to other aircraft around them. Good pilots will build a mental map by listening to the radios (party-line) and piecing the scene together. Recently, TCAS and Automatic Dependent Surveillance-Broadcast (ADS-B) have started providing pilots with a little more awareness of their immediate environment. These technologies provide the pilot with a “second set of eyes” besides the controllers. This information allows pilots to make decisions of their own, but unfortunately it is not coordinated with ATC, which has resulted in unfortunate accidents, again highlighting the importance of coordination.

Future ATM systems will enable pilots to be more autonomous and consequently will require more coordination among agents. They will have contracts like musicians have scores. Consequently, these contracts will have to be coordinated by some kinds of planners, like the composers do. From this point of view, the main difference between ATM and a symphony is that contracts may change during performance, like the play of a Jazz orchestra. Authority trading will be a major issue. Situation awareness of each agent remains a central emergent cognitive function to investigate and identify during design and development. In fact, agent’s authority and situation awareness are intimately coupled, and their identification determines the type of interaction model the agent will have with the other agents that are relevant in the operational context. Sometimes, supervision is the only interaction model that is possible, and agents will need to refer to a conductor. In other situations, they will be able to interact via contracts (scores) and trust this mediating means. Finally, it will happen that they will perfectly understand what the others are doing, and therefore will communicate directly.

Off-nominal situations are infrequent, but have a tremendous impact when they do occur. They typically induce dynamic function allocation, i.e., appropriate agents will have to be aware of the situation change (resulting in a different common frame of reference), contracts will have to be redefined and coordinated (composer role), and consequent operations will have to be coordinated (conductor role). For example, it may happen that an aircrew would not be able to make it off the runway at the high-speed exit and take a full-length landing. In a congested terminal area, the following aircraft will have to perform a go-around maneuver. First, the aircrew must realize they are not going to make the exit (situation awareness cognitive function), they must manage the landing (safety-assurance and action-taking cognitive functions), and find time to let the controller know (coordination cognitive function). Consequently, the ATCO must inform the trailing aircraft and potentially all other aircraft sequenced on the approach (coordination cognitive function). All these cognitive functions must be implemented at the right time, which might not be the case taking the extra workload during this kind of operations. Information flows are highly dynamic and can only be managed by well aware and knowledgeable agents, possibly new technology. For example, the ATCO re-sequencing traffic may also find out that there is an aircraft that is low on fuel and requires an emergency landing. Creative decision-making is consequently the appropriate cognitive function that is at stake for the ATCO. On this very simple example, we see that authority must be timely shared among appropriate agents.

One way of managing this coordination problem is to develop appropriate automation. Automation can be used to detect

when an aircraft will not make the exit and automatically signal the controller, elevating this burden from the pilot who is likely under high workload already. That same signal could automatically be sent to all the trailing aircraft. This kind of additional agent is expected to create more situation awareness among involved agents and therefore increase their common understanding of the situation (thus promoting the third interaction model). In addition, the ATCO, as a conductor, could make a single call confirming the situation and requesting reduced speeds. Each aircraft could acknowledge through their flight displays instead of using radio communications and ATCOs would see each response on their own screens. If this kind of solution seems to simplify the job of the various agents, it is mandatory to make sure that they are properly trained or fine-tuned, and use the right cognitive functions.

In these examples, we can see that cognitive function analysis using the Orchestra model enables the investigation of the various relationships among agents and the emergence of new cognitive functions, such as appropriate automation. Of course, any solution needs to be tested and further validated in HITLS or in the real world.

DISCUSSION

Systems such as air-air surveillance capabilities (ADS-B) and cockpit automation (ASAS) are being designed to enhance authority sharing between the flight deck and the ground. The evolution between what is currently done and the next generation of air-ground environments requires carefully studying function allocation and keeping automation as simple as possible, in terms of flexibility for the actors. Aircraft S&M technology remains immature and requires further investigation and development. In terminal areas, S&M currently relies on air traffic controllers’ skills and experience and is affected by weather conditions, rates of runway use, ground congestion and other factors. In the perspective of authority delegation to the flight deck, new approaches to S&M need to be invented, especially in high-density traffic situations. They will rely on new kinds of automated technology and procedures. Obviously, whenever S&M can be anticipated en route, it would be a great gain of time and workload in terminal areas. It is now important to identify required functional evolutions and cognitive functions that emerge from this evolution, taking into account a representative environment with very high traffic. Referring to the Orchestra model, new approach procedures and terminal area patterns are part of the common frame of reference, i.e., a music theory analog. Generic contracts, as scores, needs to be defined according to cognitive functions that will emerge from both new automation and organizational rules, mainly coordination rules. Contract coordination should be both anticipated (composer role) and managed (conductor role). Finally, function allocation should be thought in terms of authority sharing in the sense that several agents share responsibility and control in context. It could be a priori defined, i.e., each function represented by a contract is allocated to an appropriate agent. It should also be dynamically defined, i.e., cognitive function may be allocated with respect to the ongoing situation. As already seen, dynamic function allocation requires appropriate situation awareness, i.e., there is a constant need to look for potential hazards and understand the perception and cognitive limits of the various agents in order to compensate with additional cognitive functions and maintain an appropriate cognitive stability. Such cognitive functions could be additional resources in the form of supervisors, mediators or automated links that provide a better common understanding. Of course, their implementation and operational costs should be evaluated with respect to relevant human and

technological factors. The choice of their effective implementation in the real world depends on these evaluations.

Other approaches, such as cognitive systems engineering/joint cognitive systems (Hollnagel and Woods, 2005), consider the growing complexity of socio-technical systems, problems and failures of clumsy technology, and the limitations of linear models and the information-processing paradigm. They also recognize the need for cognitive function (Boy, 1998) “in the mind”, i.e., processes that mediate responses to events. In fact, this anthropological approach of cognition was already started with the identification of situated actions (Suchman, 1987) and distributed cognition (Hutchins, 1995). All these contributions emphasize context as the main research issue. In fact, people are both goal-driven and event-driven; they are opportunistic according to context. This is why context is so important to identify and take into account. “Situated activity is not a kind of action, but the nature of animal interaction at all times, in contrast with most machines we know. This is not merely a claim that context is important, but what constitutes the context, how you categorize the world, *arises together* with processes that are coordinating physical activity. To be perceiving the world is to be acting in it—not in a linear input-output relation (act – observe – change) – but dialectically, so that what I am perceiving and how I am moving co-determine each other” (Clancey, 1993).

Context is an extremely difficult concept to grasp and identify since it is directly associated to the persistence of situations and events (Boy, 1998); some are long enough to be captured, and some others are too short to even be perceived. This is why a scenario-based approach carried out by domain-expert professionals is necessary. The Orchestra model is a metaphoric framework that enables handling context in a functional and structured way, since the cognitive function representation includes the context attribute by construction. The identification and categorization of the possible connections and interactions among agents through their cognitive functions enables to better understand various relevant issues of situation awareness. In fact the way we identify and categorize the world is crucial in the perception of context when acting. It is clear that all metaphors are very limited, and the Orchestra metaphor has limitations when we use it to describe socio-technical systems. However, it incrementally emerged as an acceptable model of the evolution of our software-immersive environment, and the ATM environment in particular.

As already described in a previous paper (Boy, 2002), the cognitive function analysis has many similarities with the activity theory, the Russian approach to cognition, which considers that people learn from their environment, and human activity is mediated by surrounding artifacts. The concept of cognitive function is very similar to Leont’ev’s functional organs (Leont’ev, 1981). “Functional organs are functionally integrated, goal-oriented configurations of internal and external resources. External tools support and complement natural human abilities in building up a more efficient system that can lead to higher accomplishments. For example, scissors elevate the human hand to an effective cutting organ, eyeglasses improve human vision, and notebooks enhance memory. The external tools integrated into functional organs are experienced as a property of the individual, while the same things not integrated into the structure of a functional organ (for example, during the early phases of learning how to use the tool) are conceived of as belonging to the outer world.” (Kaptelinin, 1995).

Another dimension that is not extensively presented in this paper is time. Time is very important in music. The Orchestra model is a very insightful metaphor for time-wise investigation.

We have already described this dimension in another paper by describing the time sequences developed by the various cognitive functions involved in the Überlingen accident (Boy and Grote, 2009). The specificity of the Orchestra model is to encapsulate both design and performance times, i.e., the time of the composer and the time of the conductor and musicians. Information flows are important to capture in the form of useful and usable contracts (scores) designed and developed by composers at design time, and in the form of coordination patterns emerging from performance and handled by conductors at operations time.

CONCLUSION

Since context is a major concern in the design of appropriate safety-critical systems, scenarios are very good tools to support the elicitation of emergent cognitive functions. Scenario-based engineering requires to be supported by a strong conceptual model. The Orchestra model was found a good conceptual tool to categorize cognitive functions in air traffic management problems, their allocation among human and machine agents, as well as the various relevant relationships between them.

This paper presents a conceptual model for function allocation and scenario-based engineering in multi-agent safety-critical systems. This model takes into account the fact that allocation can be done a priori, but is also dynamic by nature. Indeed, relationships between agents are supported by contracts that are very similar to scores in music. In addition, when there are many agents to coordinate, these contracts (scores) need to be coordinated also; this is typically the role of a composer in music. Despite the initial planning, i.e., the coordination of contracts, there are always events that are not anticipated either because they are intentions from some of the agents that differ from the original plans, or unexpected external events. These events require dynamic re-allocation of functions, and therefore modification of initial contracts. This is typically the role of a conductor. Agents, as musicians, need not only to be competent to perform their functions; they also need to understand what the other agents are doing. This is why we need interaction models also. In the best case, they communicate between each other by common understanding, but they may require being supervised or mediated when they do not have acceptable situation awareness.

As many other contributors suggested, new technologies and automation do not have only quantitative effects, but have also qualitative shifts (Dekker and Woods, 2002), induce the emergence of new practices (Flores et al., 1988), and even may alter the tasks for which they were designed (Carroll and Campbell, 1988). The Orchestra model provides a conceptual framework that supports the elicitation of this kinds of emergences.

ACKNOWLEDGMENTS

Portions of this work were performed when the author was coordinating the PAUSA project. Many thanks to Matthew Johnson who helped on the articulation of the ATM examples.

REFERENCES

- [1] Boy, G. A. 1991. Advanced interaction media as a component of everyday life for the coming generation. Invited Speech at the JMA Marketing World Congress, Tokyo, Japan.
- [2] Boy, G. A. 1998. Cognitive Function Analysis. Ablex-Greenwood, Westport, CT.

- [3] Boy, G. A. 2002. Theories of Human Cognition: To Better Understand the Co-Adaptation of People and Technology, in *Knowledge Management, Organizational Intelligence and Learning, and Complexity*, edited by L. Douglas Kiel, in *Encyclopedia of Life Support Systems (EOLSS)*, Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK, [<http://www.eolss.net>].
- [4] Boy, G. A. 2005. Maturity, Automation and User Experience (Maturité, Automation et Experience des Utilisateurs). *Proceeding of the French Conference on Human-Computer Interaction*. ACM-Press, New York.
- [5] Boy, G. A. 2007. Perceived Complexity and Cognitive Stability in Human-Centered Design. *Proceedings of the HCI International 2007 Conference*, Beijing, China.
- [6] Boy, G. A., Salis, F., Figarol, S., Debernard, S., LeBlaye, P. and Straussberger, S. 2008. PAUSA: Final Technical Report. DGAC-DPAC, Paris, France.
- [7] Boy, G. A. and Grote, G. 2009. Authority in Increasingly Complex Human and Machine Collaborative Systems: Application to the Future Air Traffic Management Construction. In: the *Proceedings of the 2009 International Ergonomics Association World Congress*, Beijing, China.
- [8] Carroll, J. M. and Campbell, R. L. 1988. Artifacts as psychological theories: the case of human-computer interaction. IBM Research Report RC 13454, Watson research Center, Yorktown, Heights, NY.
- [9] Carroll, J. M. 1995. *Scenario-based design: Envisioning work and technology in system development*. New York: John Wiley & Sons.
- [10] Carroll, J. M. 2010 to appear. Scenario-based design. In: Boy, G. A. (Ed.). *Handbook of Human-Machine Interaction*, Ashgate, UK.
- [11] Campbell, G., Cannon-Bowers, J., Glenn, F., Zachary, W., Laughery, R. and Klein, G. 1997. Dynamic function allocation in the SC-21 Manning Initiative Program: Naval Air Warfare Center Training Systems Division, Orlando, SC-21/ONR S&T Manning Affordability Initiative.
- [12] Chapanis, A. 1965. On the allocation of functions between men and machines. *Occupational Psychology*, 39, 1–11.
- [13] Clancey, William J. 1993. Situated action: A neuro-psychological interpretation (Response to Vera and Simon). *Cognitive Science*, 17(1), 87–107.
- [14] Corso, G. M. and Moloney, M. M. 1996. Human performance, dynamic function allocation and transfer of training. In: Koubek, R. J. and Karwowski, J. (Eds.). *Manufacturing Agility and Hybrid Automation-I*. *Proceedings of the 5th International Conference on Human Aspects of Advanced Manufacturing*. Louisville, KY: IEA Press.
- [15] Dekker, S. W. A. and Woods, D. D. 2005. MABA-MABA or Abracadabra? Progress on Human-Automation Coordination. *Cognition, Technology & Work*, 4, 240–244.
- [16] Fitts, P. M. 1951. (Ed.). *Human engineering for an effective air navigation and traffic control system*. Washington, D.C., National Research Council.
- [17] Flores, F., Graves, M., Hartfield, B. and Winograd, T. 1988. Computer systems and the design of organizational interaction. *ACM Transaction on Office Information Systems*, 6, 153–172.
- [18] Grote G., Ryser C., Wäfler T., Windischer, A. and Weik, S. 2000. KOMPASS: a method for complementary function allocation in automated work systems. *International Journal of Human-Computer Studies*, 52, 267–287.
- [19] Hildebrandt, M. and Harrison, M. 2002. The temporal dimension of dynamic function allocation. *Proceedings of ECCE-11*. Pp. 283–292.
- [20] Hildebrandt, M. and Harrison, M. 2003. Putting time (back) into dynamic function allocation. *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society*. Denver, Colorado, October 13–17. Pp. 488–492.
- [21] Hollnagel, E. and Woods, D. D. 2005. *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: CRC Press / Taylor & Francis.
- [22] Hutchins, E. 1995. How a Cockpit Remembers its Speeds. *Cognitive Science*, 19, 265–288.
- [23] Kaber D. B. and M. R. Endsley 2003. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theor. Issues in Ergon. Sci.*, Taylor & Francis, 1–40.
- [24] Kaptelinin, V. 1995. Designing learning activity: a cultural-historical perspective in CSCL. *Proceedings of the Computer Supported Cooperative Learning (CSCL'95)*. Indiana University, Bloomington, IN.
- [25] Leont'ev 1981. *Problems of the development of the mind*. Moscow: Progress.
- [26] Minsky, M. 1985. *The Society of Mind*. Touchstone Simon and Schuster.
- [27] Newell, A. and Simon, H. A. 1972. *Human Problem Solving*. Prentice Hall, Englewood Cliffs, NJ.
- [28] Parasuraman, R., Sheridan T. and Wickens, C. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, 30, 286–297.
- [29] Sheridan, T. B. 1987. Supervisory control. In: Salvendy, G. (Ed.), *Handbook of human factors*, Wiley, New York, USA. Pp. 1243–1286.
- [30] Straussberger, S., Chamayou, C., Pellerin, P., Serres, A., Salis, F., Feuerberg, B., Lantes, J. Y., Guiost, B., Figarol, S., Reuzeau, F. and Boy, G. A. 2008. Scenarios in PAUSA. DGAC-DPAC EURISCO, Technical Report. April.
- [31] Suchman, L. A. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge: Cambridge Press.
- [32] Swain, A. D. and Guttman, H. E. 1980. *Handbook of human reliability analysis with emphasis on nuclear power plant applications (NUREG/CR-1278)* Washington, DC: U.S. Nuclear Regulatory Commission.
- [33] Wickens, C. D. 1992. *Engineering Psychology and Human Performance*, Harper Collins, New York.
- [34] Wright, P. C., Fields, R. E. and Harrison, M. D. 2000. Analyzing human-machine interaction as distributed cognition: The resources model. *Human-Computer Interaction*, 15(1), 1–4.

Cognitive Control in Car Driving: The Case of Brain-injured Drivers

Jean-Michel Hoc
 CNRS – IRCCyN B.P. 92101
 44321 Nantes Cedex 3, France
 Jean-Michel.Hoc@irccyn.ec-
 nantes.fr

Camilo Charron
 University of Rennes 2 – CRPCC 6,
 av. Gaston Berger 35043 Rennes
 Cedex, France
 Camilo.Charron@univ-rennes2.fr

Isabelle Milleville-Pennel
 CNRS – IRCCyN B.P. 92101
 44321 Nantes Cedex 3, France
 Isabelle.Milleville-
 Pennel@irccyn.ec-nantes.fr

ABSTRACT

Cognitive control is a key tool for adaptation in dynamic situations. This paper's main aim is to assess the relevance of a theoretical framework for cognitive control in dynamic situations, in order to understand brain-injured car drivers' cognitive impairment. The framework bears on a cognitive control multimodality based on the crossing of two orthogonal dimensions: *symbolic/ subsymbolic* and *anticipative/ reactive control*. Brain-injured (BI) car drivers' behaviour was compared to a control group (CTRL) during driving simulator scenarios. BI participants showed a more symbolic and a more reactive cognitive control than CTRL participants. Whereas CTRL participants succeeded in adapting to conflicting situations with a stable cognitive control distribution among the modalities (*cognitive compromise*), it was more costly for BI participants who had to change their cognitive compromise. In addition, BI participants were less able to process secondary driving subtasks. Some implications in terms of car-driving assistance are drawn.

Keywords

cognitive control, symbolic and subsymbolic processing, anticipative and reactive behaviour, brain-injured car-driver

INTRODUCTION

Brain injury often concerns young drivers and can result in an impediment to autonomy and mobility. After recovery, two questions are raised: how to evaluate driving ability and the possibility of rehabilitation. The present paper focuses on the first question. The population of BI drivers is heterogeneous. Frontal brain damage frequently occurs, bringing with it associated difficulties in terms of planning [1]. However, anatomical injuries can vary widely. Thus, the current research approach is functional. Some authors try to define driving simulator or real driving tests in order to identify the functional deficits related to and within the context of driving (e.g., [2]). A first experiment was undertaken that is in line with this approach and is particularly relevant for identifying cognitive control differences between brain-injured drivers (BI) and a control group (CTRL).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Cognitive control is a key tool for adaptation in dynamic situations. Hoc and Amalberti [3] have developed a model of cognitive control in dynamic situations that is based mainly on studies of industrial process control and transportation. We applied this model to the study of BI drivers who had recovered from their brain injury and had been driving again for several years. As a matter of fact, adaptation to unexpected situations is a key feature of driving skills. Although attention is frequently addressed within this context [4], it is not sufficient to account for adaptation mechanisms. We will relate our results to the attention framework, before discussing some issues in terms of car-driving assistance.

THEORETICAL FRAMEWORK

Hoc and Amalberti [3] defined cognitive control as a process of bringing into play the cognitive representations and operations required for adaptation, both in the correct order and with the appropriate intensity. In order to account for cognitive control dynamics, they considered two complementary aspects: the *cognitive compromise* and the notion of *satisficing performance* (borrowed from Simon [5]), which determines the former.

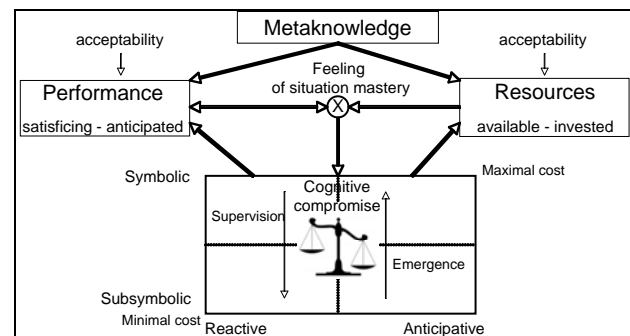


Figure 1. Cognitive control dynamics.

Cognitive control modalities are defined within a plane generated by two orthogonal dimensions (bottom of Figure 1). The first dimension contrasts symbolic control with subsymbolic control. Symbolic control bears on information interpretation, whereas subsymbolic control is directly based on the superficial (e.g., perceptive) features of information. The second dimension contrasts anticipative control with reactive control. The former bears more on internal information (mental models) and the latter more on external information (situational cues). However, within this context, control reactivity has nothing to do with processing speed. Rather, it concerns the need to rely on external information. Some parallelism is possible between several control modalities (e.g., symbolic control supervising subsymbolic control, or reactive control correcting anticipative control). The distribution of control within this plane constitutes the *cognitive compromise*.

Satisficing performance is the level of performance that is considered by the individual to be acceptable within a certain context (e.g., motivation, social acceptability, invested resources). Satisficing performance may have several criteria, with different levels of priority.

At any one time, an individual tries to reach an appropriate cognitive compromise. The distribution of cognitive control among its diverse modalities enables the individual to achieve a satisficing performance at an acceptable cost in terms of invested resources (e.g., symbolic control is more costly than subsymbolic control). Resources include energy as well as cognitive processes, ranging from high-level to routine skills, and attentional processes.

The dynamics of cognitive control for adaptation is determined by metaknowledge that enables the individual to evaluate anticipated performance and resources needed. If anticipated performance is lower than satisficing performance or/and anticipated resources need larger invested resources, there is a feeling that situation mastery is low. In this case, the individual can reduce the satisficing performance level or increase the invested resources. Then the cognitive compromise can be adjusted. For example, if the individual experiences a high rate of action slips (bad performance), there is a feeling of low situation mastery. The individual can then decide to invest more resources, exerting a more symbolic control of the activity.

One of the main methodological difficulties when attempting to identify the distribution of control among its diverse modalities is the use of non-invasive indicators. The importance of visual information in car driving allowed us to base our identification on some parameters of eye movements. However, other kinds of variables were recorded, such as speed, speed variability, and so on.

Symbolic control needs deeper processing, as it is based on interpretation. This is consistent with the literature [6], which considers that symbolic control can be identified by a longer mean fixation duration than subsymbolic control.

Anticipative control can be identified by fixation distance, as is shown by an abundance of data on visual activity during car driving. Healthy and experienced drivers have mostly far fixations. Any near fixations are of quite a short duration and occur in straight lines [7]. In bends, they spend a large amount of time fixating an area around the tangent point, inside the curve. This is a way of anticipating the road curvature and the appropriate steering wheel angle [8, 9]. However, other distant points on the road can play the role of the tangent point in anticipation [10]. Thus, the fixation distance and position can be an indicator of anticipation when driving.

In this experiment, we investigated cognitive control and adaptation differences between BI and healthy drivers. This was carried out from within a context of arbitration between conflicting performance criteria in the driving situation, namely speed and safety. A speed or safety instruction is supposed to directly act on the driver's definition of satisficing performance. Events occurring during the experimental scenarios are assumed to provoke conflict between performance criteria and possible changes in the cognitive compromise. For example, a fast-moving vehicle approaching to the rear may cause a driver to speed up, despite there being a safety instruction aimed to encourage the adoption of a low speed. In order to overcome such conflict, the driver has to invest more resources. The cognitive compromise may or may not be robust enough to allow the driver to do so. In the second case, the driver may simplify the definition of satisficing performance by relaxing some constraints or changing the cognitive compromise.

METHOD

Participants

The BI group comprised seven male drivers who had agreed to participate in the experiment. They had all recovered from a brain injury caused by a trauma and were aged between 35 and 50 years. They had had a score equal or below 8 on the Glasgow Coma Scale (GCS) and a coma of at least 48 hours' duration. On the whole, neuropsychological tests showed that they had difficulty in dividing their attention and that they had also experienced problems in planning and anticipation. Each of the BI participants had gained their driving licence at least two years before the brain injury occurred. All of them had recovered and were driving again. On average, they each had 11 years of driving experience.

The control group (CTRL) comprised six male participants with no impairment. They were recruited from within IRCCyN and were aged between 36 and 50 years. On average, they obtained a driving licence 23.6 years ago.

All the participants had normal or corrected-to-normal vision and driving experience of more than 30,000 km.

Apparatus

The driving simulator software, Sim2 (developed by the MSIS team at INRETS) was used, coupled with a FAROS fixed-base driving simulator. It was equipped with an automatic gearbox, a steering wheel fitted with force feedback, brake, accelerator and clutch pedals, and a speedometer. The visual scene was projected onto a screen (3.02 m in height x 2.28 m in width, which corresponds to a visual angle of 80° height and 66° width). A 3.5-km main road, forming a circuit, was simulated with traffic and with about ten bends of various directions.

An eye-tracker, IviewX (SMI), was used to investigate visual exploration. This eye-tracker consists of a hardly invasive, lightweight head-mounted camera that captures images of the subject's eye and field-of-view.

Procedure and Experimental Design

Following a familiarisation stage, participants had to complete six laps. Two of these were base laps. These comprised simple scenarios, one without a car in the lane occupied by the participant, and one where a slow car was present to act as an incentive for the driver to slow down (or to overtake). Then, each participant had to perform four experimental laps, generated by crossing two binary and independent variables. For each lap, pedestrians were present on the road verge. The independent variables are as follows.

- [1] *Type of Instruction (INSTR)*. Half of the experimental laps were performed with a *Safety Instruction (SAI)*: "imagine there is a child with you in the car and you must be very careful on the road"; the other half with a *Speed Instruction (SPI)*: "imagine you have a very important appointment, for example a job interview, and you are late".
- [2] *Type of Scenario (SCEN)*. With each type of instruction, there were two types of scenario. In one of them, only the slow-moving car interacted with participants, acting as an incentive to slow down, thus defining a *Safety Scenario (SAS)*. In the other scenario, when participants approached the slow-moving car, a fast-moving car caught them up. This car was visible in the mirror. It followed them for the remainder of the lap and was an incentive for them to speed up, thus defining a *Speed Scenario (SPS)*. Two

distinct but similar scenarios of each type (1, 2) were presented in order to avoid familiarisation.

The order of presentation of the four experimental conditions was balanced over participants. Before the first lap and after each lap, a questionnaire was submitted to participants in order to collect their performance and situation mastery assessments (see below).

Data Recording and Analysis

Within the restricted scope of this paper, only the main results will be presented. Thus, only some of the recorded variables will be defined below.

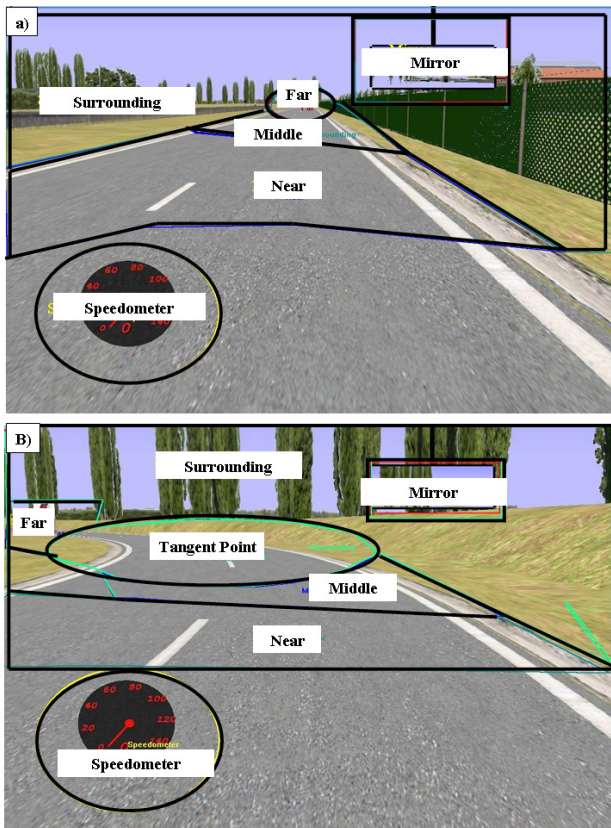


Figure 2. Areas of interest: (a) straight lines; (b) bends.

- *Eye fixation parameters.* (a) Mean duration of fixation (high if symbolic processing); (b) Percentage of time spent in particular areas of interest in straight lines and in bends (near area related to reactive processing; mirror related to time-sharing between trajectory control and traffic interaction management). (Figure 2); (c) Number of gazes devoted to pedestrians on the road verge (related to time-sharing).
- *Driver-car behaviour.* In this paper, we will only consider speed and speed variability adjusted by mean speed (as a co-variable: when mean speed increases, the difference between straight line and bend speed also increases). Speed will be used to identify the instruction effect. Adjusted speed variability will be interpreted in terms of reactivity.
- *Subjective assessments.*
 - *Baseline satisficing performance.* Before driving the simulator, each participant was invited to choose the

most important and then the least important driving performance assessment criteria: speed, regulation compliance and safety.

- *Situation mastery assessment.* After each lap, the participant was invited to answer on a four-point scale from very low to very high.
- *Memory of pedestrians.* After each experimental lap, two photographs were shown to participants. These showed part of the road and were taken from the driver's point of view. In the first photograph, no pedestrian was present. The second photograph was then displayed, in which a pedestrian had appeared. The participant was invited to say whether the photograph was the same as had been seen before. In the case of a negative reply, the participant was asked to say what was different. Memory of pedestrians was analysed in relation to the number of gazes to pedestrians and to the mirror, in order to evaluate time-sharing skills within the driving task.
- *Perceived performance assessment.* After each experimental lap and after the situation mastery assessment, each participant was invited to say which performance criteria was the most satisfied and which was the least satisfied (see baseline satisficing performance given above).

As usual, for numerical variables and comparisons with one degree of freedom, in order to conclude whether a sample effect (δ) is non-null on the basis of an observed effect (d), a Student's t -test of significance was calculated. The t -tests were associated with an observed two-tailed threshold (p). However, to draw conclusions in terms of population effect sizes and go beyond a conclusion in the sole terms of non-null effects, a variant of Bayesian statistical inference (fiducial inference: [11, 12, 13]), which considers test power, was used. On the basis of a maximal a priori uncertainty, the technique enables the user to emit a probabilistic judgement on the population effect size. For example, if the observed effect (d) can be considered as large, then a conclusion such as: "there is a high probability (guarantee γ) that the population effect is larger than a notable value" is tried ($P(\delta) > a = \gamma$; shortly $\delta > a$). Conversely, if the observed effect is negligible, the expected conclusion is that "there is a high probability that the absolute population effect is lower than a negligible value", ($P(|\delta|) < \epsilon = \gamma$). All fiducial conclusions below will be given with the guarantee $\gamma = .90$. When no relevant conclusion could be reached, at least with this guarantee, we have noted this as "no gen.", meaning that no generalisation in terms of population effect size could be reached. When the comparison has more than one degree of freedom, the effect indicator selected is λ , the quadratic mean.

The variables from the subjective assessment were analysed with the binomial law. The exact binomial test with a two-tailed threshold was used to discover when the answers differed from random events. The hypothesised probability of success chosen was $1/r$ (where r is the number of possible responses, that is, 3 or 4) and the number of trials was given by the size of the considered group (BI or CTRL). On the other hand, the group differences were tested with the Schwarz Bayesian Index, BIC (Schwarz, 1978). This index permits the selection of the best model from amongst those tested (the null hypothesis vs. the group difference model). The model retained is the one that obtains the smallest BIC; that is, the model that presents the best compromise between likelihood and economy (minimum number of parameters).

Main Hypotheses

Instruction Effects

Instruction type is assumed to directly determine the satisficing performance. The effect of instruction on actual performance is expected to be massive.

Scenario Effects

The effect of scenario on performance is not direct. It is mediated by whether or not conflict is generated between what actually happens and the satisficing performance, and whether or not there is an arbitration facility. However, the scenario type could have a global, if slight, effect on performance.

Combination of Type of Instruction and Type of Scenario

Table 1 sums up our expectations when combining the two independent variables.

Table 1. Conflict between type of instruction and type of scenario. Arbitration between performance criteria.

	Speed Scenario (slow car in front + faster car to the rear)	Safety Scenario (only slow car in front)
Speed Instruction	CONFLICT between reality (slow car) and satisficing performance WITH FACILITATION (faster car in favour of speed)	CONFLICT WITHOUT any FACILITATION
Safety Instruction	CONFLICT between reality (faster car) and satisficing performance WITH FACILITATION (slow car in favour of safety)	NO CONFLICT

The interaction between instruction and scenario is related to a possible conflict between the reality and satisficing performance. Conflict resolution can be facilitated by a particular property of the scenario in itself. Such a property can modify the satisficing performance in accordance with the instruction, with a possible change in cognitive compromise. This is the case for the speed scenario, whatever the instruction. The most difficult scenario to manage is the combination of speed instruction and safety scenario. Finally, there is no conflict when combining the safety scenario and the safety instruction.

Group and Interaction Effects

In line with existing literature, BI drivers are expected to adopt more reactive control, in relation to their difficulty to anticipate and plan. As far as symbolic control is concerned, the question remains open. Possible differences between the two groups in terms of the effects of instruction and scenario could be related to differences in cognitive control. The question of adaptation skill is also open.

RESULTS AND DISCUSSION

First, we will present the results of questionnaires on performance and situation mastery. Then, we will briefly examine the instruction and scenario effects. Finally, this paper will focus on three main results related to cognitive control dynamics. BI participants appeared to be more symbolic, more reactive, less

skilful at time-sharing, and adaptable to a certain extent by modifying their cognitive compromise.

Subjective Assessment of Performance and Situation Mastery

First of all, the BI group was more concerned with regulation compliance. In the baseline satisficing performance questionnaire, the regulation compliance criterion was seen as the most important for five out of the seven BI participants (binomial test: $p \leq .05$). Conversely, just two out of the six CTRL participants made the same choice. However, the answers given by the CTRL group did not appear to differ from random events (binomial test: $p > .99$).

Secondly, all the participants felt comfortable with the simulator. The situation mastery was always evaluated above 2 on the 4-level scale. Furthermore, for the experimental scenarios, the situation mastery was equal to or greater than 3 for half of the participants from both groups. With the BIC index, there was found to be no difference between groups.

Thirdly, for the CTRL group, the perceived performance was always clearly consistent with the instruction: participants (5 out of 6 for each scenario type, $p < .02$) declared that they had more than met the speed criteria in the speed condition and the safety criteria in the safety condition (5 out of 6, $p < .02$; and 4 out of 6, $p < .11$, respectively for the speed scenario and the safety scenario). Although the perceived performance was consistent with the instruction type for the BI group too, this consistency was weaker. Across the remaining laps, with the exception of the combined speed instruction and speed scenario, two or three BI participants said that they best satisfied the regulation compliance criterion. On the other hand, the CTRL participants, with one exception, did not choose this criterion. The difference in choices between groups was significant for two scenarios (BIC = 9.50 and BIC = 7.58 respectively for the null hypothesis and for the difference model). The results obtained for perceived performance are compatible with the BI group's concern for regulation compliance.

Thus, subjective assessments have shown that all participants had the feeling that they achieved a satisficing level of performance. Consequently, it may be possible, later on in the performance analysis, to interpret group differences as an indicator of satisficing performance set by each group during the driving activity in a given scenario. Whereas no major criterion of baseline satisficing performance was detected for the CTRL group, the BI group did seem to be more concerned with regulation compliance. Furthermore, during the experimental laps, some of the BI group members stated that regulation compliance was the best criterion to be achieved. On the other hand, the CTRL group mostly gave the best criterion as being either time, safety or speed, depending on which of these was emphasised in the instruction.

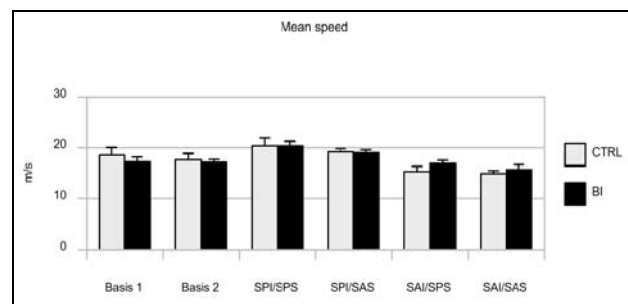


Figure 3. Mean speed.

Instruction and Scenario Effects

Figure 3 shows that group performances in terms of mean speed were quite similar for BI and CTRL groups. As expected, the speed instruction produced a higher speed than in the base laps ($d = 2.06$ m/s; $t(11) = 3.15$; $p < .01$; $\delta > 1.17$ m/s), whereas the safety instruction lowered the speed ($d = 1.95$ m/s; $t(11) = 3.47$; $p < .01$; $\delta > 1.18$). Thus, the instruction had a clear effect on the satisficing performance. The speed scenarios induced a higher speed than the safety ones (respectively 18.23 m/s vs. 17.20 m/s; $d = 1.03$ m/s; $t(11) = 2.86$; $p < 0.02$; $\delta > 0.54$ m/s). Thus, the faster car to the rear played a major role in the arbitration of satisficing performance than the slower car when they are used together (speed scenario).

Symbolic Control

The mean fixation duration times obtained by each group in straight lines are presented in Figure 4. The location of control on the subsymbolic/symbolic dimension, measured by mean fixation duration, shows some contrasts in straight lines and small differences in bends.

Figure 4 shows that, in straight lines, cognitive control appeared to be more symbolic for the BI group since this group had a longer mean fixation duration ($d = 98$ ms; $t(8) = 2.10$; $p < .07$; $\delta > 32.81$ ms). The location of cognitive control on the symbolic/subsymbolic dimension remained quite stable for all laps for the CTRL group ($l = 18.64$ ms; $F(5,25) = 0.23$; $p > .95$ ns; $|\lambda| < 61.08$ ms). (λ is the quadratic mean, which can be interpreted as the fluctuation of the mean duration from one lap to another.) However, it varied for the BI group ($l = 73.94$ ms; $F(5,15) = 1.08$; $p > .41$ ns; $\lambda > 64.47$ ms). Compared with the CTRL group, the cognitive control of the BI group was much more symbolic in the first base lap ($d = 178.27$ ms; $t(8) = 2.02$; $p < .08$; $\delta > 55.24$ ms) and in the experimental lap containing the safety instruction and the safety scenario ($d = 126.76$ ms; $t(8) = 3.31$; $p < .02$; $\delta > 73.28$ ms), contrary to all others ($d = 71.16$ ms; $t(8) = 1.32$; $p > .23$ ns; no gen.). So, the BI group's cognitive control was always more symbolic than that of the CTRL group, particularly in the laps where the driving task was less demanding.

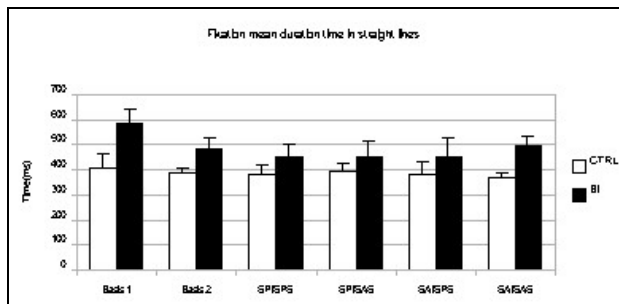


Figure 4. Mean fixation duration in straight lines with standard errors.
SP: Speed; SA: Safety. I: Instruction; S: Scenario.

The analysis showed that the group difference was negligible in bends ($d = 1.86$ ms; $t(8) = 0.09$; $p > .93$ ns; $|\delta| < 38$ ms) with no particular effect in each lap (from the six comparisons: $d_{max} < 60$ ms; $t_{max}(8) < 0.40$; $p > .68$ ns; $lim_{max} |\delta| < 84$ ms).

The results in straight lines show that the location of the cognitive control along the symbolic/subsymbolic dimension, measured by the mean fixation duration time, remained stable for the CTRL group. On the contrary, it varied for the BI group, being always more symbolic (longer fixations) than for the CTRL group, particularly in the laps with fewer constraints

(base and SAI/SAS). This means that BI participants can be more subsymbolic in conflicting situations (in all others).

Reactive Control

The percentage of time spent in particular areas of interest for straight lines is presented in Figure 5. The CTRL group spent more time looking at the far area than the BI group ($d = 25.14\%$; $t(8) = 1.65$; $p < .14$; $\delta > 3.83\%$). The main group difference relates to the near area; the BI group fixated more on this ($d = 27.42\%$; $t(8) = 2.98$; $p < .02$; $\delta > 14.55\%$). In the curves, the CTRL group looked slightly more at the tangent point ($d = 8\%$; $t(9) = 1.95$; $p < .09$; $\delta > 2.45\%$) than the BI group.

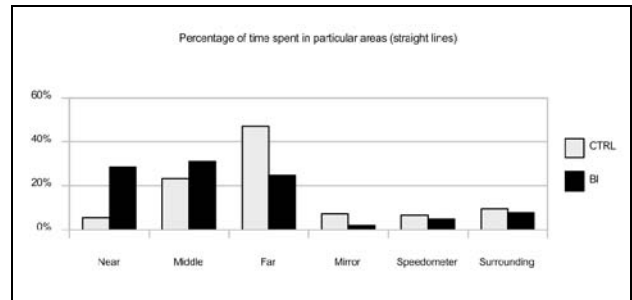


Figure 5. Percentage of time spent in particular areas of interest (see Figure 2) in straight lines.
SP: Speed; SA: Safety; I: Instruction; S: Scenario.

These results show that BI participants devoted fewer fixations than CTRL participants to anticipation areas. Such behaviour can be interpreted as more reactive, which was confirmed by the analysis of speed parameters.

Figure 6 presents the percentage of time per lap that was spent looking in the near area in straight lines. It can be seen from this figure that the cognitive control reactivity of the BI group was not stable over the laps. In comparison to the CTRL group, it decreased in the only non-conflicting experimental lap (safety instruction and safety scenario: $d = 4.54\%$; $t(8) = 0.63$; $p > .55$; $|\delta| < 5.61\%$). This was the lap with the safety instruction and the safety scenario. It was stronger in all others ($d = 32\%$; $t(8) = 3.02$; $p < .02$; $\delta > 17.19\%$). No other particular significant differences were found.

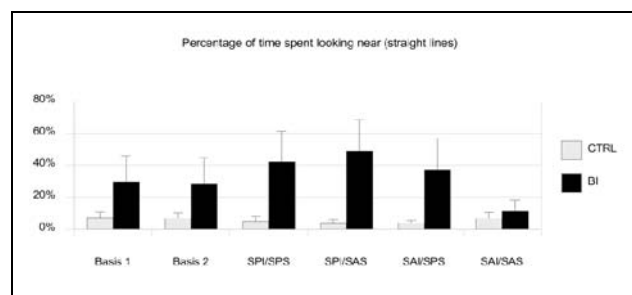


Figure 6. Percentage of time spent looking near in straight lines with standard errors.
SP: Speed; SA: Safety; I: Instruction; S: Scenario.

Figure 3 shows that the effect of the faster car to the rear is clear for the BI group. With the safety instruction, the BI group drove slightly faster in the speed scenario than in the safety one ($d = 1.22$ m/s; $t(6) = 2.13$; $p < .08$; $\delta > .39$), whereas for the CTRL group this difference is small ($d = 0.36$ m/s) and non significant ($t(5) = 0.49$; $p > .65$; no gen.). Thus, the BI participants were sensitive (reactive) to the pressure of the faster car to the rear, modifying their satisficing performance in favour of the speed criterion.

The mean speed variability, adjusted by mean speed, was increased by the speed instruction (Figure 7), in comparison with the base laps ($d = 1.09$ m/s; $t(8) = 2.02$; $p < .08$; $\delta > 0.34$ m/s). On the other hand, the safety instruction produced a small difference with the base laps ($d = 0.06$ m/s; $t(8) = 0.22$; $p > .83$ ns; $|\delta| < 0.49$ m/s) and the effect of the scenario type was small ($d = 0.26$ m/s; $t(8) = 1.13$; $p < .08$; $|\delta| < 0.58$ m/s). This result shows that the effect of the instruction factor previously detected in the performance analysis modified the reactivity of the drivers' cognitive control. Consequently, they became less anticipative with the speed instruction.

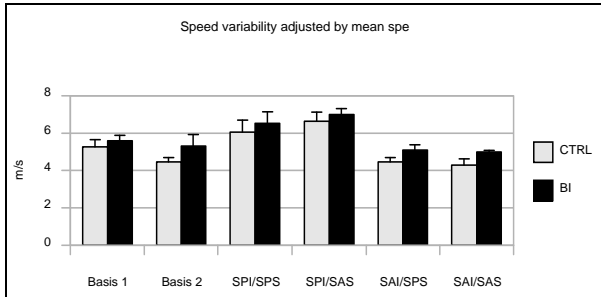


Figure 7. Speed variability adjusted by mean speed.
SP: Speed; SA: Safety; I: Instruction; S: Scenario.

Moreover, the mean speed variability was slightly greater for the BI group ($d = 0.55$ m/s; $t(8) = 1.48$; $p > .18$ ns; $\delta > 0.03$ m/s). This difference was mostly due to the laps where a safety instruction was given ($d = 0.64$ m/s; $t(8) = 1.58$; $p > .16$ ns; $\delta > 0.07$ m/s and $d = 0.71$ m/s; $t(8) = 2.02$; $p < .08$; $\delta > 0.22$ m/s respectively for the speed scenario and the safety scenario). This result is in accordance with a more reactive (less anticipative) cognitive control for the BI group. This was particularly the case when the safety instruction was introduced, where the BI group did not reduce its reactivity to the same extent as the CTRL group.

Time-sharing

In straight lines, the CTRL participants looked in the mirror more ($d = 5.66\%$; $t(8) = 4.52$; $p < .01$; $\delta > 3.91\%$) than the BI participants. This can be interpreted as a weaker ability to switch to a secondary part of the task.

As we can see in Figure 8, the time spent looking in the mirror was higher in the experimental laps than in the base laps. However, the percentage changed from one lap to another more for the CTRL group ($l = 6.68$; $F(5,25) = 5.91$; $p < .001$, $\lambda > 5.60\%$) than for the BI group ($l = 1.80\%$; $F(5,15) = 1.56$; $p > .24$; ns, $|\lambda| < 3.11\%$). That is to say, the CTRL participants looked in the mirror more during the speed scenario ($d = 7.86\%$; $t(5) = 4.70$; $p < .005$, $\delta > 5.40\%$) compared to the BI participants ($d = 1.14\%$; $t(3) = 1.05$, $p > .37$; $|\lambda| < 3.12\%$). This result suggests that the BI group was less able to process a secondary part of the task, especially when it was necessary to check on the behaviour of the faster car to the rear.

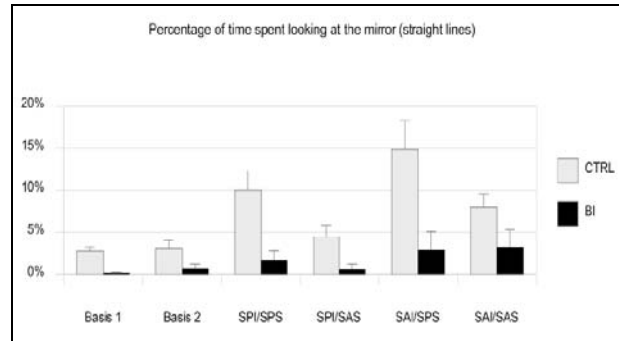


Figure 8. Percentage of time looking at the mirror (straight lines).
SP: Speed; SA: Safety; I: Instruction; S: Scenario.

Overall, all participants made at least one gaze at the pedestrian on the road verge. However, the control participants made more than twice as many gazes (3.21) than the BI group (1.54). The difference is significant and notable ($t(10) = 2.25$; $p < .05$; $\delta > 0.65$). Half of the CTRL participants remembered having detected the pedestrian on the edge of the road. However, none of the BI participants stated that they had detected the pedestrian. This difference is quite significant (Fisher exact probability: $p = .07$). Thus, the BI participants seemed to be less able than the CTRL participants in detecting an unpredictable but relevant event, such as a pedestrian on the road verge. This result is consistent with the difficulty in processing information acquired from looking in the mirror.

CONCLUSION

To sum up, five main results can be gathered from this experiment.

1. BI participants favoured regulation compliance, whereas CTRL participants did not express any priority among the performance criteria. This could possibly be related to the BI participants' feeling that there is a risk of losing their driving license by breaking the rules. BI participants evaluated their situation mastery as positively as CTRL participants. Thus, the participants considered that their adaptation to the instructions and scenarios was acceptable. They also considered that their satisficing performance was reached. The participants found their perceived performance consistent with instructions, but BI participants were less affirmative than CTRL participants. Overall, the actual performance can be interpreted as the satisficing performance and adaptation was successful, leading participants to situation mastery.
2. Instruction and scenario types produced the expected effects on the satisficing performance. Thus, participants succeeded in modifying their satisficing performance to adapt to these factors. Although the main BI participants' baseline satisficing performance criterion was regulation compliance, they were able to give priority to speed when needed.
3. BI participants adopted a more symbolic control, especially in easy situations. However, in order to adapt to more difficult situations they needed, but also showed their ability, to change their cognitive compromise toward a less symbolic control. CTRL participants were able to face every kind of situation with the same cognitive compromise, with less symbolic control than BI participants. Thus, on this basis, CTRL participants' cognitive compromise appeared to be more robust.

4. BI participants were more reactive than CTRL participants in several respects. They spent more time looking in the near area. They were also more sensitive to the car in the rear than to the safety instruction. The observations show that this was not the case for CTRL participants, although without generalisation. BI participants adopted a more variable speed. They were able to reduce their reactivity (near area fixations) in the lap where there was no conflict. In other words, BI participants were less adaptable in terms of reactivity than in terms of symbolic processing. Again, CTRL participants showed a more robust compromise from the reactivity point of view.
5. BI participants were less able to process secondary subtasks within the driving task (looking in the mirror and at pedestrians). Multitasking is a typical feature of driving. The drivers must navigate (find their route), control the trajectory in the medium term, handle various pieces of equipment inside the cockpit, converse with passengers, and so on. However, the trajectory control task is composed of several subtasks. The main subtask is maintaining the car in the lane, as far as possible. Many other trajectory control subtasks are also implied; for example, those related to managing interaction with traffic (pedestrians, traffic to the rear, etc.).

BI participants showed a more symbolic and more reactive control. This result has a theoretical value because it validates the model put forward by Hoc and Amalberti in which two orthogonal dimensions are considered. Symbolic control is not necessarily anticipative; it can also be reactive. The BI participants' difficulties in multitasking can be explained by their cognitive compromise. Symbolic control is costly in terms of cognitive resources, and slow. Reactive control implies frequent information-gathering for the same subtask. These two features can explain why BI participants were confronted with difficulties in multitasking.

Another validation of the model is the dynamic feature of cognitive control when adapting to various situations, especially conflicting ones in terms of satisficing performance. BI participants were able to modify their cognitive compromise in terms of symbolic/ subsymbolic tradeoff as well as in terms of reactivity. CTRL participants kept their distribution of control along the symbolic/ subsymbolic dimension more stable than BI participants. Overall, the participants were able to change their satisficing performance in order to adapt to experimental situations. However adaptation could be more costly for BI participants than for CTRL participants, because it also implies a change in cognitive compromise.

Although attention was not directly addressed in this study, albeit globally through psychological testing, some discussion is needed. Following the approach made by James [15], we consider attention as a filtering resource to manage the information bombing we experience in order to process the most relevant part. There are three main attentional processes [4]: selective attention (the main filtering process), divided attention (the ability to filter distinct parts), and sustained attention (the duration of the activation/inhibition process). From our point of view, attention is not cognitive control but a resource for cognitive control.

The symbolic feature of BI participants' cognitive control could be related to selective attention difficulties. BI participants could have deliberately invested more resources in the processing of information from the main subtask, in order to avoid being distracted by signals coming from their surroundings and to counter difficulties in receiving relevant information. However,

another possible explanation for the fact that BI participants relied on symbolic control more than the CTRL participants could be that the former have suffered from a decrease in processing speed. This is incompatible with an over-reliance on routines, but is acceptable for symbolic processing. The BI participants' reactivity is clearly related to frontal damages, which are correlated with planning and anticipation deficits. The BI participants' difficulty to manage multitasking could be related to divided attention difficulties; however, it could also be a direct consequence of a symbolic and reactive control. A new research program has been established in order to find answers to these questions. A better understanding of these mechanisms is needed if one wants to derive implications in terms of rehabilitation.

In the short term, some implications can be drawn in terms of driving assistance. Trajectory control support devices are currently being developed [15] for longitudinal (e.g., Cruise Control) as well as lateral control (e.g., Lane Keeping Warning or Assistance Systems). They are designed not only for comfort but also for safety, and in such a way that they are also useful for normal drivers. Car manufacturers are interested in equipping vehicles with driving assistance devices that are relevant to all drivers, not just those that are disabled. Apart from rehabilitation training sessions, the type of assistance currently in use, both under development and under study, could be of interest for BI drivers.

For whatever reason, if BI drivers need to exert a symbolic control of the trajectory, and consequently are confronted with multitasking difficulties, driving assistance could improve the situation. For example, a Lane Keeping Assistance System can correct the trajectory automatically, saving processing time for the driver, and allowing the driver to manage other subtasks such as interaction with the traffic. Another research program should be devoted to the evaluation of "normal" driving assistance for this type of driver. Today, elderly drivers are frequently used for the evaluation of driving assistance devices: such evaluations could be extended to include other populations.

Besides, there could be some commonality of functional impairment among various populations, including: the elderly, BI people, and Alzheimer and stroke sufferers. In this case, driving assistance devices could be of interest for several populations. This is one of the reasons for comparing BI drivers with young drivers and elderly drivers. The other reason is to gain an understanding of the role of metaknowledge in adaptation. Young drivers are not skilful at driving, as a more symbolic control of the task is needed before developing routines. In addition, they do not have a rich metaknowledge and cannot avoid situations where they perform poorly as drivers. Elderly drivers could also be more symbolic, for similar reasons to BI drivers, but with the addition of metaknowledge. It could be relevant to identify to what extent BI drivers have metaknowledge of their impairment and implement coping strategies. This point could be of importance in designing rehabilitation sessions.

ACKNOWLEDGMENTS

This work has been done within the French research network GDR CNRS 3169 Psycho Ergo. The experiment was carried out with the collaboration of ARTA and Johanna Pothier, who invited the BI drivers to participate in the study on the basis of neuropsychological testing. We thank the participants for their contribution to the study. We are grateful to René Amalberti for early discussions on this experimental project and to the INRETS MSIS team for their technical assistance. We would like to thank Susan Watts for English-language revision.

REFERENCES

- [1] Debelleix, X. 2001. When and how should people be authorised to drive after severe head injuries? *Europa Medicophysica*, 37, 201–208.
- [2] Lew, H. L., Poole, J. H., Lee, E. H., Jaffe, D. L., Huang, H. C. and Brodd, E. 2005. Predictive validity of driving-simulator assessments following traumatic brain injury: a preliminary study. *Brain Injury*, 19, 177–188.
- [3] Hoc, J. M. and Amalberti, R. 2007. Cognitive control dynamics for reaching a satisficing performance in complex dynamic situations. *Journal of Cognitive Engineering and Decision Making*, 1, 22–55.
- [4] Parasuraman, R. and Nestor, P. G. 1991. Attention and driving skills in aging and Alzheimer's disease. *Human Factors*, 33, 539–557.
- [5] Simon, H. A. 1969. *The science of the artificial*. MIT Press.
- [6] Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- [7] Underwood, G., Chapman, P., Brocklehurst, Underwood, J. and Crundall, D. 2003. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, 46, 629–646.
- [8] Land, M. F. and Horwood, J. 1995. Which parts of the road guide steering? *Nature*, 377, 339–3405.
- [9] Land, M. F. and Lee, D. N. 1994. Where we look when we steer. *Nature*, 369, 742–744.
- [10] Mars, F. 2008. Driving around bends with manipulated eye-steering coordination. *Journal of Vision*, 8(11), Art. 10, 1–11, <http://journalofvision.org/8/11/10/>, doi:10.1167/8.11.10.
- [11] Lecoutre, B. and Poitevineau, J. 2005. Le logiciel « LePAC » [The software “LePAC”]. *La Revue de Modulad*, 33 (whole volume). French and English version. Retrieved [29.06.09] from: <http://www.univ-rouen.fr/LMRS/Personne/Lecoutre/PAC.htm>.
- [12] Rouanet, H. 1996. Bayesian methods for assessing importance of effects. *Psychological Bulletin*, 119, 149–158.
- [13] Rouanet, H. and Lecoutre, B. 1983. Specific inference in ANOVA: From significance tests to Bayesian procedures. *British Journal of Mathematical and Statistical Psychology*, 36, 252–268.
- [14] James, W. 1890. *The principles of psychology*. Cosimo Classics.
- [15] Hoc, J. M., Young, M. S. and Blosseville, J. M. 2009. Cooperation between drivers and automation: implications for safety. *Theoretical Issues in Ergonomics Science*, 10, 135–160.

Distributed Cognition in Ship Navigation and Prevention of Collision

Thomas Koester
FORCE Technology
Division for Maritime Industry
Hjortekaersvej 99
DK-2800 Kgs. Lyngby,
Denmark
tsk@force.dk

Nikolaj Hyll
FORCE Technology
Division for Maritime Industry
Hjortekaersvej 99
DK-2800 Kgs. Lyngby,
Denmark
nnh@force.dk

Jan Stage
Aalborg University
Department of Computer
Science
Selma Lagerlöfs Vej 300
DK-9220 Aalborg East,
Denmark
jans@cs.aau.dk

ABSTRACT

In this paper we investigate how technology can help the navigator to a better performance. We use three examples based on observations onboard three ships to show, how technology can support the work of the navigator and thereby enhance the performance. Our analysis is based on the paradigm of distributed cognition. The analysis shows that information persistence and information availability are essential parameters in optimization of the performance. Therefore the designer of maritime equipment could benefit from identifying need for persistence and need for availability. Some information which could be a great help to the navigator could in fact already be present as data in the cognitive system, but because it has not yet been processed, transferred, recorded or displayed in the right way it is not persistent enough or not available when needed for the navigator to use it and take advantage of it.

Keywords

distributed cognition, cognitive system, maritime technology, information persistence, situation awareness

ACM Classification Keywords

H.1.2 [Models and Principles] User/Machine Systems – *human factors*; H.1.2 [Models and Principles] User/Machine Systems – *human information processing*; H.5.3 [Information Interfaces and Presentation] Group and Organization Interfaces – *theory and models*.

INTRODUCTION

It is in the navigation of a ship important that the navigator on the bridge of the ship knows where the ship is (the position) and where it is going (the route). It is also in some cases important to know where it has been (the track). Calculating the position of the ship, the route and the track is among the most important work tasks of the navigator. Throughout the history of navigation means and methods used in these work

tasks has improved. New technologies support the navigational work and add ease, speed and accuracy.

Another important work task for the navigator is to avoid collision with other ships according to the international set of rules for prevention of collision at sea – in daily terms called colregs. In order to prevent collision the navigator needs to know where other ships in the surrounding area are (their position) and where they are going (their route). This task is also facilitated through new and advanced technology. The navigator should also be able to communicate with navigators onboard other ships about positions, routes and intentions. This communication is important in the prevention of collision. It is possible to communicate about intentions and plans for the route and thereby to take necessary precautions well in advance in order to prevent collision.

The technological development in navigation has grown in parallel with the increased speed of ships, growing intensity of traffic, higher focus on safety and demands for operation in weather and visibility conditions which in earlier days of navigation would cause delays. Technology is developed to assist the efficient and safe navigation, and the availability of the technology will open up for higher speeds, less delays and operation with smaller safety margin. The technology will thereby assist the navigator in the optimization of the route and manoeuvres. Navigation and prevention of collision are as mentioned two of the most important work tasks of the crew on the bridge of the ship. And they are – with some variation – the same for all kinds of ships: Fishing vessels, ferries, cargo vessels, naval ships etc. Hutchins has in his book “Cognition in the wild” [5] made a comprehensive analysis of the work tasks of the crew on the bridge of a naval ship. The analysis is based on the paradigm of distributed cognition.

We are in this paper looking at work tasks within navigation and prevention of collision. This work is performed by the navigator on the bridge of the ship. The question we would like to answer is Q1: How can technology help the navigator to a better performance? We will answer that on the basis of empirical examples. Through the analysis of these examples we will try to answer the following question Q2: How can new maritime technology be designed in order to support the work tasks of the navigator even better? Our answer to Q2 is based on the analysis of the examples in Q1 and our findings in this analysis. The analysis points at two important features or functions in technology, which are essential for the enhancement of the performance and the optimization of the operation.

The examples used in this paper are by no means exhaustive. We could make more observations and find a lot of other examples. But the examples are illustrative and the analysis of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

the examples points out two important features or functions which are important to take into account when designing new technology for maritime use. We took advantage of the paradigm of distributed cognition and the concept of situation awareness because it helped us in the analysis and on our way to the conclusion.

The focus of the paper is three specific problems related to navigation and prevention of collision. For each problem we have looked at how technology is or potentially could be used to compensate for the problem. We are thereby looking closer at three different technologies on the bridge which supports safe and efficient navigation and prevention of collision: Electronic chart displays, AIS and the VHF replay function. The effect of the technologies is illustrated by the three empirical examples. The electronic chart display helps the navigator in the precise manoeuvres and navigation involved in a search and salvage operation, and the AIS and VHF replay function helps the navigator prevent collision while still maintaining optimized manoeuvres and routes. A common theme in all three technologies is information related to the past, present and future in navigation and prevention of collision at sea. It is as mentioned important to know where the ship has been (the past), where it is (present) and where it is going (future). It is further important to know where other ships in the vicinity have been, where they are and where they are going. The electronic chart provides information about where the ship has been in the past. This information is used by the navigator to plan for the future. The AIS provides information about present positions, speed and course of other ships. This information is used by the navigator in the planning of future manoeuvres. And the VHF replay provides information about what has been said by other ships in the past about future intentions and choice of route. This information can be used by the navigator to plan own route and optimize according to routes and choices made by the other ships.

We have analysed each of the three empirical examples from the perspective of distributed cognition. The analysis of the cognitive system is also discussed in relation to how the need for information persistence and information availability is handled by the system. In addition to that we are also discussing how the technology supports the situation awareness of the navigator. The examples show how the introduction of certain technologies in the cognitive system has the potential of improving efficiency and safety by compensating for problems in the performance of the task. It also shows how the technology support the situation awareness of the navigator and how it support the need for information about past, present and future and thereby assist the navigator in the planning of future manoeuvres and routes. From the analysis of the examples we conclude that persistence and availability are important parameters when it comes to the design of new technology supporting the work tasks of the navigator. By adding information persistence and information availability the navigator is helped to a better performance.

Theoretical Background

The examples we have collected in our studies are primarily analysed on basis of the paradigm of distributed cognition. These analyses illustrate the effect of adding a new technology to the cognitive system. We have also analysed our examples in relation to the concept of situation awareness.

Distributed Cognition

In the distributed cognition approach we take a very broad view of the entire system involved in navigation and

prevention of collision. In relation to the concept of situation awareness we have the cognition of the navigator in focus, and look from the perspective of the navigator on how cognition is supported by technology. The difference between the distributed cognition perspective and the situation awareness perspective is that technology is viewed as a part of the system in the first approach while it is viewed as an aid to the navigator in the latter approach. Sharp, Rogers and Preece [12, p. 130] describe the distributed cognition paradigm as follows: "This way of describing and analyzing a cognitive activity contrast with other cognitive approaches, e.g. the information processing model, in that it focuses not on what is happening inside the head of *an* individual, but what is happening across a system of individuals and artefacts".

Since the work onboard a ship in most cases is a team performance by the crew, and since the crew is interacting with other people such as for example the pilot and crews on board other ships, the distributed cognition approach seems to be well suited as instrument for the analyses. Further the crew usually works with a wide selection of equipment, technology and artefacts. The equipment and technology can be designed and produced by different companies and therefore utilize different user interfaces. The introduction of new equipment on board old ships can result in a mixture of old and new equipment covering a very broad range of user interfaces. Having a situation far from "one person, one computer" the distributed cognition approach is expected to have a better explanation power when used in the analysis than approaches based on for example the traditional information processing paradigm [12].

The distributed cognition approach describes a cognitive system "...which entails interactions among people, the artefacts they use, and the environment they are working in" [12, p. 129]. The *people* are in this study the crewmembers on board the ship as well as the crews on board other ships in the vicinity. Crewmembers on board the ship communicate directly verbally and using internal communication systems and mobile radios. Communication with crews on board other ships is maintained through VHF radio and using signalling horn. The open channel 16 on VHF radio is often used for short announcements. The advantage is that messages sent out on VHF can be heard by all ships in the nearby area. It is not dedicated to one ship or person only such as for example mobile phones. Further, actions performed by crews through manoeuvres of the ships can be viewed as a sort of communication. The manoeuvres can be observed directly visually – if the weather and visibility permits – or indirectly by the use of radar or AIS. A ship can for example indicate that it will give way to another ship by performing a turn which is large enough to be observed visually and on radar and AIS. On radar the turn is indicated by a vector showing the speed and direction of a target ship. On AIS the same information is shown by numbers changing or – if AIS-information is transferred to electronic charts or radar – graphically on the electronic charts or radar screen. The *artefacts* the people use are the ship and the equipment, technology etc. on the ship. The crew use different equipment and technology on the bridge. In this paper focus is on electronic charts, AIS and VHF. Electronic charts are used for navigation. Own ship's position is shown on the electronic charts screen and it has different functions for navigational support. AIS is used to identify other ships and show information about them; for example their course and speed. VHF is used for verbal communication with other ships, pilot stations, harbour authorities and other land bases stations etc. The *environment* the people are working in is the area in which they navigate their ship. Parameters in the environment are weather, visibility

and current, other ships in the area (other cognitive systems), ports and other topographical elements. It can be discussed if the ship itself is an artefact or a part of the work environment. Since the ship is manipulated and controlled by the crew (under influence from weather and current) we have in this study considered it to be an artefact.

The interactions in the cognitive system are described “in terms of how information is propagated through different media” [1, p. 130]. Information can for example move from artefact to people when they look at a display, or it can move from people to artefact when something is written on paper. This process where information is transformed is in the paradigm of distributed cognition referred to as changes in representational state. In this paper our focus is on the identification of the elements in the cognitive system and the analysis and description of how information is transferred and transformed between these elements.

Situation Awareness

The concept of situation awareness has been studied by – among others – Mica Endsley [2]. She developed the so-called SA-model describing three successive levels of situation awareness. The model is often illustrated like this:

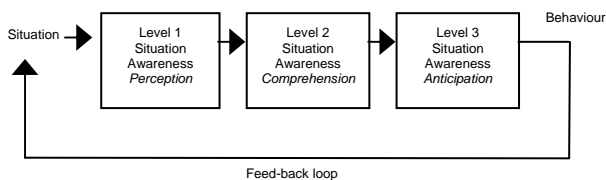


Figure 1. The traditional way of illustrating the SA-model.

The model has three successive levels of situation awareness: Perception, comprehension and anticipation. And a feed-back loop from behaviour to situation.

This means that the behaviour will change the situation (making it a new situation) and new situation awareness on level 1, 2 and 3 needs to be obtained.

The model and the concept of situation awareness have been used by many authors. It has for example been used to explain nuclear power plant accidents [3] and in maritime risk analyses [14]. But it has also been subject to critique. Sarter & Woods [11, p. 45] says that situation awareness is a ubiquitous phrase and that a commonly accepted definition is missing and Dekker [1, p. 49] explains that “...situation awareness remains ill defined”.

METHOD

The method used in this study is inspired by the work by Margareta Lützhöft and Ed Hutchins. Both authors use onboard interviews and observations as resource for data collection. Hutchins methods are described as “anthropological” [5]. The method used by Lützhöft is described as “problem-oriented ethnography” [8, p. xi]. The observer’s domain knowledge is important according to Hutchins: “Ed Hutchins emphasizes that an important part of doing a distributed cognition analysis is to have a deep understanding of the work domain that is being studied. He even recommends, where possible, that the investigators take steps to learn ‘the trade’ under study and become an accomplished pilot or sailor (as he has done himself in both cases).” [12, pp. 394–395]. The observations in our study were made by two of the three authors. Both observers have an MSc in psychology. With respect to domain knowledge

one observer has – apart from the MSc in psychology – certificates and experience as captain including work on ferries. The other observer has not a commercial seagoing background, but he has gained extensive domain knowledge through field studies on board 23 different ships (ferries, fishing vessels, cargo vessels etc.) conducted from 2000-2009. Apart from that he has in the same period been teaching as instructor on human factors and ship simulator courses for crews from several ship-owners. He also holds Danish certificates as Yacht Master of 3rd, 2nd and 1st class and has 25 years of sailing experience with pleasure crafts. The work onboard a modern ferry or container ship is very different from pleasure craft sailing, but the basic principles of navigation, the waters in which they navigate and the rules for prevention of collision they have to obey are the same.

The collected data is notes taken by the observers and photos taken during the observations. The study is based on observations on board ships in normal operation. Even the presence of the observer on the bridge of the ship is part of normal everyday routine: The crews reported that they very often have visitors to the bridge talking and discussing with the crew. It can be discussed if the *reliability* is high or low. Can the observations be made again? On board the same vessel? On board other vessels? To test for this we asked the crews how they evaluated the scenarios we had observed. Was it something they had never ever experienced before or was it seen very often? The examples used in this paper are all evaluated by the crew as being very common on their particular ship in the specific operation. It can still be discussed – even when the scenarios are evaluated as being very common – if we can generalize from them. We actually do not know if our observations are special for the specific ship in that particular pattern of operation. This discussion is from a scientific viewpoint very relevant. It is also relevant when it comes to the discussion of technological support for the task. Are the observed problems frequent on many ships or are they limited to very few ships? The commercial interest in development of technological support could vary depending on this. But for our purpose – to demonstrate the role of the three different technologies in cognitive systems – we only need the three observed scenarios which we have used as illustrative examples. The power of our findings could of course be enhanced by adding more observations and examples, but this goes beyond the limitations of this paper. To express it in another way: We only need to see one example to know that it exists. From that we can not conclude how common it is, but we can conclude that it exists. When it comes to validity, we claim that the *ecological validity* of our study is high: We have made our observations on board real ships, in everyday situations, with real crews etc. The setting is realistic and naturalistic, and the observations can therefore be characterised as naturalistic rather than experimental. The result is a high ecological validity.

THE OBSERVED PROBLEMS AND THE TECHNOLOGICAL SOLUTIONS

We were in our observations able to identify three very specific problems related to specific work tasks. Technology can be used to compensate for and solve these problems and thereby support the work of the navigator and enhance the navigator’s performance. This can lead to general enhanced efficiency and safety of the operation. In the two first examples we observed the use of the technology. In the third example we observed a situation where technology available on the market (but not on the particular ship) potentially could have been used. By looking at these specific examples we see how technology can support the work of the navigator and help the navigator to

improve his or her performance. Apart from that we also identify the most important features or functions in the technology. From that knowledge it is possible to find directions on how technology should be designed and what really matters. In *the first example* the problem is to know exactly where the ship has been and where it has not been in a search and salvage operation. On some vessels it is extremely important to know exactly where the ship has been in the past – the track sailed – and where it has not been yet. The obvious example is mine sweeping where the track can be considered clean from mines and thereby open to navigation of other vessels. But also on board fishing vessels it can be important to know exactly where the ship has been. If the fishing net is destroyed or lost due to contact with some kind of underwater obstruction (such as a ship wreck) it could be of high value to the captain of the ship knowing exactly where that happened in order to avoid loss of fishing nets in the future. In our study we have observed a special purpose vessel searching for lost chain on the seabed and the use of an electronic chart display for this task. In *the second example* the problem is that the crew wants to optimize the departure from the port by coordinating it with the other ship leaving port at almost the same time. This is difficult because the line of sight between the two ships is obstructed by buildings on the quay. It is therefore not possible for the crew to see when the other ship starts moving visually or on radar. In this case we observed how the crew used the AIS-system to compensate for the problem. In *the third example* crossing traffic announce intentions and choice of route on VHF-radio in advance so that ferries can prepare and take precautions. The problem in relation to that is that the announcement is not persistent and therefore there is a risk that it is forgotten, and it is difficult to have the message repeated if it is misunderstood. Further there is the risk that the crossing traffic deviates from what is previously announced – by mistake or if plans are changed in the last minute. In this particular example the crew could have taken advantage of using the replay function in the Sailor RT-5022 VHF-radio, but this device was available on the ship we observed. We are not claiming that the three problems we have identified so far are the only problems present on board the ships we have observed or on ships in general. There are indeed many other problems. We are also not claiming that the three problems are the most important problems. The identified problems should be seen as illustrative examples on the basis of which the function of and elements in the cognitive system can be analysed and discussed. The problems illustrate how and to which extend work tasks are supported by technology on the bridge. Knowing about the problems we can discuss if available technology could be improved in order to counteract or compensate for the problem. And we can discuss how this should be done. Better understanding of the tasks and problems related to them can help manufacturers of maritime equipment in the development and design process of new technology for navigation and prevention of collision. Eventually this could lead to more efficient and safer operation.

THE THREE EXAMPLES

We have made observations on board in total 23 ships over a period of nine years, but for this study we have chosen specific scenarios from observations on three of these 23 ships: (1) M/V Poul Løwenørn – special purpose vessel, (2) M/V Mercandia IV – ferry on route between Helsingør, Denmark and Helsingborg, Sweden and (3) M/V Mercandia VIII – ferry on same route as Mercandia IV. We are going to use the three scenarios to illustrate and discuss the effect of three different types of technologies when they are introduced in the cognitive system:

(1) Electronic nautical charts, (2) AIS and (3) the replay function on the Sailor RT-5022 VHF radio. We will also – on the basis of the examples – discuss how the technologies in focus support the situation awareness of the navigator and how the technologies contributes with persistence by recording past events or assist the navigator in the prediction of future events.

Example 1: M/V Poul Løwenørn

The special purpose vessel M/V Poul Løwenørn operated by The Royal Danish Administration of Navigation and Hydrography serves as a buoy maintenance vessel. When the observations took place, the vessel was – among other things – engaged in search for lost anchor chain from buoys on the seabed. Lost anchor chain on the seabed could be an obstruction especially in relation to fishing vessels, and the chain is valuable because it can be used again. The chain can for example be lost if the buoy breaks loose and the rest of the chain sink to the seabed. The search was performed by dragging an anchor in a chain (a grapnel) over the seabed in order to catch the chain. The captain had to manoeuvre the vessel around in the area in order to drag the grapnel. The search was supported by electronic equipment including an advanced 3D seabed scanner and high precision GPS navigation equipment. The chain with the grapnel went from a winch on the deck over the rail in the starboard side of the vessel and down to the seabed. A crewmember on the deck held his foot on the chain to feel if the grapnel got hold of an object on the seabed. The search for lost chain was made in two positions. It did not give any result in the first position, but in the second position, the grapnel caught a chain on the seabed, and a significant amount of chain and concrete blocks (used as anchorage of buoys) was salvaged. The operation was a success. The task itself could be compared to “finding a needle in a haystack”. The sea surface looks the same, and when manoeuvring around in the area it is without special equipment almost impossible to find out if the search covers the area or if it is only done in the same spot over and over again.



Figure 2. The electronic chart display showing the track of M/V Poul Løwenørn in the search and salvage operation.

The captain manoeuvring the vessel needs assistance in the question “where have I already done my search?”. From the answer to this question, the captain decides where to search next. Information about the past is in this task important for decisions about the future. The task could be compared to for example the task of mine sweeping at sea. It is also in mine sweeping important to collect information about where mines have been swept and where it has not been done yet. The captain is therefore using an electronic system showing on a map display the exact track of the search (see Figure 2). Using this system, he is able to see where he has already searched in the past and from that decide where to search next.

Analysis of the Example

The elements in the cognitive system in this particular task are: (1) People: The captain, (2) Artefacts: Electronic nautical chart based on GPS information (and also 3D seabed scanner, but this is not in focus in the analysis of the example), and (3) Environment: The area and the seabed. If we look at how the information about the past is managed in the cognitive system we see that the electronic equipment is of great importance. Without the electronic display it would be practically impossible for the captain to get information about where he had already searched, and this could make search much less effective by adding the risk that the same areas are searched multiple times and that there are “blind spots” in the search. The information on the electronic display about where the ship has already been matches in this case the captain's need for information in that specific work task perfectly. In the case of searching for lost chain it is important that the captain have information about what has happened in the past (where and how he has already searched) in order to make decisions about what to do in the future. The information is available through an electronic system, and it is transformed from this system through a graphical representation on a display to a mental representation. The critical component in the cognitive system in this particular task is the graphical display. The position can be calculated by navigational techniques – even traditional techniques without electronic equipment. The position is much more precise with modern electronic navigation such as GPS, but in theory it is also possible to calculate the position using techniques from traditional navigation. If the position is calculated and recorded repeatedly when the ship is moving, the set of positions form a track. To get a precise track you need to calculate and record positions with a high frequency for example using GPS. Basically the GPS is only showing where you are at a given moment in a system of co-ordinates with latitude and longitude. The GPS system delivers the coordinates for the actual position with good and adequate precision and with a high frequency. The position can be transferred manually by the navigator to the nautical paper chart. This takes some time and therefore the position set out in the chart has some delay. When sailing ocean voyages over large distances this small delay is not critical, but when manoeuvring with high precision e.g. in the search for lost chain on the seabed it could be very critical. The position can also be transferred electronically to an electronic nautical chart on a display. Then you can immediately – without having to set out the position in a paper chart and without the risk for delay and possibly lack of precision (or making errors) – see the position of the ship in the chart. Without the electronic chart display information was transferred from the GPS to the navigator's mental representation and from that to a representation on a cognitive artefact (the nautical paper chart) and from that again to mental representation of the location of the ship. Introducing electronic chart displays in combination with GPS the information of the position is transferred directly from the GPS to the graphical representation and from that to the navigator's mental representation. The introduction of new technology in the cognitive system – in this case the electronic chart – makes the work task easier for the navigator and enhances the performance. Now information is automatically transferred to a graphical display. Beforehand this task has to be performed manually by the navigator. Or it was not performed at all making the search less effective. The electronic chart makes the information about the present position persistent by recording it graphically on a display. The recorded information represents the track sailed in the past. And having this track available it becomes easier for the navigator to make an efficient search. It is of critical importance that the GPS and

electronic chart presents data with high accuracy. It is also important that the system has a high update frequency. Otherwise the track will be less detailed. Both the high accuracy and the high update frequency differentiate the system from traditional methods where calculation of position and the plot of positions in a nautical paper chart could be a time consuming process with high potential for errors and lack of accuracy. The system we observed on board Poul Løwenørn had sufficient speed and accuracy for the purpose according to the crew. The electronic chart supports the need for persistence in the cognitive system. The information about where the ship has been in the past is lost if only GPS is used, but with the electronic chart it can be recorded as a track showing where the ship has sailed in the past. The electronic chart can be considered as a tool for preservation of information about past positions. This information is used by the navigator in the decisions made about where to go next. The information on the electronic chart is used to plan the future manoeuvres. Without this tool the operation in relation to the specific task of searching for lost chain would be much more difficult. It is also assumed that it would suffer from less efficiency and success. The electronic chart supports situation awareness at level 1 because it makes it possible to perceive the track of the ship. This track would – without the electronic chart – be invisible for the human eye. The improved situation awareness on level 1 could contribute to improved comprehension of the entire search for chain situation and thereby put the captain in a better position when he anticipates the remaining search procedure and decides where to search next (situation awareness level 3).

Summary of Findings in Example 1

The electronic chart presents data recorded from the GPS on a display in a way that they represent and illustrate the track sailed by the ship. This is done with high speed and accuracy compared to older and traditional methods. The speed and accuracy of the system assist the navigator in efficient and successful search for lost chain on the seabed because it makes it possible for the navigator to monitor where the search already has been done and where there are areas not yet searched. It is essential that the system both record and present data. It thereby supports persistence: Data about the track is kept on the display (until deleted by the crew). By making the in the environment invisible track visible on the display it contributes to the captain's situation awareness on level 1.

Example 2: M/V Mercandia VIII

The ferry Mercandia VIII sails between Helsingør in Denmark and Helsingborg in Sweden. Three ferry companies and a total of 7 ferries are operating on this route. For the ease of the passengers departure times from the two ports are often set to times like XX:00 or XX:30 making it easier for the passengers to remember. This means that up to three ferries are leaving the port at the same time. The departure is coordinated in that way that the first ferry announcing departure on a special radio channel called the K-channel has the right to leave port first etc. This is done to avoid congestion at the harbour entrance. Having a short crossing (about 15-18 minutes) crews on board the different ferries tries to optimize and save time. This means that they do not waste any time at departure but manoeuvre in a way that they follow right behind the first ferry out. The problem here is that there is a time delay between when the intention of leaving the port is announced on the K-channel and when the action is performed. It takes some time for the ferry to handle mooring equipment and leave the quay, and it takes some time from when it has left the quay and until it has gained speed and is sailing out of the basin. This problem is illustrated

by a situation observed onboard the ferry Mercandia VIII. Just before Mercandia VIII was leaving the port of Helsingborg, and before Mercandia VIII had announced departure on the K-channel, the other ferry Siletta announced departure. According to the rules agreed between the three ferry companies this means that Siletta had the right to sail out of port first and that Mercandia VIII should follow. The crew on board Mercandia VIII had interest in following Siletta close to optimize crossing time. They also had the experience that it usually takes a few minutes from the announcement that Siletta is leaving port until it is actually sailing out of the harbour entrance. The visual sight between the berth of Mercandia VIII and Siletta is obstructed by large buildings. It is therefore not possible for the crew on Mercandia VIII to see visually when Siletta has started moving. Since the buildings are also creating a so-called radar shadow it is also impossible to get this information from the radar. Having only observations based on visual line of sight and radar, the crew needs to wait for Siletta to show up behind the buildings before they start their own departure procedure. Even though Siletta has announced departure they can not be sure exactly how long the departure procedure takes for Siletta, and they are not able to predict for certain when Siletta is moving to the harbour entrance and thereby when it is safe for Mercandia VIII to start. The relatively (compared to for example radar) new technology AIS is used to identify ships and show specific information about them such as course and speed. The advantages of the system include the ability to “see” ships without being in direct line of sight for example near a river bend or in ports where buildings can obstruct the direct line view necessary for visual observation or observation by radar. In the case of Mercandia VIII and Siletta the AIS system can be used by the crew on Mercandia VIII to see when Siletta starts moving, and where in the harbour basin Siletta is in order to optimize their own departure. Without AIS the crew should wait for visual observation (or observation by radar) before they could start the departure procedure. With AIS they can start the departure procedure when they by use of the AIS observe that Siletta is moving and on the way to the harbour entrance. In that way they can potentially save a few minutes because they do not have to wait. A few minutes do not sound as a big deal, but in a crossing of about 15-18 minutes every minute counts. The example is not unique. We also observed that when leaving Helsingør Harbour crossing south bound traffic could be hidden visually and from radar behind the buildings of the castle Kronborg. Also here the crew made use of AIS to find ships hidden behind the castle. These ships should according to the rules give way to the ferry, but because of the geography of the area (narrow strait) the crew on the ferry prepared the departure and tried to plan it in a way that not would bring the crossing traffic in a difficult situation at that spot. If ships were hidden behind the castle there was a pretty good chance that they would be right outside the harbour exactly at the time when the ferry was sailing out of the harbour entrance. By knowing this in advance the crew would have the option to adjust the exact time of the departure, adjust the first course when leaving the harbour or adjust the speed.

Analysis of the Example

Since there are two ships involved in this example it makes sense to talk about two cognitive systems: One system for each ship. The components in the cognitive systems are: (1) People: The navigators on board the ship (Mercandia VIII and Siletta), (2) Artefacts: AIS, and (3) Environment: The port of Helsingborg, the buildings obstructing the view. The AIS makes information available that was not available before when only visual observation and radar could be used. The information from AIS about the present situation and status of

other ships support situation awareness on level 1 because it makes something in the environment perceivable: It makes it possible for the navigator to see the movements of Siletta even before the movements are visible on radar or visually in the environment. Thereby it helps the navigator in the prediction and anticipation (situation awareness level 3) of the future manoeuvres of Siletta. The navigator can for example predict exactly when Siletta is leaving the harbour or exactly when Siletta will show up behind the buildings and act accordingly. This new information provided by the AIS helps the crew in optimization of the departure. Without AIS they had to wait until visual observation or observation by radar was possible, and this would make the performance less optimal. AIS information about the present status and actual manoeuvres of Siletta is used by the navigator on Mercandia VIII to plan the departure and time it according to the departure of Siletta. The example illustrates how the crew in a particular situation can benefit from the use of AIS technology.

Summary of Findings in Example 2

The AIS present data about other vessels movement even if the other ship is not in line of sight. By making the – in the environment and on radar invisible – movements of other ships visible on the display, it contributes to the navigator’s situation awareness on level 1. The new information provided by the AIS helps in this particular example the crew in optimization of the departure. Having the new information the crew can do something they could not do without this information. The conclusion we can draw from that is that the crew can benefit from having more information available when it is relevant in the work task. In this case they used the additional information provided by the AIS in their optimization of the departure. The AIS function they use in this particular situation is that AIS is able to transmit information about position, speed and course. This information is available locally delivered by the GPS on one ship (cognitive system), and it is transmitted to other ships (cognitive systems) in the area by the AIS system.

Example 3: Mercandia IV

The ferry Mercandia IV is the sister ferry to Mercandia VIII and it is operating on the same route. The route they sail is busy with up to 7 ferries running at the same time. The route also runs across The Sound which has intense north and south bound traffic. The Sound is the strait between Denmark and Sweden and thereby one of the entrances to The Baltic. When crossing The Sound the ferries have to pay attention to the crossing traffic and navigate according to the rules in order to prevent collision. In praxis this means, that they when they are eastbound has to give way to traffic from the south and when they are westbound they have to give way to traffic from the north. This is the general rule, but from time to time they agree on VHF-radio about deviations from this rule. The ships – especially those with local pilots – sailing north or south in The Sound are aware about the intense ferry traffic. It is therefore common practice that they announce their intentions and choice of route. For example southbound ships often announce in advance which one of two possible routes they take: East or west of “Disken”. By the announcement of that they help the ferries in their route planning. When the navigator on the ferry knows if a ship takes the eastern or western route he or she can prepare and take the necessary precautions and by the adjustment of course and/or speed ensure that collision is prevented. The crews onboard both the Mercandia IV and VIII reported in informal interviews that they actually used the information from crossing ships about intentions and choice of route in their planning. The information from the north/south

bound ships is given verbally on the VHF-radio. It is often given well in advance. This means that there could be up to several minutes between the message on the VHF and the manoeuvre described in the message. Messages given on VHF are not persistent. This means that they disappear immediately [6, 7]. There are some ways in which the information can be stored or recorded (and preserved). It can for example be recorded using pen and paper (cognitive artefact). Another option is to use the replay function in the Sailor RT-5022 VHF. This function makes it possible to play a message received on the VHF again if it was not heard the first time or simply to repeat the content of the message if it has been forgotten. None of the two ferries in the study had this type of VHF. Depending on availability of technology the information can be stored using different elements in the cognitive system. It can for example be stored in the navigator's memory. Storing information in the navigator's memory is not necessarily safe and efficient – especially not if there is a time delay between the broadcast of the information and the time at which it is to be used and especially not in an environment where disturbances and disruptions could affect the memory of the navigator [4, 9, 10, 13]. The use of cognitive artefacts such as pen and paper or technology such as the replay function in the RT-5022 VHF radio is more reliable. During the observations it happened several times that a north or south bound ship announced intentions or choice of route on the VHF, but particularly one example is especially illustrative. The ship M/V Valtellina was on a southbound route through The Sound. Mercandia IV was westbound and therefore they had to give way to Valtellina according to the international rules. Well in advance the choice of route east or west of Disken was announced from Valtellina. Two navigators were on the bridge of Mercandia IV. One of them said to the other shortly after the announcement on the radio: "Didn't he say west [of Disken]?". He thereafter expressed that it could be necessary to call Valtellina on VHF to have it confirmed, but that was never put into action. A few minutes later Valtellina passed east of Disken. The situation was not even close to be critical, but it illustrates quite well that storage of information about other ships intentions and choice of routes in the cognitive system could fail. If the RT-5022 VHF had been available it could in that particular situation have been of some help to the crew because they easily could access the message from Valtellina and hear it again. The example also illustrates the problem about "saying and doing". What if Valtellina actually had announced that they would sail west of Disken by mistake? Or what if they changed their mind in last minute and forgot to inform other ships in the area about that? The example where the message is misunderstood and east and west is confused is not unique. We have in our on board observations also seen examples where ship names are misunderstood or confused. For example, if a group of ships is south bound in The Sound some of them could announce passage east of Disken while others announce passage west of Disken and others again make no announcement at all. In that case the risk would be that some ships are confused with each other.

Analysis of the Example

Since there are two ships involved in this example it makes sense to talk about two cognitive systems: One system for each ship. The components in the cognitive systems are: (1) People: The navigators on board the ship (Mercandia IV and Valtellina), (2) Artefacts: VHF radio (and also AIS and radar, but these are not in focus in the example), and (3) Environment: The water between Helsingør and Helsingborg and other ships in the area. The captain on Valtellina sends out a message about his intentions regarding route east or west of

Disken. This message is received on Mercandia IV and other ships in the area. Even when received the message is not necessarily perceived by the crew on Mercandia IV. In this example the message was misunderstood, and the only possibility they had – given the available equipment on the ship – if they would correct it, was to call Valtellina again and ask. The problem is that a message – a piece of information – normally disappears from the cognitive system right after it has been broadcasted on VHF unless it has been recorded in the memory of people in the system and perhaps transferred to a cognitive artefact such as a piece of paper. The replay function in the Sailor RT-5022 VHF automatically records the message, and the information is thereby persistent for a while. When the information is persistent, the crew can have it replayed and thereby transferred from the VHF to their memory if the transfer was obstructed (for example misunderstood or not heard) in the first place. This process is much easier than calling the other ship again. In this particular example the crew could have made use of the replay function and have the misunderstanding solved without calling and involving the other ship. They could then have used the information about the route of Valtellina in their optimization of the route if necessary. Another problem – apart from the misunderstanding we observed in this example – could be that the information is transferred well in advance; sometimes even many minutes in advance. The longer the time is between the transfer of the information and the time when it is to be used, the greater risk is that the information is forgotten – that it disappear from the memory of the navigator and thereby from the cognitive system on the ship. To counteract for that information can be recorded in other places in the cognitive system for example on a piece of paper. This is often more reliable and less vulnerable to disturbances and disruptions than human memory. The replay function of the RT-5022 provides automatic recording of information in the cognitive system. Thereby the information becomes persistent for a while. RT-5022 provide information persistence, and the information persistence supports the work task especially when it comes to the anticipation of the manoeuvres made by other ships – so-called level 3 situation awareness.

Summary of Findings in Example 3

Information about future actions is transferred between ships (cognitive systems). The verbal information is – unless it is recorded manually or automatically – not persistent in itself. The VHF replay adds information persistence to the cognitive system. Using the replay function it becomes possible to retrieve information about what has been said as verbal messages by other ships in the past about their future intentions and choice of route. This information can be used by the navigator to plan own route and optimize according to routes and choices made by the other ships.

CONCLUSIONS

Our observations and examples illustrate the effect of the introduction of new technology in the cognitive systems constituted by the crews, the technology and artefacts they use and the environment. Our first question was Q1: How can technology help the navigator to a better performance? We have seen how performance in a search and salvage operation is improved by the introduction of an electronic chart display. We have seen how performance in the departure of a ferry is improved by the introduction of AIS. And we have described how performance in route planning potentially could be improved by the introduction of the VHF replay function. Information persistence and information availability is essential

in our examples. Both the electronic chart system and the VHF replay function add persistence to the cognitive system when they act as storage for information. And the AIS system makes new information available in the cognitive system; information which is not available through other information sources. It is interesting here that both electronic chart, AIS and VHF replay improve the cognitive system not by adding new information from new kinds of sensors but simply by transferring, processing and/or displaying information already available in the cognitive system provided by other technologies and types of equipment. The electronic chart is based on information from the GPS, AIS information is based on a technology where information from GPS on other ships is received through radio link, and VHF replay is a simple recorder of transmitted verbal messages. Improvement of performance and support to the work task of the navigator could as shown through our examples simply be a matter of how data is processed and presented, how information is recorded and displayed. Our second question was Q2: How can new maritime technology be designed, in order to support the work tasks of the navigator even better? The conclusion we draw in that respect is, that our examples show the importance of information persistence and information availability. Both persistence and availability could be important parameters in the design of new supportive technology. Therefore the designer could benefit from identifying need for persistence and need for availability. Some information which could be a great help to the navigator could in fact already be present as data in the cognitive system, but because it has not yet been processed, transferred, recorded or displayed in the right way it is not persistent enough or not available when needed for the navigator to use it and take advantage of it. Finding and identifying these potentials is not a desk top exercise. It can only be done through comprehensive on board studies. Our studies show some examples on how this process can be approached, and the paradigm of distributed cognition could appear to be as valuable for the designer as it have been in our analysis.

ACKNOWLEDGMENTS

Our thanks to the crews onboard the ships we visited for letting us observe their work and environment and for answering our questions.

REFERENCES

- [1] Dekker, S. W. A. 2000. Crew Situation Awareness in High-Tech Settings: Tactics for Research into an Ill-Defined Phenomenon. *Transportation Human Factors*, 2(1), 49–62.
- [2] Endsley, M. R. 2002. Theoretical Underpinnings of Situation Awareness: A Critical Review. In: Endsley, M. R. and Garland, D. J. (Eds.). *Situation Awareness Analysis and Measurement*. Mahwah, New Jersey: Lawrence Erlbaum Associates. Pp. 3–32.
- [3] Flin, R. 2005. Safe in Their Hands? Licensing and Competence Assurance for Safety-Critical Roles in High Risk Industries. Industrial Psychology Research Centre University of Aberdeen. http://www.abdn.ac.uk/iprc/papers%20reports/Safe_In_Their_Hands_Revalidation_Report.pdf.
- [4] Gillie, T. and Broadbent, D. 1989. What makes interruptions disruptive? A study of length, similarity and complexity. *Psychological Research*, 50, 243–250.
- [5] Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, Massachusetts: MIT Press.
- [6] Kjeldskov, J. and Stage, J. 2003. The Process of Developing a Mobile Device for Communication in a Safety-Critical Domain. In: 9th IFIP TC13 International Conference on Human Computer Interaction, *Interact 2003*.
- [7] Kjeldskov, J. and Stage, J. 2006. Exploring “Canned Communication” for Coordinating Distributed Mobile Work Activities. *Interacting with Computers*, 18(6), 1310–1335.
- [8] Lützhöft, M. 2004. The technology is great when it works. *Maritime Technology and Human Integration on the Ship’s Bridge*. Linköping Studies in Science and Technology. Dissertation No. 907. Linköping, Sweden: University of Linköping.
- [9] Miyata, Y. and Norman, D. A. 1986. Psychological issues in support of multiple activities. In: Norman, D.A. and Draper, S.W. (Eds.). *User-Centered System Design: New Perspectives on Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- [10] Monk, C. A. 2004. The effects of frequent versus infrequent interruptions on primary task resumption. In: *Human Factors and Ergonomics Society 48th Annual Meeting*.
- [11] Sarter, N. B. and Woods, D. D. 1991. Situation Awareness: A Critical but Ill-Defined Phenomenon. *International Journal of Aviation Psychology*, 1(1), 45–57.
- [12] Sharp, H., Rogers, Y. and Preece, J. 2007. *Interaction Design: Beyond Human-Computer Interaction*. 2nd ed. Chichester: John Wiley & Sons Ltd.
- [13] Trafton, J. G., Altmann, E. M., Brock, D. P and Mintz, F. E. 2003. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International J. of Human Computer Science*, 58(5), 583–603.
- [14] U.S. Coast Guard 2006. Captain of The Port Long Island Sound Waterways suitability report for the proposed Broadwater Liquefied Natural Gas Facility. <http://www.uscg.mil/d1/units/seclis/broadwater/wsrpt/WSR%20Master%20Final.pdf>.

Alarm Handling in Rail Electrical Control

Nastaran Dadashi
Human Factors Research Group
University of Nottingham
epxnd2@nottingham.ac.uk

Sarah Sharples
Human Factors Research Group
University of Nottingham

John R. Wilson
Human Factors Research Group
University of Nottingham
Network Rail

ABSTRACT

Recent technological advances have led to more complex control rooms and several ergonomics research studies have focused on understanding this complexity and the consequent roles of the operators [4,6,8]. Many of the key aspects of behaviour in cognitive systems are not easy to assess, these include reasoning, problem solving, prioritising, etc. This study is the first in a programme to examine the human factors of remote condition monitoring (RCM), and particularly the knowledge requirements to support successful strategies for operators to diagnose, prioritise and initiate action. Alarm handling in an Electrical Control Room is the focus of this study. In this paper the data collection, analysis, and interpretation are reported only as they inform and provide insight into the work of the ECRO and their handling of alarms and the consequences of this. The aim of this paper is to identify the artefacts associated with alarm handling and to conduct an exploratory contextual investigation of alarm handling.

Keywords

cognitive system engineering, rail, alarm handling, electrical control rooms

ACM Classification Keywords

H.5.1.2 [User/Machine Systems].

INTRODUCTION

Recent technological advances have led to more complex control rooms and several ergonomics research studies have focused on understanding this complexity and the consequent roles of the operators [4,6,8]. However the diversity of activities in control rooms can impose a severe challenge for understanding the work of the operators and designing effective systems within complex socio-technical environments. One route to such understanding is to treat and study the work system as a joint cognitive system (JCS). In such studies it is essential to understand operators' (in rail, controllers and signallers) behaviours while interacting with complex supervisory systems in a real setting. Similar work has been done in domains such as nuclear power plants [10] and conventional

power plants [12]. Although in transportation control systems there has been work to analyse safety and human error [1], the application of cognitive system analysis for predictive display design has been neglected. Electrical control rooms in railway are no exception.

Alarm handling is one of the main responsibilities of electrical control room operators. Poor alarm handling may lead to critical accidents such as the Milford Haven explosion in 1994 [13] or the Channel Tunnel fire [3]. This study is a first field investigation of rail Electrical Control Room Operators (ECROs), specifically their handling of alarms. Because relevant behaviours are not easy to observe or for an investigator to explain, care was taken over a methodological approach which drew from features which have been defined previously as of significance for a JCS.

This study is the first in a programme to examine the human factors of remote condition monitoring (RCM), and particularly the knowledge requirements to support successful strategies for operators to diagnose, prioritise and initiate action. There are three reasons for focusing on alarm handling in this study: 1- it is a particular example of RCM in place at the moment (as opposed to the potential future system which are being proposed); 2- it is a good medium through which to investigate the skills, knowledge, expertise and strategies of this group (ECROs); and 3- it is an important and under-studied topic in its own right.

One additional element to the work carried out was to examine different cognitive models as a framework for all our investigations into the performance and information support of people interacting with RCM systems. In human factors various forms of model have been proposed for understanding the cognitive activity in control tasks: the classic Stimulus, Organism and Response model (SOR), Rasmussen's decision ladder [11], Knowledge Acquisition models [12], Joint Cognitive Systems [6], and Bainbridge's human information processing model [2] are a few examples of such models and theories. According to Bainbridge, models that represent human information processing can be categorised into sequential and contextual models. "*Contextual models focus on the overview, the temporary structure of inference built up in working storage to describe the task situation and how this provides the context for later processing and for the effective organisation of behaviour*" ([2], p. 352)."

Sequential models on the other hand model the cognitive processing as a sequence of processing stages. Rasmussen's decision ladder is one such model. Every thought is directly followed by an action through context free strategies. Decision ladder and SOR models are considered to be sequential models whereas Knowledge Acquisition models are contextual models.

The full study has involved assessing different analysis routines to work within Bainbridge's [2] alternative models, i.e., sequential and contextual, and this explains some of the data collection and particularly analysis approaches taken.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

However for the purpose of this paper the data collection, analysis, and interpretation are reported only as they inform and provide insight into the work of the ECRO and their handling of alarms and the consequences of this. The focus of this paper is to identify the artefacts associated with alarm handling and to conduct an exploratory contextual investigation of alarm handling. Field study generally, observational checklists and video recordings were applied as the main techniques to collect the data.

ELECTRICAL CONTROL ROOMS

Rail Electrical Control Rooms (ECRs) in the UK were originally integrated from a number of adjacent Railway Traction Power Supply Systems. Since 1932 Electrical Control Room Operators (ECROs) are responsible for remotely opening and closing equipment, instructing staff on the operation of manual switches and leading the maintenance and fault finding on electrification distribution and DC traction equipment.

Lewisham ECR (Figure 1) has three workstations with similar information available to all three. Two ECROs are active at one time and the third workstation is used for emergency situations. Apart from dynamic information displays on their desks there is a static board covering one wall of the ECR. This board shows the links and platforms of the area under control. Although the board is out dated, some less experienced operators use this to familiarise themselves with the area.

According to Network Rail’s “Specification for remote control equipment for electrical distribution systems” handbook [9], control displays in ECR are human machine interfaces which display either a mimic diagram operated through physical keys and push buttons or full graphic visual display units operated through keyboard and mice. Figure 2 shows the information layout on the operational displays in ECR.

The screen layout of the high resolution graphic VDU displays is: Alarm banner, Menu bar/area, System information bar (including date, time, operator log in identity), Command/error message bar, Picture display area, System overview and Detailed operating /outstation page or general/system-wide pages (e.g. index pages, lists, event logs, trend displays, data communication network and status page, Current alarm log or AC system overview)

There are three displays on each of the workstations and they are normally arranged as below (Figure 3):

Left hand screen→ supply system overview

Central screen→ Operational display (Figure 4)

Right hand screen→ alarm information or AC overview (where supply system is DC)

The operational display as its name suggests is to apply procedures. The other two displays are used to identify and interpret the problem.

Alarms are important features of control systems. They provide the stimuli to attract operators’ attention towards abnormal situations. One of the features of alarm presentation on ECR displays are their priorities. ECR alarms have priorities between one and six. System failures are priority six and the rest of alarm priorities are engineer configurable. Alarm priority is shown by its colour.

Priority 1→ extra high→ not applicable in ECRs yet, Priority 2→ High→ Red, Priority3→ Medium→ Pink, Priority 4→ Low→ Brown, Priority 5→ Extra Low→ for future use, Priority6→ System alarms→ Light grey, Clearing alarms→ Green

Any unacknowledged alarm appears on the alarm banner, the panel can contain up to seven alarms and if there is more than that at one time an arrow will be displayed at the right hand side in the colour of the highest priority alarm not displayed. Number of alarms varies in different hours and not all of them require operator’s intervention. Figure 5 shows the number of alarms generated in Lewisham ECR in one week from 29/01/2009 to 05/02/2009 which totalled to 1884 alarms. Alarms and operator actions are shown in different hours of the week; 168 hours in one week. Normal alarm/operator actions are around 3 alarms/operator actions in one hour, whereas the actual average of the alarms/operator actions in the observed week are around 10 per hour.



Figure 1. Lewisham ECR.

Mode		Time and date
Alarm Banner		
Display area		Menu
Page control buttons		
Text entry		

Figure 2. operations screen configuration.



Figure 3. ECR workstation.

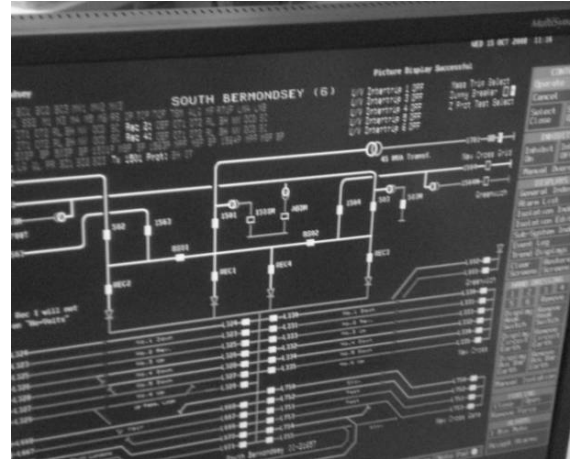


Figure 4. Operational display.

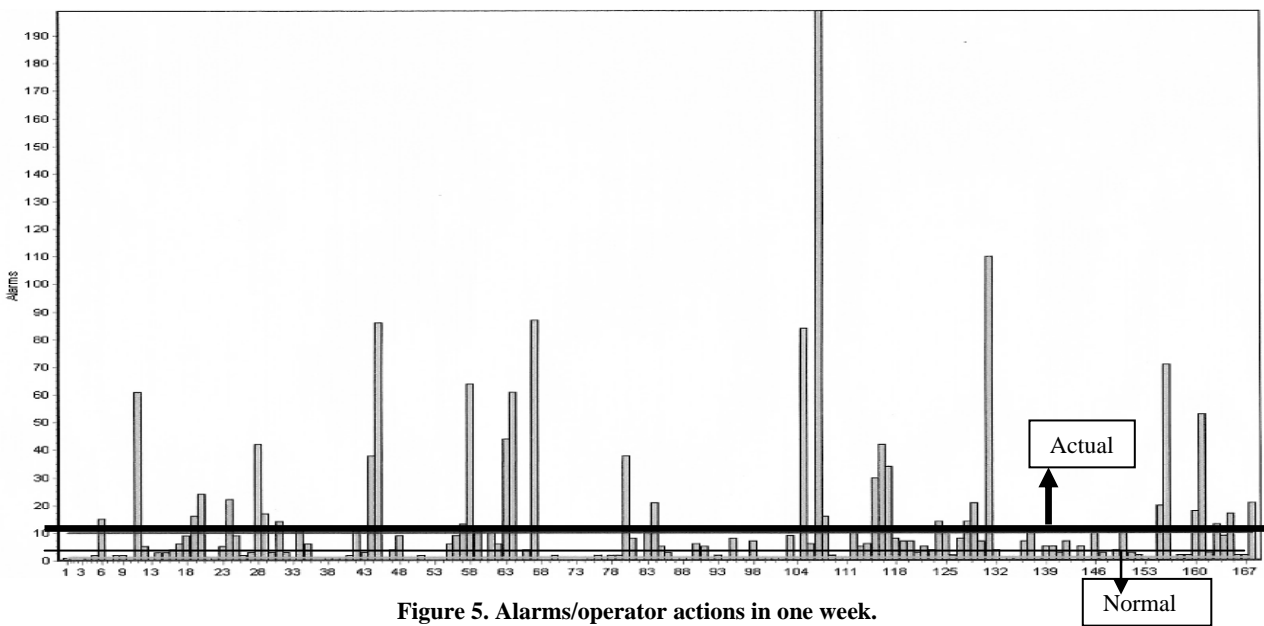


Figure 5. Alarms/operator actions in one week.

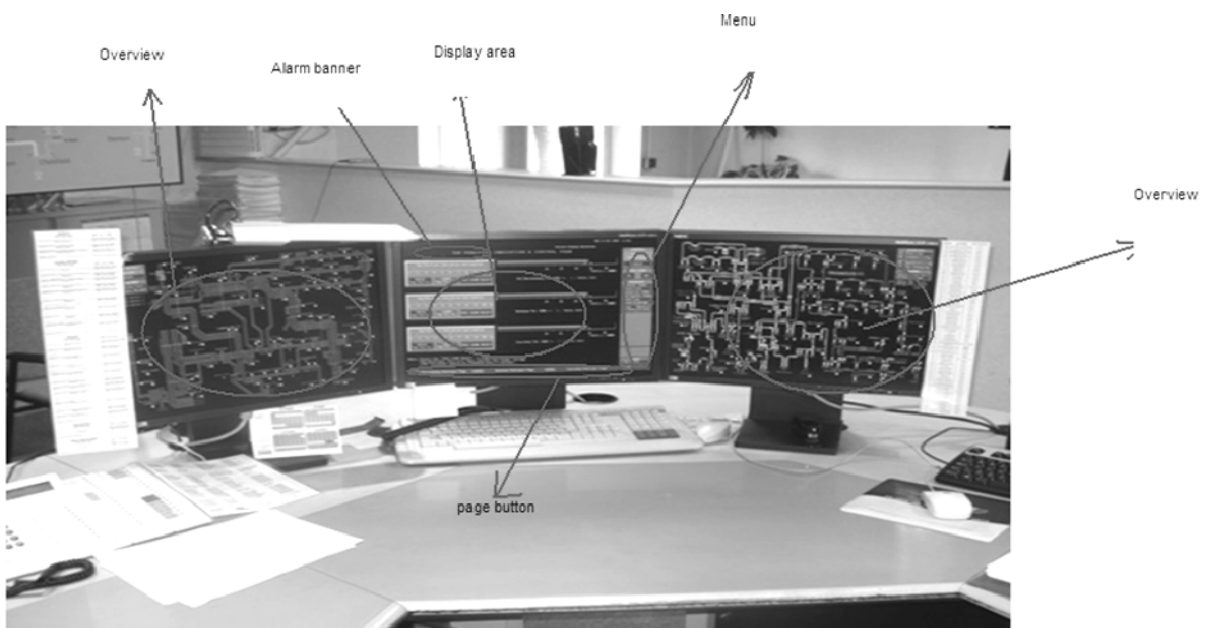


Figure 6. Dynamic and static artefacts.

METHODS

Field Study

Many of the key aspects of behaviour in cognitive systems are not easy to assess. These include reasoning, problem solving, prioritising etc. There are also more general aspects to understand within human factors study, such as expertise and mutual support. Therefore this first field study was exploratory and without pre-conceptions.

In order to support the completeness and validity of acquired data it was necessary to apply more than one qualitative method. Field observations (including operator commentary), video ethnography [5] and semi structured interviews were applied for the purpose of this study. To support observation techniques Lipshitz [7] suggests checklists of behaviours which require fewer judgments on the part of the observer and also recommends focusing on a small portion of the problem at a time. In this study a spreadsheet was developed to provide a checklist to log by eye observations of operator alarm handling.

Semi-structured interviews were conducted to familiarise the researcher with the ECR and its associated tasks. Operators were asked to explain the tasks they perform throughout a shift and their main responsibilities as well as the artefacts they use to conduct those tasks. These interviews also enabled the operators to get to know the researcher, helping build the relationship between the two and easing access for field observation to become less obstructive.

Lewisham electronic control room near London was observed for the purpose of this study. A pilot study was conducted to brief the ECROs and for the researcher to familiarise herself with the control room. This pilot study was followed by four observational sessions of 4.5 hours each at different times of the day to reduce the temporal bias. Two of the sessions occurred during the day (one early morning and one mid day) and the other two sessions occurred on the night shifts. It was believed that the number of alarms increases during night time due to maintenance operations on track. A total of 18 hours of observation enabled the study to collect introductory information regarding the ECRO tasks and to identify the artefacts they use.

In observation studies there is inevitable interpretation bias on the part of the investigator. To reduce bias, video recordings were taken and operators' comments and behaviours regarding the task were used as the basis for interpretation. Furthermore operators were asked to describe some parts of the problem solving (talk aloud) when they had time.

A Sony camcorder was used to record alarm handling. Alarm handling was recorded from the moment the alarm was generated until it was cleared. The aim was to capture detailed information such as the artefacts used, the screen the operator is attending to, time spent thinking without clicking on any of the buttons, etc. Furthermore the operator's comments regarding the alarm during and after handling the alarm were recorded.

Operators were asked to explain the procedure they followed afterwards and give comments regarding the complexities they were facing while handling the alarm. These comments were then used for categorising alarms into two groups: those where any complexity in handling the alarm comes from high levels of information (i.e. alerted in several different ways) and those where any complexity may come from an absence of

diagnostic or other information (high information and low information respectively). Note that these levels of complexity did not eventually cause any major issues since all of the alarms were cleared in time.

During the 18 hours of observation 22 alarms occurred, half of them were due to the maintenance work and the operator was informed beforehand, these are referred to as 'expected' alarms. Since no diagnostic procedures are required for these alarms they are used as base cases for analysis. The rest of the alarms occurred unexpectedly; referred to as 'unexpected'.

In this paper the first phase of the data analysis is presented, addressing three main questions: 1) What is the difference between expected and unexpected alarms in terms of use of artefacts? 2) What is the relationship between system complexity and operators control? 3) What is the association between applications of different types of artefacts?

The non-directional hypotheses and variables are:

H1: There is a significant difference between the use of artefacts unexpected and expected alarms.

Independent variable: type of alarm → expected and unexpected.

Dependent variables: Number and type of artefacts used → dynamic (telephone, face to face communication, alarm banner, menu, display area, page button, overview displays); and static (board and paper).

H2: There is a significant difference in use of artefacts between high information and low information.

Independent variables: Type of complexity → high information and low information.

Dependent variables: Number and type of artefacts → dynamic (telephone, face to face communication, alarm banner, menu, display area, page button, overview displays), static (board and paper).

H3: There is a significant difference in operators' control over the alarms between various complexity types.

Independent variables: type of complexity → high information and low information.

Dependent variables: control → combination of three factors: time, knowledge and resources (see below).

Observation Checklist

A Microsoft™ Excel spreadsheet was designed to structure the results obtained from field studies and to address the objectives of this study. Table 1 shows the columns of the spreadsheet. A JCS approach views the cognitive system as a result of human-computer co-agency rather than human computer interaction. Hence what matters is the performance of the whole unit rather than the individual. In other words a JCS is not defined by what it is but it focuses on what it does. Features of significance in the JCS approach are said to be the use of artefacts, control and complexity [6], although here the definition of control is modified to fit the context of the study.

Artefacts are assessed by the number of times the operator attends to them during alarm handling. The other two JCS features, control and complexity, are assessed through the interpretation of various data.

Table 1. Observation checklist.

JCS feature	Artefacts		Control			Complexity	
Categories	dynamic	Static	Time	Resources	Knowledge	High information	Low information
Assessment	Quantitative assessment: the number of use for each of the artefacts		Qualitative assessment: scaled by low, medium and high and assessed through the field study.			Qualitative assessment: Filed study and direct questioning	

Use of Artefacts

“An artefact is a device used to carry out or facilitate a specific function that is part of work ([6], p. 66)”. Understanding the artefacts used by operators will provide the designer with a guideline of how these artefacts should be presented in the system. In an ECR, artefacts can be classified as dynamic or static. Table 2 presents static and dynamic artefacts in ECR (shown in Figure 6).

Table 2. Dynamic and static artefacts.

Dynamic							Static		
Communication		Operational display				Overview			
Phone	Face to face	Display Area	Menu	Page button	Alarm banner	supply system overview	AC overview	Paper	Board

Every 15 seconds artefacts that were used were counted and the total use of that artefact for each of the alarms is the total number of times they have been counted for each alarm.

Control

According to Hollnagel and Woods, [6], a human operator is in control if four conditions are satisfied: time, resources, knowledge and competence. He/she should have sufficient time to diagnose the problem, enough experience and information to anticipate the future events and to recognise the present state of the system. Moreover the operator should be competent enough to choose clear procedures. As one of the operators in Lewisham ECR commented:

“You can really fall into it when you have too much information on your displays... I know the area and what is wrong on our displays, majority of our emergencies are main concern, and we have to decide there and then and you don’t get much time planning and you have to be 100% correct”.

In this study three of these factors were taken into account: time, knowledge and resources. Competence was left out since ECROs who participated in this study were all experienced and competent according to Network Rail’s training procedures. Time, resources and knowledge have been interpreted as low, medium or high based on the observations. Time is not defined as the alarm handling duration, but as the time that operator thinks he/she has to assess the alarm. It can vary based on the

alarm priority on the alarm banner and how busy the operator is while the alarm is generated.

Resources are defined as the information available to the operators to assess the alarm; this is not only the information presented through the display but the operator’s knowledge of the area under coverage as well as the information obtained through communication with track workers and other ECROs.

Knowledge refers to how well the operator knows the cause of the alarm (the why). Operators’ comments regarding alarms were used for assessing this factor. Tables 3, 4 and 5 respectively show examples of comments and illustrate reasons for the assessment of time, resources and knowledge levels.

Complexity

This is when the operator faces multiple channels of information which require continual allocation and division of their attention and this will consequently cause complexity [6]. The amount of information to specify one problem can be more than enough, for example if an operator gets alerted to an alarm through a siren as well as a phone call and flashing buttons on his display; this is ‘high information’. On the other hand some of the alarms have no indication at all, for example alarms known as ‘ghost alarms’ occur without specific reason and neither of the displays shows any indication of what caused it; this is ‘low information’. Operators were directly questioned to determine if they were dealing with a case of ‘high information’ or ‘low information’.

RESULTS

Expected and Unexpected Alarms

Independent sample t-test was used for the statistical analysis. Telephones and display area were found to be significantly different depending on the type of alarm. There was a significant difference between the number of times operators used the telephone in unexpected (Mean = 0.131, SD = 0.340) and expected alarms (Mean = 0.592, SD = 0.050); $t(86) = -5.044, P < 0.01$. Also there was a significant difference between the number of time operators attend the display area in unexpected (Mean = 0.524, STD = 0.503) and expected (Mean = 0.222, STD = 0.423) conditions; $t(86) = 2.721, p < 0.01$.

Figure 7 shows the mean frequencies of the artefacts used in unexpected and expected alarms. Use of face to face communication, menu, display, page button, overview and paper increased in unexpected alarms whereas use of phone and alarm banner was more in unexpected alarms.

Table 3. Sample of time level assessment.

Alarm no	Time Level	Actions and comments
1	Low	Operator was conducting his weekly testing when a high priority alarm was generated. Only two seconds after the first alarm another alarm was generated. There were two things that he needed to keep in mind, first to attend the alarm and second to remember how far he was in the testing, hence he wants to finish the alarm clearance as soon as possible and return to the testing before he forgets the last tested circuit.
2	Medium	A high priority alarm was generated. It took the operator 3 seconds to start handling and he seems to have sufficient time, however he had to go to another page on his operational display and the system took 4 seconds to load, he mention that this wait in some cases can increase up to 30–40 seconds and there is nothing that you can do about it.
3	High	Operator is monitoring the display when a high priority alarm occurs. Since he is free when the alarm was generated, he seems to be attending it without any rush.

Table 4. Sample of resource level assessment.

Alarm no	Resource Level	Actions and comments
1	Low	Operator cannot figure out the alarm because the display is lacking some specific information as mentioned by the operator: <i>“computer lost indication and it does not know whether it is open or close and that why it actually displaying a question mark on the breaker”</i> He reaches the point that he can only fix the alarm with trial and error since he could not enough information to deal with it properly.
2	Medium	The operator cannot find the reason from the display, he go to the event log to find more information and he is successful.
3	High	The operator can easily assess the alarm; he has all information required at hand.

Table 5. Sample of knowledge level assessment.

Alarm no	Knowledge Level	Actions and comments
1	Low	This is an alarm which he cannot figure out, the first procedure was not useful so he keep murmuring the codes and site numbers in a puzzled voice, he spent nearly 15 seconds staring at the screen and trying to figure out why it has failed the second time.
2	Medium	The operator is guessing: “for some reason, the computer lost indication of that circuit breaker there and generate an alarm what is probably caused it is that when this circuit breaker is open is probably sitting next to that one and the vibration moves something in the generator and that’s why it generated an alarm”. But his action cleared the alarm.
3	High	The operator completely knows the cause of the alarm and start applying procedures right away.

Complexity Types

Artefacts

In order to investigate the difference of application of artefacts between high information (Mean = 0.38, STD = 0.49) and low information (Mean = 0.84, STD = 0.37) an independent sample t-test was applied; the results revealed that the display area attendance are significantly higher in alarms due to high information; $t(59) = -3.63, p < 0.01$. Figure 8 shows the mean of application of artefacts in different complexity types.

Control

From the total of 11 unexpected alarms, 6 of them were high information and 5 of them were low information. Figure 9 shows the scale (low, medium, high) of the control factors in different complexity types. A Mann Whitney U test was conducted to investigate the difference between the two complexity types. No significant difference was found between the complexity types.

From a total of 915 seconds of observed alarm handling, 630 seconds of it were due to alarms with high information and 285 seconds of it were alarms with low information (Figure 10). From the total of 11 alarms, complexity types was spread fairly, 5 cases of low information and 6 cases of high information. This difference suggests that the alarm handling duration for cases with high information is more than cases with low information.

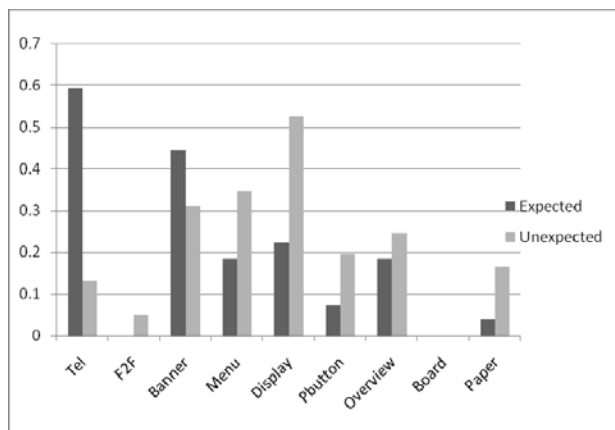


Figure 7. Artefacts in expected and unexpected alarms.

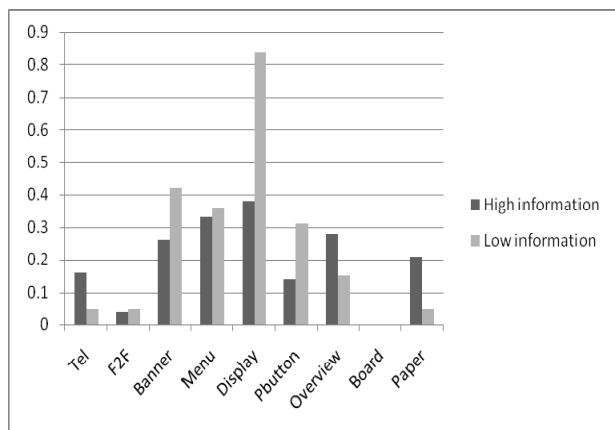


Figure 8. Artefacts in different complexity types.

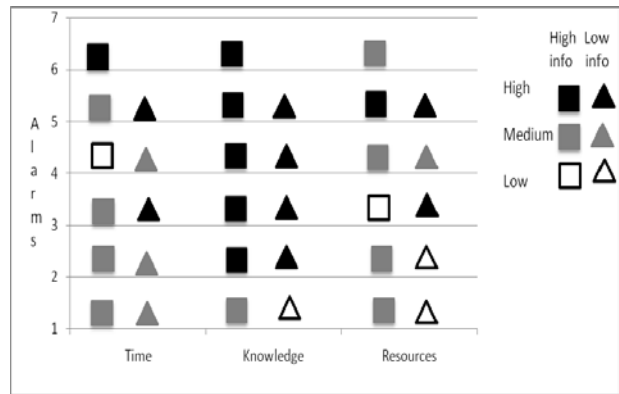


Figure 9. Control and complexity.

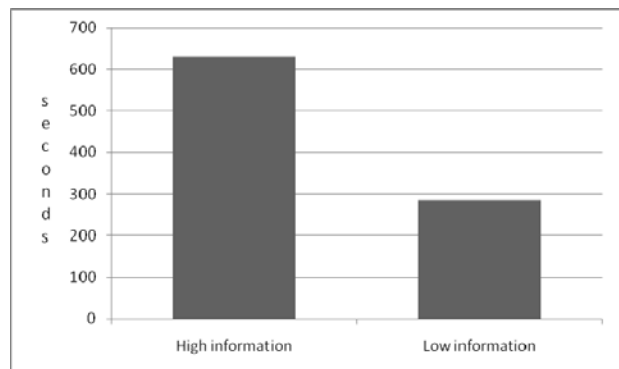


Figure 10. complexity and alarm duration.

DISCUSSION AND FUTURE WORK

An exploratory study was conducted in Lewisham ECR to investigate ECROs' alarm handling cognitive processing. This is the first of a set of investigations conducted in the rail ECR; the aim is to apply field based data to develop an information processing framework for understanding supervisory control systems in rail of particular relevance to remote condition monitoring.

Results regarding the difference between the expected and unexpected alarms are interesting. Operators used the phone more when they were attending to an expected alarm and attended to the display area more when they were dealing with an unexpected alarm. Also they used paper more while assessing unexpected alarms; this can suggest the limitation of the display which cannot provide alarm indications in one page. Hence operators are required to write indication codes from, for example, an event log and switch back to the display area to complete the alarm assessment.

ECROs' control seems to be similar in both high and low information types of alarm complexity. This might be due to the fact that all of the alarms during the study were cleared and operators' definition of high and low information was not extreme enough to be counted as complexity. Also the artefacts used in low information are mostly within the operational display which suggests that operators knew that the information required is somewhere in the system whereas they mostly used phones and papers while attending high information and looked over the overview display to organise the high amount of information and assess it. Laboratory-based studies and more rigid quantification of these factors are recommended to investigate this in more detail.

Furthermore use of artefacts also varies between the two cases of high information and low information. In high information situations, operators used the overview displays and paper to organise the required information, whereas in cases of low information they spent most of the time exploring the display area to find an indication for the alarm. Once more, since the cases with low information are less time consuming, it suggests that artefacts associated with these alarms are more straight forward and easy to follow.

High information seems to be inevitable in today's control rooms; however the findings of this paper strongly suggest that this should be avoided as much as possible. Information requirement assessment is one of the major steps in system design, however in complex control rooms such as ECR, it should be not only to identify the required information but also to identify the sufficiency level for information.

This paper is the first phase of a cognitive system analysis, the next level is to apply the context specific findings of this paper to identify operator's strategies and provide guidelines to support the operator's preferred strategies and their associated information types.

ACKNOWLEDGMENTS

This research was supported by Network Rail (UK). The authors wish to thank the crew in Lewisham ECR for the time they put in as participants.

REFERENCES

- [1] Amalberti, R. 2001. The paradoxes of almost totally safe transportaion systems. *Safety Sceince*, 37, 109–126.
- [2] Bainbridge, L. 1997. The chenage in concepts needed to account for human bahaviour in complex dynamic tasks. *IEEE Transaction on systems man and cybernetics – part A: systems and humans*, 27.
- [3] Brown, S. 1999. The role of alarm systems within safety-related systems-the links with IEC 61508 and IEC 61511. *IBC Seminar on Alarm Systems*. London: IBC UK Conferences Ltd.
- [4] Cordiner, L., Wilson, J. R. and Robberts, S. 1999. Evaluation of control room human factors: A case study in the Railway industry. *People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centers*, 21–23 June 1999.
- [5] Farrington-Darby, T. and Wilson, J. R. 2009. Understanding social interaction in complex works: a video ethnography. *Cogn Tech Work*, 11, 1–15.
- [6] Hollnagel, E. and Woods, D. D. 2005. *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. CRC Pr I Llc.
- [7] Lipshitz, R. 2004. There is more to seeing than meets the eyeball: the art and science of observation. In: Montgomery, H., Lipshitz, R. and Brehmer, B. (Eds.). *How professionals make decisions*. Lawrence Erlbaum Associates Inc.
- [8] Luff, P. and Heath, C. 2001. Naturalistic analysis of control room activities. In: Noyes, J. and Bransby, M. (Eds.). *People in Control: Human Factors in Control Room Design*. London: IEE Press. Pp. 151–167.
- [9] Network Rail. 2008. *Specification for remote control equipment for electrical distribution systems*. London: Network Rail.
- [10] Patrick, J., James, N. and Ahmed, A. 2006. Human processes of control:tracing the goals and strategies of control room teams 1295-14. *Ergonomics*, 49, 1395–1414.
- [11] Rasmussen, J. 1986. *Information processing and human-machine interaction*. United States: Elsevier Science Pub. Co. Inc., New York, NY.
- [12] Sundstrom, G. A. and Salvador, A. C. 1995. Integrating field work in system design: a methodology and two case studies. *IEEE Transaction on systems, man and cybernetics*, 25, 385–399.
- [13] UKHSE. 1997. *UK Health and Safety Executive (UKHSE): Explosion and fires at Texaco refinery, Milford Haven July 24, 1994*. HSE books.

Session 7: Developing Usability Evaluation

Familiarity with Mobile Phones

Phil Turner
Centre for Interaction Design,
School of Computing
Edinburgh Napier University, Edinburgh, UK
p.turner@napier.ac.uk

Emilia Sobolewska
Centre for Interaction Design,
School of Computing
Edinburgh Napier University, Edinburgh, UK
e.sobolewska@napier.ac.uk

ABSTRACT

Heidegger describes familiarity as the readiness or preparedness to cope with the world, the equipment it comprises and other people. However, and more usefully for the current discussion, Dreyfus extends this and argues that familiarity comprises “know-how” and involvement which is a proposition susceptible to empirical testing.

In our qualitative study of 21 regular mobile phone users we find evidence of “know-how” in the form of a range of cognitive heuristics which people use to make sense of how to use their phones and (when asked) to explain how it operates. We also found evidence of involvement which we understand to mean comportment, that is, an orientation towards the technology. To our surprise peoples’ comportment was largely confined to the phones’ ease of use rather than its aesthetics or brand. We were also able to distinguish between the phone as an *artefact* and the phone as a *means to an end* and some suggestion that this distinction may be age related.

We conclude that familiarity offers a coherent conceptual platform from which we might reason about how people use, conceive of, feel about and select interactive technologies.

Keywords

heidegger, familiarity, involvement, know-how

ACM Classification Keywords

H.5.1.2 [User/Machine Systems].

INTRODUCTION

A discussion of familiarity with mobiles phones, indeed any kind of interactive technology, promises to be prosaic. Such an expectation is a consequence of simply understanding familiarity to mean knowledge of, or acquaintance with, something. But familiarity is more complex than this. So with what are we familiar? We are familiar with the world, people and their practices; with technology and equipment; with the affordances the world provides; we are familiar with human-made worlds of plumbers, carpenters, academics and designers (cf. communities of practice).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

The perspective we have adopted to investigate familiarity draws directly on Heidegger’s work and that of one of his most respected commentators Hubert Dreyfus. However before we launch ourselves into the world of hermeneutic phenomenology we should take a moment to consider other views of familiarity. What is striking is that very few researchers have actually considered it directly. Aside from treating familiarity as a narrowly cognitive phenomenon (e.g. prior knowledge or ‘perceptual set’) which fails to capture the breadth of the concept, there are very few directly investigations of it (in truth we could not find any, but to say there are none is tempting fate). However there are a number of theoretic perspectives which seem close to Heidegger and Dreyfus’ descriptions. Perhaps Polyani’s [1] description of tacit knowledge does appear to be a close synonym for familiarity. Tacit knowledge or “know-how” involves learning and skills which cannot easily be quantified except through their articulation. Riding a bike, for example, is best learned through direct experience, by closely observing others, or being guided by an instructor. Familiarity also appears as *learned perception* or “know-how”. Wartofsky [2] has shown that perception can be thought of as an historical process (as opposed to a purely cognitive one) and one which is an expression of individual or group familiarity. He argues that perception is a functional aspect of the interactions between animals and their environments, observing like Gibson that there is a reciprocal relationship between the animal and its environment. The perceived world of the animal may be internalized as a ‘map’ while the animals senses are themselves are shaped by the purposive interactions which the species has with the environment. Extending this argument to human-made worlds, Wartofsky, writes, “Rather, the very forms of perceptual activity are now shaped to, and also help to shape an environment created by conscious human activity itself. This environment is the world made by praxis – nature transformed into artefact, and now embodying human intentions and needs in an objective way.” Examples of ‘nature transformed into artefact’ can be found in Bucciarelli [3] observations about the perceptions of different kinds of designers: “...consider this page in front of you. [...] A naïve empiricist would sense its weight and estimate its size; another reader might note its colour or texture; a chemist on the design team would describe its resistance to discoloration, its acidity, and its photosensitivity. A mechanical engineer would be concerned with its tear or its tensile strength, its stiffness, and its thermal conductivity. An electrical engineer would speak of its ability to conduct or hold a static charge. All of these attributes of the object, the same artefact, are understood within different frames of reference, and they might all contend in a design process...” (p. 71).

Further evidence of Wartofsky’s observations can be found in, for example, Goodwin and Goodwin’s [4] study of operational staff at an airport. Their study demonstrated how perceptions of flight information displays, and so forth and their perceived properties or characteristics are shaped by the histories of both the personnel involved and the artefacts themselves. Expert

chess player also show evidence of being able to see patterns of pieces which the novice cannot (e.g. [5]); and the tennis pro is able to predict the direction of an opponent's serve better than a novice [6]. Turner *et al.* [7] have also reported changes in perception in a longitudinal study of seniors learning to use a personal computer. Changes in their perceptions of the technology they were using were evident with their growing experience (familiarity) with it. For example, information technology was initially perceived as alien and having a language of its own but became a means of warding off or at least delaying many of the effects of aging (loss of mobility, contact with the outside world, communications with friends and relatives) with growing familiarity.

HEIDEGGER ON FAMILIARITY

Central to Heidegger's examination of the nature of being is the observation that a fundamental characteristic of human being (which he calls *Dasein*) is it being-in-the-world. *Dasein* is neither outside the world, nor is it next to it, "There is no such thing as the 'side-by-side-ness' of an entity called *Dasein* with another entity called 'world'" [10 (55)]¹². Further, *Dasein* does not enter the world but is always already in it. In defining being-in-the-world, Heidegger begins as follows [Being and Time, 54]:

"In stems from *innan-*, to live, *habitare*, to dwell. "An" means I am used to, familiar with, I take care of something. It has the meaning of *colo* in the sense of *habito* and *diligo*. We characterized this being to whom being-in belongs in this meaning as the being which I myself always am. The expression "bin" is connected with "bei". "Ich bin" (I am) means I dwell, I stay near... the world as something familiar in such and such a way. Being as the infinitive of "I am": that is, understood as an existential, means to dwell near... to be familiar with."

The way of being of human being is to inhabit a world, i.e. to dwell in practices and their equipment and *to be familiar with them*. Familiarity is the readiness to cope with people, technology and situations. For Heidegger, familiarity is not a body of knowledge or a capacity; to be familiar is to be constantly accommodating or adjusting to the situation or use of technology.

Heidegger describes this readiness as "the primary familiarity, which itself is not conscious or intended but is rather present in [an] un-prominent way" [11 (189)] and in his *Basic Problems of Phenomenology* [8] he calls it the "sight of practical circumspection [...], our practical everyday orientation" (163).

Instead of adopting un-interpreted Heidegger we have found Dreyfus' exposition of familiarity to be both tractable and testable. For Dreyfus [9] familiarity comprises understanding and involvement and it is a discussion of these two concepts to which we now turn.

Understanding as "Know-how"

Understanding, he suggests, should be interpreted as 'know-how'. Dreyfus notes that "This know-how ... is more basic than the distinction between thought and action" and describes human beings as "We are such skills", thus directly equating humans with our know-how. This know-how is not to be found in knowledge, i.e. it is not to be found in what we think but in what we do: "We always conduct our activities in an understanding

of being" [10 (5)]. Understanding of being is embedded in human activities and shows up with them, "we understand ourselves and our existence by way of the activities we pursue and the things we take care of" [8 (159)]. Our understanding of being is embedded in human activities because it is fundamental to our ability to do things, our skills. We use these skills to cope with the world which, according to Heidegger, has three key characteristics.

Firstly, it comprises the totality of inter-related pieces of equipment. Each piece of equipment being used for a specific task – hammers are for driving nails into wood; a word processor is used to compose text.

The second 'component' of the world is the set of purposes to which these tasks are put. Of course, while we cannot meaningfully separate out purposes from tasks in these (non-Cartesian) worlds we can recognize that the word processor is used to write an academic paper for the purpose of publication and dissemination.

Finally, in performing these tasks we acquire or assume an identity (or identities) as designers, academics and so forth.

Involvement as Comportment

Involvement is a fact of our being and is not susceptible to further analysis. It is perhaps best understood as comportment, that is, an *orientation towards*. It is fair to say that involvement with, in this instance, is significantly less well understood as compared with "know-how". While involvement and engagement tend to be used inter-changeably in the human-computer interaction literature, our reading of Heidegger suggests that involvement is primarily social in character. Thus we also believe that comportment comprises a repertoire of social activities and awareness. Comportment is then observed as various forms of engagement which may or may not be explicitly social but have, nonetheless, their roots in the social. This is the position we have adopted for the purposes of this investigation.

We also make a further distinction, namely, between voluntary and involuntary involvement. The presence of voluntary involvement can be inferred from explicitly people taking pleasure in, liking (aesthetic appreciation), finding easy to use, fun and playfulness. Involuntary involvement or *thrownness* to use Heidegger's terminology refers to our unwitting participation in a situation. When we find ourselves thrown into a situation we cannot, for example, choose not to understand our native language; in the context of a meeting most of us cannot let a clearly incorrect assertion go by without objecting to it; we cannot help jumping at scenes in scary movies. All of these examples illustrate the fact that we cannot help but be involved in certain situations. This is not to suggest that we cannot 'tune' out, direct our attention elsewhere and be voluntarily involved.

We now consider the empirical evidence for familiarity.

THE EMPIRICAL STUDY

Twenty-one people (11 men and 10 women) were recruited for this study. Their ages ranged between 20–58 years.

The participants were told in general terms the purpose of the study ("what people think about their phones") and gave their consent to the use of the interview data for research purposes and subsequent publication.

The participants were interviewed individually by the authors. An audio recording, using a Sony DAT recorder, was made of the interviews which in turn was processed using Audacity

¹² The figure in parentheses is the section number in that particular volume.

(www.audicity.com) to create a .wav file which in turn was converted into an MP3 file as this is easily manipulated using iTunes™. The MP3 file was then transcribed into text.

The interviews themselves varied in length with some lasting as long as 40 minutes. The interview procedure was designed to constrain the interviewee as little as possible and started with the question, “Your mobile phone – what’s the story?”

The participants were also prompted to discuss:

1. how they thought their phone worked;
2. how they selected it; and
3. what they used it for.

Subsequent prompts of the form “go on” were used as required. Our approach was based on Payne’s approach [12] to the qualitative investigation of the mental models used to understand the operation of automatic tellers.

Evidence of “Know-How”

Our definition of “know-how” was based on our earlier work [13] where we found that people surprised us by talking about their personal technology in a wide variety of different ways. In all we found evidence that peoples’ “know-how” exhibited as fragments of propositional knowledge, mental models and the occasional use of anthropomorphism, that is, treating interactive technology as though it was a ‘friend’. This last point is discussed in Section: “Anthropomorphism”.

Propositional Knowledge

We regard fragmentary items of factual information regarding the operation of the technology as being evidence of propositional knowledge, for example, in answer to the question “What is in a mobile phone?”, we find:

“Many things, circuit boards, chips, transceiver [laughs] battery ... a camera in some of them”

Participant 4¹³

Mental Models

Mental models [12, 13, 14] are proposed complex cognitive representations which mediate the use of interactive artefacts. These models are very diverse and include systems (e.g., our knowledge of banking), devices (e.g. the operation of pocket calculators), physical forces (e.g., the nature of electricity), and concepts (e.g. the administration of justice). HCI mental models also have had the dual role of being used to reason about how to create an interactive system and as a means to represent people’s understanding of a particular interactive device or system. Along with other researchers we infer their use by our participants from coherent narrative accounts of the operation of devices. Again in answer to the question “How do you think your mobile phone works?”, we find:

“Well eh [long pause] I would explain to them ... that the cell phone, well over here it is called the mobile phone but in the U.S. it is called the cell phone and I think the American version of the name really gives an indication to how it works. The ... eh eh from now on I will call it a cell phone, the cell phone allows to communicate with other individuals that are attached to the generic phone network but while you move. The cell phone works by locating you within a cell, a cell is a geographic location, which are mapped out by masts, which send out an electromagnetic

signal, and the signal is a wireless signal which allows you to communicate through the air without any sort of connection to the mast. The mast is then connected by wired or wireless system to the normal telephone network, the normal telephone network being the systems of phones that are connected via cables and non-wireless technologies or any intervening such technology. And now, what you do is, you tap in the number as in a normal phone, I should probably describe to them what a phone is anyway, I mean, the only difference between a wireless and a wired phone is the transport medium between the handset and the receiver. In a normal wired network, it obviously goes down a cable, and with a wireless phone the only difference is the cell idea. A telephone allows you to communicate with someone by picking up the phone and dialing a unique number, which identifies them in this region. Telephone numbers are split up into regional codes and then there is a hierarchy of regional codes, then there’s categorization by country, so for example if I wanted to phone someone in the U.S., first of all I would have to call there country code, then there area code and then there local code after that, and that’s all there is really to it”

Participant 6

Anthropomorphism

Reeves and Nass [17] were among the first to recognize that the way in which we treat interactive technology, television, and other new media is essentially social. In their *The Media Equation*, they show in a surprisingly wide variety of ways that the apparent blurring of real and mediated life is commonplace. They have presented evidence that we interact with media in the same way we respond to other people, using the same rules that govern face-to-face interpersonal interactions.

In these data, some of the interviewees used anthropomorphism as a causal explanation:

“I suppose ‘cos because the phone... The phone is always checking in with the network, so, it’s always sending something back and forward... and it has to keep it internal clocking things running... and I’m sure it has some other business to attend to, when it’s sitting there... if it’s getting styling

Describing the use of anthropomorphism as a cognitive heuristic seems legitimate enough but it may also be evidence of the overlap between “know-how” and involvement. This is an issue which requires further investigation. However, we now turn to the evidence for involvement.

Evidence of Involvement

Evidence of involvement follows into two categories namely the social and the *ludic*. Evidence of predominately social involvement with mobile phones we have recognized as the phone being treated primarily as a ‘means to an end’. This is not the Heideggerian description of the inter-relatedness of the equipment, which comprises the world (e.g. hammers are for driving nails into wood in order to hang a picture etc). Instead it is, as we have suggested, the social basis of engagement. This in turn which brings us to the evidence of engagement itself, for example, people reporting that they took pleasure in (their phones...), liked their phones, had fun with..., and found them easy to use and so forth.

Means to an End

What follows are a series of extracts under the broad heading ‘means to an end’ where the phone is not being described in

¹³ We use an ellipsis to indicate a pause in the interview.

turns of its remarkable functionality or quality of design but in a matter of fact manner.

“I got a pay as you go, it’s orange and I’m being with the same... I got the same number since I first bought the first mobile 12 years ago, or something like that, and I... I was... I’m very happy with my orange pay as you go, I’m very... I’m reluctant to change it (...) I’m quite happy with what I got. I think phones should be just the phones, you know.”

Participant 15

People demonstrated a remarkable disinterest in a phone’s additional features such as a camera:

“[Interviewer] Do you regret it doesn’t have a camera, or any other options?”

No I wouldn’t be interested. That’s the reason why I went for the cheaper one, with very less options. Just to phone, text and that’s it.”

Participant 14

“Ok. First of all, the simplest model as possible. No extra features. If it’s... nothing extra... funky tunes or anything. Just simple brick if possible. Because at least I won’t lose it I will feel it in my pocket I have phone... eh... like... the simplest thing. I have a Samsung – the cheapest ever, because it’s a normal screen, no camera, or anything. No mp3 player, because I have my own. Plus the sound from those kind of mobile phone mp3’s is not as good as others, so, with my sound background and knowledge about electronics I don’t believe everything in one device is a good thing because you lose your mobile phone you lost your mp3 player as well. If you lose your phone you have an mp3 player in your pocket anyway. So you see, there you go – advantages and these are advantages of it. Eh... simplest. No cameras, no extra features, no games, never use the pictures, never them. That’s it. I can show you mine – it’s really very basic... Yeah sure... the problem is, it’s very hard to find a phone without extra features”

Participant 12

“Uuum, ok. Because... I don’t I don’t know. Because it’s something else that I would have to get to use and then become dependant on it and I actually don’t need it. So it’s almost like using something, and then creating a need for it, from it when actually I don’t need it. I don’t need it. If I wanted a camera I would carry a camera, surely. But I mean, and I have, the phone I had previously did have a camera in it and yes I did snap pictures on it. I never look at them and I never download them. So actually, did I need to take them? Because if they meant anything, I would have used them. And so they don’t. So there’s a lot of extraneous, meaningless umm Use... of... phones.”

Participant 13

“Eee... because... they have got lots of unnecessary things. I don’t need the access to internet, that’s why I don’t have it with this one, because I’ve got my computer in the house. I don’t need really need the camera, even though it has one, because I have got a digital camera. And am... I do have an mp3 player, you know... this has radio which I don’t use. I think these days this mobiles have too many things that are probably not necessary for me. A phone has to be just functional. Oh, it doesn’t appeal to me at all. Because I don’t think I would use the extras. And am...: It would just make the phone very, very confusing for me to operate. And I got in this one everything that I need, which is my

alarm clock, which is what I use. And an alarm clock is just... you know, my texts, I can make my phone calls and... The rest of the features that it has like games and all these extras, and Internet access... I got it but I don’t use it, ‘cos I don’t see the point.”

Participant 15

Interestingly from the above quotations, is that when people reach a certain age (perhaps around 30 yrs), the mobile phone stops being a gadget or a toy, and becomes a means to an end:

“That’s what the phone is for. That’s why it’s a mobile phone – it’s a phone to call and nothing else. The name of it just tells everything. It doesn’t have to make you a coffee or anything like that. It’s a phone.”

Participant 12

Aesthetic Appreciation

We recognized a recurring tendency to express considerable appreciation of the aesthetics of their mobile phones, particularly among younger participants. In 7 out of 21 interviews the appearance of the phone: its shape, its feel in the hand and its styling in general were described as the primary reason for its purchase. All 7 were under 30 years old and were the youngest members of the sample.

“Because it’s incredibly small and it looks like a calculator [laughs], and it’s nice, and easy, and very unusual because of the shape, and nobody else has this kind of phone, so it makes it special. And because it’s so small you can fit it to the purse.”

Participant 3

“It’s nice, it’s sexy, it’s thin, and it’s got a big screen and big buttons and I liked the way it lights up ... and the way it flips open and it’s black, and that it’s really tactile, the buttons, the buttons are metal and so it’s the case, so it feels nice in your hand it doesn’t feel plastic – y like, like the new ones.”

Participant 1

The look and feel of the phone, however important, does have to be complemented with functionality and this balance of features was a frequent requirement expressed by all the participants.

“Eeee... [pause] touch screen basically... I would say it’s because of touch screen, ... and... it’s because it looks... lets say, when I’m choosing a phone I probably base it more on looks than functionality... but I wouldn’t sacrifice functionality just for looks, so, it has to be some sort of balance”

Participant 2

A very similar attitude is evident in the following quote from participant 16:

“No, no not primary not, because most of them, most of them have camera, so no it wouldn’t be a primary selling point. I think for me it would be... it would be if it counts, the primary point for me would be how many features you get for your money. So, if there was a phone that was probably cheapest but still looked good, [laughs] you always want it to look good, it’s like, you know, something that is quite small, these little silly things like I like the ones with the top flips out... I don’t know, with no particular reasons I suppose, I think they cool, that’s, you know... so, that’s probably what it looks like like it’s probably more important then whether it has a camera or

not... You don't want something that looks... you know... I wouldn't like something that would look like you know the size of I don't know a brick, and those pink for example [laughs]"

Participant 16

Fun and Playfulness

Again, primarily among the younger participants, there were interesting examples of the enjoyment derived from ownership of the mobile phone. This sense of fun appeared to be predicated on a considerable interest in the 'additional' features present of their mobile phones. The following extracts illustrate these points

"Eh, well..., it's a walkman phone so that's what I actually use it most for and then I use it as a phone second I'd say probably, and then texting, that's pretty much it really..."

It's got other uses but I really don't bother with them, it's kinda (sic)..."

Participant 3

"Other than calling, I use the phone for obviously SMS; I use it for Internet access umm... I use a variety of things there I use it Very much as a portable media device. I... I use it as a walkman, basically. Umm, and it's also got GPS antenna in it so I use it for tracking where I am, so, umm... So I don't get lost. ... Umm... Technical features. I'm familiar with ahh... Well, I mean... that's... it its pretty much my hobby is various bits of gadgets and kit so umm... Yeah, I like it. I... I know all the details about it. Umm... What it can do and how well it can do it. I know that it's the second version of the same phone umm... the N95 original version was... had very poor battery life and so on... so this is a much better quality phone. Oh yea, and it's also got a very good camera on it, which I do use."

Participant 7

"It had a nice feel to it in my hand, I'm a very touchy-feely person and... its pink... I was feeling girly"

Participant 20

On the other hand however, as was already visible from the quotation above (participant 7), for some of the participants, especially those technologically oriented, the idea of having the appliance itself was already a source of strong enthusiasm. This can be also seen in the quotation below:

"I have an O2 XDA, which is manufactured by a company called HTC, which manufactures these smart phones for various manufacturers around the world, well for phone companies around the world. It's a classification of a, uh well it's a type of phone called a smart phone, which basically runs an operating system [pause] along the similar lines to an operating system you would find on a computer. So the so-called smart phone will allow you to synchronize with your operating system on your computer at home, and will allow you to play movies, will allow you to play songs *eh* that kind of thing. I use my phone every day and I use it for a number of purposes, I use it to keep in contact with friends, it acts as a memory device so if I need to remember to do a task then it goes in there. Eh, it acts as a diary, so if I have an important appointment it will tell me where I have to go and when I have to be there and hopefully beforehand as well, oh and it also wakes me up in the morning."

Participant 6

Simplicity and Usability

There is also evidence of mobile phones are treated like a 'necessary evil' (see also the next section). While expressing negative opinions about them they were also appreciated as playing an important role in their everyday live:

"I hate those devices is the... to be honest with you, I'm using my phone only to phone somebody, or receive the phone, send message or receive. Nothing else. I hate those extra features that all those kids are looking for and... Camera, never use it and I never will. I have a special digital camera to make a photos. I hate mobile phones, I despise them."

"Yes, I certainly do have a phone because like I said, it is useful thing but just for sending messages. I don't need to send a picture to somebody – I can always... describe it. I hate texting them – it takes far too long. I don't understand people sitting, walking and texting. My wife does it all the time. Just bought something and texting. To me it's... I don't know."

"Disable mobile phones! I know they are useful devices and for emergencies they are fantastic, and for communication, yes but let's not make it ... big deal out of it. Phone is a phone, you use it. Why normal house phones do not have games or anything? They are used just for phoning people, yes? Why mobile phones cannot be like that?"

Participant 12

Participant 14 expressed similar sentiments but less emotionally:

"All right... eeee.... ok... where we start... mobile phones are good. Sometimes. They are not good sometimes, because when you want to relax, when you want to nap... they make a lot of noise and then you wake up, whatever, but in general it's good because you can.. The latest mobile phones you can eee... kind off keep on track on what's happening, what events, you can do the... you can mark your appointments, you can play some games,... you can go on the Internet, so...I don't know, that's about everything... My relationship with my phone... em... I can't stay without a mobile phone, because I feel just left out. But as I said to you, sometimes it's like a wee bit too much..."

Participant 14

The usability, simplicity and functionality of the phone were the features that seemed to be most important for number of participants. Again, these comments were made by the older participants in this study. Usability issues included external features, like button size, for example:

"Right, ok. For me, numbers. The size of the numbers on the keypad because I was having to sort of pull it away from me to type it in and once I typed it in it was gobbledygook because I hadn't actually seen the letters and umm... so for me it was the numbers. I needed to see the numbers on the keypad really clearly. So what else did I like about this phone that I've got now? It's simple, it's not too complicated uhhhhh, very basic... yeah, numbers I think. The size of numbers."

Participant 13

Simplicity was an essential quality stated by every participant in this group:

“It’s the simplicity and it’s cheap. And I wanted something pay as you go which is not so expensive, so I went for this one.”

Participant 14

Functionality was frequently mentioned in conjunction with the price:

“[Interviewer] Ok, what phone do you have and why this one?”

Why, the one I’ve got is... I got am... (wonders, and is looking for a phone in her locker) what I have... I don’t remember now what the handset is... I got it because it does everything that I need, and it’s cheap, and it works fine.

[Interviewer] And what does it do that you need?

Well, I don’t use it... it has camera which I don’t need, and I don’t use because I have my digital camera, but ..Am.. I mean for £30 that costs me, you know... em... I needed basically for texting more than anything, because I use, ... I use my landline for most of my phone calls anyway..

[Interviewer] So, how did you chosen this phone

I went to Carphone Warehouse and I looked for what they have available and... am... this is a Sony Eriksson, and am... it was only £30, and... It works fantastically well, so... I’m quite happy... I chosen it basically because of the price. Because I don’t believe in paying for very expensive handsets.”

Participant 15

Involuntary Involvement: Thrownness

The involuntary use of mobile phones is also a recurring theme in these interviews. Examples of thrownness often centered around the question of ownership. Ownership has been imposed by others¹⁴ or prevailing circumstances:

“Umm, it was given to me as a present, so I didn’t choose it.”

Participant 5

“No, actually, well, it’s a present from my son, for my Christmas actually, so but my wife knew exactly what I want. No extra features, just the simplest thing ever. And she knew that I hate gadgets and I don’t use phone for anything else. Just to phoning. So, that was no??? I did not consult anybody or anything. I just wanted the simplest thing. The simpler, the better.”

Participant 12

For others it was a work necessity:

“My job currently also means that I travel a lot, which is obviously one of the reasons why I have my mobile phone. The phone I have is a company phone, and, emm, I’ve had that phone really ever since I’ve had the job, which is about 6½ years. In those 6½ years it’s only the second phone that I’ve ever had... eee... it has a limit on it so, that am.. I’m not using it

all the time, so we don’t waist the money when we could use a land line, if I’m in the office instead, or that kind of stuff.”

“Yeah, I was given the phone. I fill out form, saying I need a mobile phone and whatever, the cost etc, this sort of thing and then so many days later the phone arrives in the post. I don’t have any choices to what type of phone it is”

Participant 16

A third group of people succumbed to peer pressure:

“No, seriously, because I was generation that didn’t grow up with mobile phones. In the 80’s they were these huge things and only sort of yuppies had them in their sort of VW Golf GTI’s hanging out with this sort of huge elbow of a phone and so I didn’t grow up with that and so... and I found them quite intrusive and so I didn’t really want one and I fought it for a long time and then through work... I ended up needing one because other people I was connecting with had them and it was like “what’s your mobile phone number?” “I don’t have one” “why not” and I was tired of explaining it so I got one and then I became dependant and my phone is... the basic idiot-proof Nokia with large numbers so I don’t have to put my glasses on to text. Umm... I do carry it with me. I don’t carry it in my pocket because I’m also prey to conspiracy theories about microwaves and should I have it near my kidneys.. Or should I... I shouldn’t tuck it in the brim of my hat, it might fry my brain. I don’t listen... I don’t speak too long on it, *and*, yeah. So I’m kind of an antiquated user, but it’s alright”

Participant 13

DISCUSSION

“Familiarity consists of dispositions to respond to situations in appropriate ways”.

[9, p. 117].

Heidegger would have us believe that familiarity is the means by which we make sense of the world and the equipment its comprises. Being familiar with the world is a necessary condition or consequence of being-in-the-world. And we necessarily make sense of the unfamiliar by relating it to the familiar.

The phenomenological roots of familiarity means that there is no psychology of familiarity as such but this does not preclude its empirical exploration. At the beginning of this paper we set out to examine whether familiarity does indeed comprise, as Dreyfus has suggested, understanding (as “know-how”) and involvement (as comportment).

As we expected analysis of our interview data revealed evidence of know-how as a mixture or collage of cognitive constructs including propositional knowledge and mental models. We also found evidence of what we might describe as cognitive heuristics, namely, the use of anthropomorphism. It is likely that the use of anthropomorphism, that is, ascribing human attributes to the technology by our interviewees may be the source of Norman’s observation about mental models being “unscientific and exhibiting “superstitious” behavior [16].

Contrary to our expectations, evidence for involvement was less clearly defined. We had expected a fairly clear split between, “know-how” as cognition/mental models and involvement/comportment as aesthetics, fun, pleasure. But the data do not support this simple, perhaps even naïve split. Instead we found that for most people comportment centered around usability. People liked their phones because they were easy to use; and preferred a particular brand because it’s

¹⁴ A day after the interview, one of the participants who had described having been bought a mobile phone by her daughter reported the following. She had been out to lunch with her daughter who berated her for not using the phone the daughter had bought her. The mother excused herself in a variety of ways (“it is sitting on the kitchen draining board”) until the daughter accused her mother of being selfish in that had the daughter suffered an accident she would be unable to call her mother for help!

functionality. However one third of the participants (exactly 7 of the 21) did express aesthetic, playful and fun-oriented comportment in addition to remarks about the devices' usability. These participants were the younger members of our sample (< 30 years)¹⁵. This (suspected) age horizon also manifested in terms of older people (30+ years) consistently expressing a preference for mobile phones with fewer features. Older people saw the technology as merely being a 'means to an end', that is, for making calls, keeping track of children, do their jobs. While this is not to suggest, that anyone over the age of 30 is incapable of appreciating the aesthetic (or whatever) of a particular phone, but does seem to point at what the neo-Heideggerian philosopher Albert Borgmann calls the device paradigm [18, 19]. The device paradigm is an account of technology use in which people treat it purely instrumentally, that is, as merely a means to an end (and with little regard for the means). Technology, for example, makes the procurement of goods "instantaneous, ubiquitous, safe, and easy" [18, p.41]. As part of this discussion he also distinguishes between commodities and things, a commodity is a context-free entity isolated from traditions and customs. A thing, in contrast, is capable of engaging and connecting with us. So, for example, a hamburger bought from a chain is an example of a *commodity*, whereas a home cooked dinner is a *thing*. Hamburgers are seen to be uniform, safe, reliable and quantifiable (though may be seen to be contributing to the homogenization of society); while a home cooked dinner relies on the skill of the cook, the availability of ingredients, time, effort and is an experience not easily susceptible to quantification. One is packaged and delivered (often in an opaque wrapper) while the other is open to inspection, modification and even participation. In the context of interactive technology we now buy and use mobile phones and other consumer electronics which are known to be safe, reliable and usable without necessarily having the faintest of ideas of how they work.

Further Work

There are two lines of further work at which this current exposition points us. The first is to investigate whether there is indeed an age related divide between those who buy a mobile phone for its 'looks' and those for its utility. Another Borgmann's concept that of 'technological horizon' might guide us here. Everyday exposure to technology also establishes a 'horizon' - a *technological horizon* which is the baseline of our familiarity with it. So, for example, shopping has traditionally involved going to a store, selecting goods and so forth. This practice is one which many of us will see as natural, authentic and, of course, familiar. This everyday practice establishes our technological horizon for shopping. Shopping at a dotcom is quite a different experience, which may, in its own way establish the technological horizon for the current generation.

Secondly, Heidegger tells us that our identity (or identities) as designers, academics or being the coolest kid on the block are a consequence of how we exploit or use the equipment of which the world is comprised. From this Heideggerian perspective we are defined by our skills. This too is worthy of further consideration.

¹⁵ The one exception to this in our data was one individual who had chosen a pink mobile in order to be 'girly' specifically recognizing that such phones were designed for young and teenage girls.

REFERENCES

- [1] Polyani, M. 1983. *The Tacit Dimension*. Peter Smith, Gloucester, MA.
- [2] Wartofsky, M. 1979. *Models: Representation and Scientific Understanding*. Reidel Publishing Company, Dordrecht, Holland.
- [3] Bucciarelli, L. L. 1994. *Designing Engineers*. MIT Press, Cambridge, Massachusetts.
- [4] Goodwin, C. and Goodwin, M. H. 1998. Seeing as a situated activity: Formulating Planes. In: Engestrom, Y. and Middleton, D. (Eds.). *Cognition and Communication at Work*. Cambridge: Cambridge University Press.
- [5] Chase, W. G. and Simon, H. A. 1973. Perception in chess. *Cognitive Psychology*, 4, 55–81.
- [6] Farrow, D. and Abernethy, B. 2003. Do expertise and the degree of perception – action coupling affect natural anticipatory performance? *Perception*, 32(9), 1127–1139.
- [7] Turner, P., Turner, S. and Van De Walle, G. 2007. How Older People Account for their Experiences with Interactive Technology. *Behaviour and Information Technology*, 26(4), 287–296.
- [8] Heidegger, M. 1982. *The Basic Problems of Phenomenology*. Bloomington: Indiana University Press.
- [9] Dreyfus, H. L. 1991. *Being-in-the-world: A Commentary on Heidegger's Being and Time, Division 1*. Cambridge, MA: MIT Press.
- [10] Heidegger, M. 1927/1962. *Being and Time*. (Translated by J. Macquarrie and E. Robinson). New York: Harper Collins.
- [11] Heidegger, M. 1985. *History of the Concept of Time*. Bloomington: Indiana University Press.
- [12] Payne, S. J. 1991. A Descriptive Study of Mental Models. *Behaviour and Information Technology*, 10, 3–21.
- [13] Turner, P. and Sobolewska, E. 2009. Mental Models, Magical Thinking, and Individual Differences. *Human Technology*, 5(1), 90–113.
- [14] Gentner, D. and Stevens A. (Eds.). 1983. *Mental Models*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.
- [15] Johnson-Laird, P. M. 1983. *Mental Models*. Cambridge: Cambridge University Press.
- [16] Norman, D. A. 1983. Some observations on mental models. In: Gentner, D. and Stevens, A. (Eds.). *Mental models*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates.
- [17] Reeves, B. and Nass, C. 1996. *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Cambridge, UK: Cambridge University Press.
- [18] Borgmann, A. 1984. *Technology and the Character of Contemporary Life*. Chicago: The University of Chicago Press.
- [19] Borgmann, A. 2000. Reply to My Critics. In: E. Higgs, A. Light and D. Strong (Eds.). *Technology and the Good Life?* Chicago: University of Chicago Press.

User Testing when Test Tasks are not Appropriate

Sirpa Riihiaho
Helsinki University of Technology
P.O. Box 9210, 02015 TKK
Finland
Sirpa.Riihiaho@tkk.fi

ABSTRACT

This paper presents two usability evaluation methods called informal walkthrough and contextual walkthrough. Both the methods are intended for situations where specific prepared test tasks are not appropriate either for the use context or the goal of the evaluation. Informal walkthrough is intended for evaluating novel systems whose concept is familiar to the users and whose intuitiveness is of major concern, for example gaming slot machines. Contextual walkthrough, on the other hand, is intended for evaluating systems that cannot be separated from their real use environment and use context, for example call center applications. In addition to presenting the methods, this paper will give some examples of their use, and assessment of how effective they would be in testing ubiquitous systems.

Keywords

usability evaluation methods, user testing, usability test

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – ergonomics, evaluation/methodology, user-centered design.

INTRODUCTION

Traditional usability testing methods are strongly based on carefully prepared test tasks that the test users are asked to perform in a very controlled conditions (e.g. [15]). Thereby, all participants perform the same tasks under as similar conditions as possible to make gathering of quantitative data easy and reliable. These methods are appropriate for settings with well-known tasks and outcomes, and are highly effective for influencing decision making in organizations, but they are limited in their ability to gather data on true user tasks, task flows, user profiles and context of use [19]. As Greenberg and Buxton [6] state, quantitative empirical evaluations give us something that appears to be scientific and factual instead of expressing opinions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

The increasing mobility and ubiquity of devices make predicting the context of use, and thereby its usability, more difficult. For example, Scott claims that usability testing must be mobile, modular and contextual, and should also employ user-driven tasks to be accurate and relevant [23]. Also other practitioners (e.g. [20]) and researchers (e.g. [5]) suggest that usability evaluation methods should provide room for the user experience and their arguments, reflections or intuitions about a design. Contextual walkthrough and informal walkthrough methods that are presented in this paper try to respond to these challenges.

Although the need for new evaluation methods is recognized and admitted, the trend in developing the methods is decreasing. For example, Barkhuus and Rode [1] analyzed trends in the approach to evaluation taken by CHI papers in the last 24 years. They chose papers from five years, about six years apart, starting from 1983. They were primarily interested in how big a portion of the papers included a usability evaluation and of what type. They classified papers whose contribution was primarily an evaluation method in their own category. They found a clear linear decrease in the percentage of these papers ending in zero in year 2006. Hopefully, this paper helps in recovering this research area.

BACKGROUND

We have been doing research and given courses on usability evaluation at Helsinki University of Technology (HUT) since 1993 [12]. During these years, we have evaluated nearly 200 systems in our research projects and course assignments. The evaluated systems have included desktop systems for professional use, gaming slot machines, virtual worlds and smart products such as heart rate monitors and televisions. The goals of the evaluations have varied as well as the systems, so we have needed a wide range of user testing methods to reach our goals.

Our usability research and teaching at HUT is strongly based on iterative user-centered product development process presented, for example, in the ISO 13407 standard [10] (see Figure 1). Our research projects are usually done in cooperation with companies, and the course assignments contribute either to the research projects or are commissions from the companies. Therefore, the methods that we teach and apply in our research projects always try to influence the development of a certain product and are tied to a certain phase of a product development process.

The goals of our evaluations fall into two major categories:

1. Finding usability problems in a product under development to make it better before releasing it.
2. Finding problems in products already in use to get ideas for improvement in the next releases or new products.

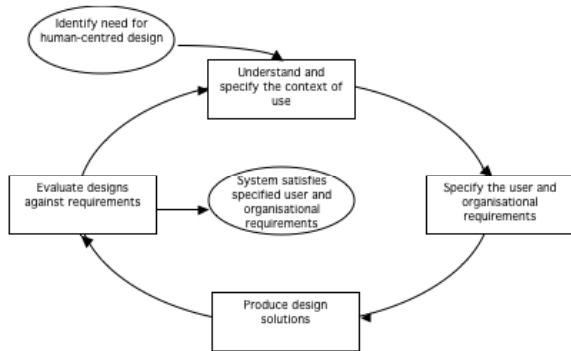


Figure 1. Iterative user-centered design activities.

We divide the usability evaluation methods into two categories by the extent to which users are involved: those that require users’ attendance, i.e., user testing methods, and those that can be used without the users, i.e., usability inspection methods (see Figure 2). On each assignment we use multiple methods: usually one or two inspection methods (typically a heuristic evaluation [15] and a cognitive walkthrough [25]) and one or two user testing methods supplemented with interviews and questionnaires.

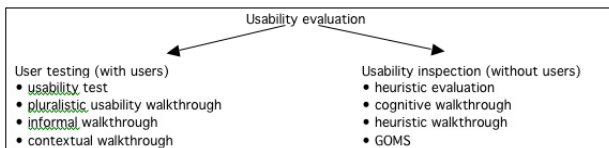


Figure 2. Two categories of usability evaluation methods.

The aim of my research has been to give usability experts and system developers a set of usability evaluation methods, especially user testing methods, wide enough and flexible enough to offer a suitable set of methods for each situation. As for example Nielsen [15] and Gould [5] present various methods for different phases in the product development cycle for the usability practitioners’ methodology toolbox, my research has focused on variations of usability testing in laboratory conditions or in the field. The other methods we have applied mostly as they have been presented in the literature.

For example, in situations where prepared test tasks have not been suitable in user testing, we have developed methods called informal walkthrough and contextual walkthrough. The methods are not precise and do not give detailed steps for interacting with the users, but they present the atmosphere that the test sessions should have and the goals that they are intended to. This paper will give a brief description of the methods, give some examples of their use, and assessments how suitable they would be in testing ubiquitous systems.

USABILITY TESTS

Both the informal walkthrough and the contextual walkthrough methods have been developed from the traditional set up of a usability test. Therefore, the phases of traditional usability testing and traditional usability test sessions are presented briefly in the following to ease the comparison of the methods.

10 steps for Usability Testing

Usability testing starts from determining the test goals, proceeds to designing the test, conducting the test sessions and analyzing the tests and their results, and ends to reporting and

communicating the results to the development team. Hansen [7] gives ten steps for a successful usability testing:

1. Get the background information about usability evaluation.
2. Make a test plan.
3. Design the test.
4. Arrange a test environment and equipment.
5. Conduct a pilot test.
6. Recruit participants.
7. Set up the test room.
8. Conduct the test.
9. Compile and analyze the results.
10. Recommend changes.

10 steps for Conducting a Test Session

Conducting a test session is a subtle task and requires respect to the users and their effort on improving the system. Gomoll [4] gives ten steps for conducting a session of user observation, such as user testing:

1. Set up the observation: the tasks, the participants and the context of use.
2. Describe the purpose of the observation in general terms to the participants.
3. Tell the participants that they may quit at any time and still get the fee that has been agreed.
4. Introduce the personnel and the equipment in the room.
5. Explain how to think aloud.
6. Explain that you cannot provide help during the test.
7. Describe the tasks and introduce the product.
8. Ask if the user has any questions, and then begin the observation.
9. Conclude the observation.
10. Use the results.

It is very important that the participants know and remember that it is the product that is tested, not the users. Other main ethical considerations include that everything is ready before the participants arrive, and the participants are informed about the state of the system and of the confidentiality of the results [15].

INFORMAL WALKTHROUGH

Informal walkthrough is a mixture of usability testing, thinking aloud, observation and interview. The most significant difference to traditional usability testing is in the test tasks: in an informal walkthrough, there are no specific, pre-defined test tasks. Instead, the test moderator has a list of features that the users should walk through in their own pace and order [16].

Compared to the ten steps given by Gomoll [4], the informal walkthrough method differs mainly in step number 7, during which the purpose of the system is explained in general level instead of giving specific tasks or presenting the system in detail. Pre-defined test tasks are only designed for those features that are of special interest in the evaluation, and they are given to the users only if they do not spontaneously use these features during the walkthrough. Sometimes, use scenarios are used instead of test tasks to lead the users to these features.

In informal walkthrough sessions, the users are asked to explore the system in the way they would if they were alone, and to think aloud while exploring the system. They are also encouraged to comment the system at any time. The test moderator may interrupt the users for questions, but mainly just observes them. With some systems, such as gaming slot machines or multi-user games, the setting is more natural and more relaxed if there is a pair or a group of users exploring the system at the same time instead of a single user.

History of the Method

We developed the informal walkthrough method as we were evaluating the usability of a news-on-demand service running on a television [16]. The system was on the level of a functional prototype and it could be operated with the television remote control. The main goal of our evaluation was to study how intuitive the news-on-demand system was, i.e., how good an affordance it had. In addition, we wanted to assess how well the users noticed and spontaneously used a new feature we added to the system. The problem with traditional usability testing in this kind of a situation is that the given tasks indicate that the system provides the features required for the task, and the users are gently “forced” to use the features. The system we studied was meant to be entertaining and self-explaining needing no manuals. Therefore, we wanted the users to find the essential features themselves and to figure out their meaning. If some parts were not found, this would indicate that these features are not wanted or they have a poor affordance.

We wanted the users to explore the system and to test the features they were interested in. Therefore, we did not give them any prepared tasks in the beginning of the test. Nevertheless, we wanted to hear all the users’ comments about some features that we expected to have problems based on our own inspections, and on the new features that we had added to the system. Therefore, we made a list of all the features the system included and marked the ones that we wanted every user to visit (see Figure 3). As the users explored the system, the moderator of the test marked the visited features. We used three types of markings: “X” for the situation where the users found the features themselves, “A” for finding the feature by accident and “-” for missing the feature. After the users had explored the system, the moderator checked whether there still were some features that had to be visited. If the users had not yet found some compulsory features, the moderator started to ask questions that guided to these features or gave specific, pre-defined tasks.

Function, button or a concept	The participant found it himself (X)/ no (-)/ by accident (A)	Correct use	The participant was helped
Newspaper concept			
Navigation in the table of contents	X	X	
Navigation: next and previous news	X	X	
Page numbers	A		
Finding background information	-		X

Figure 3. A checklist of the system features. The compulsory parts have been encircled.

Requirements and Guidelines for Use

The informal walkthrough method requires a finished product or a functional prototype. The product concept must be familiar to the users so that they are able, willing and as relaxed as possible to explore the product. If the product is evaluated in such an early phase that there is only a partly functional prototype available, we recommend that the system developers participate in the tests. The early prototypes are not very steady and the developers are usually the only ones who know how to fix the problems quickly. In these situations, extra care should be taken to keep the developers and their opinions in the background during the test sessions.

Since the users do not get any specific tasks, each session is unique. All the users have their own ways of learning and

experimenting new things, in addition to having various needs and expectations for the system. Therefore, the test moderator should let all the users behave in their own way and proceed in their own order without disturbance. This way, we get realistic results of how easy the system is to learn, what features the users easily find, which ones they use first and what features they long for. The users are asked to think aloud and comment the system as much as possible. Therefore, the method is not recommended to measuring efficiency or effectiveness. Instead, it is a very good method to evaluate how intuitive, easy to learn and easy to navigate the system is.

Informal walkthrough method can be used both in the laboratory settings and in the fields. If the system is still on the level of first prototypes and needs complicated settings to function, it is better to keep the testing environment controlled and focus on finding the biggest obstacles in its use as well as the most enjoyable and rewarding features. In addition to its openness to the test environment, informal walkthrough method does not restrict the number of test users in a test session. Instead, it is recommended to let the users act as they normally would, i.e., alone, in pairs or in groups.

An Example of a Typical Use

We have often used informal walkthrough method in evaluating gaming slot machines, because the concept of the machines is familiar to the users, and because the interpretations that the users make of the machines and the games have been of main interest instead of performance times. Since the slot machines are quite often used in pairs or in groups, we usually include one or two paired-user sessions to the set of single user tests. Other researchers have used informal walkthrough method for example to usability evaluation in real context involving participants with cognitive disabilities [13].

To ensure that the users try some interesting functions, we prepare scenarios that lead to situations involving the use of these functions. For example, we may tell the users that their bus is about to leave in a few minutes. This means that they have to finish their games, collect their wins and possibly cash the sum they did not have time to spend in gaming. These subgoals are not specified, but left for the users to decide.

CONTEXTUAL WALKTHROUGH

Contextual walkthrough is a combination of usability testing and contextual inquiry. In contextual walkthrough, the real context and real tasks are essential as in contextual inquiry, but in contextual walkthrough, the goal of the observations is to reveal usability problems in the evaluated system instead of gathering ideas for new development projects and understanding the work as a whole. Usability experts conduct the walkthroughs instead of designers, and the analysis of the results resembles more the analysis of usability testing than analysis of contextual inquiry.

Compared to the ten steps given by Gomoll [4], also the contextual walkthrough method differs mainly in step number 7. The tasks arise from the use context and the product is usually already familiar to the users and need no introduction. Also step number 1 differs in that respect that in contextual walkthrough the context and tasks are authentic instead of set ups. Explaining how to think aloud (step 5) is only made if thinking aloud is appropriate in the use situation.

We have applied contextual walkthrough, for example, in evaluating the usability of call center applications. These applications are used while talking to a customer on the phone,

so thinking aloud is not appropriate. The call center assignments have differed in whether the operators have had enough time to review the calls between the calls or only later during a brake with the help of the video recordings.

History of the Method and an Example

We developed the contextual walkthrough method during an assignment, in which the customer wanted us to study the efficiency of a new application and to measure the operators’ performance times also in the low-level actions in the call center. Therefore, we studied the use of GOMS [3] in the beginning of our work and decided to apply the method in analyzing our observations and videotapes.

As we went to the call center and observed the operators’ work, we soon realized that the performance times of the low-level actions in the application were irrelevant compared to the other tasks that the operators typically did during the calls and especially after the calls. Therefore, we gave up on the use of GOMS and decided to observe the operators as they worked, and switch more to contextual inquiry [2], [8].

The operators’ work cannot be interrupted as they talk with the customers on the phone. Therefore, we could not apply contextual inquiry as such. As Beyer and Holtzblatt [2] suggest, we observed the operators quietly during the calls and waited for the moment when the operator could explain the contents of the call and the consequences it had. We recorded the sessions on videotape, but we did not review the tapes with the operators.

The operators took the calls just the same way as normally. After the calls, they gave a short summary of the call and explained what they had to do next. When they had finished the tasks, we talked about the task for a while. At this point, we could confirm our findings with the user and ask further questions.

The day we visited the call center was quiet, so the operators had also time to present their work beyond the calls. For example, they dug up some reports they had to check and fill in with the new application just to explain that part of their work to us.

Contextual inquiry is intended to gathering information about the users’ work for the basis of design work. Our initial goal was not to study the users’ work but to study the usability and especially the efficiency of an existing application. If we had just measured the performance times of the tasks that operators did with that specific application, we would have missed the findings that the view of contextual inquiry revealed. Although the application had some usability problems, they were irrelevant in the work overall. The biggest problem in the operators’ work was the need for several separate applications and the physical location of these applications instead of the usability problems of the new application.

COMPARISON TO OTHER METHODS

The usability evaluation methods can be classified in numerous ways. Some attributes for division include:

- Goal of the study (understanding the work, requirements gathering or usability evaluation of a specific system)
- Object to be studied (culture, work or leisure time, or systems or services)
- Environment for the study (laboratory or fields)
- Number of participants (single, pair or group)
- Type and origin of the tasks (pre-defined tasks, only goals or informal).

Ethnographic studies are used to understand the patterns of life and culture that people have. They require longitudinal studies and usually imply only a few participants. Contextual inquiry is originated from ethnographic studies but concentrates on the work involving the systems that the customer company is developing instead of the lives of the employees.

As contextual inquiry goes one level lower in the goal of observation compared to ethnographic studies, contextual inquiry and informal walkthrough go even lower: they concentrate on the usability of the system to be evaluated. In contextual walkthrough, the evaluation is made in the real context giving the opportunity to assess the role of the system in the big picture of the work or tasks at hand. Informal walkthrough can be conducted both in laboratory conditions and in the fields. Table 1 summarizes these differences between the methods.

Table 1. Some methods for usability studies classified according to the aim and object of the study.

Aim of the study	Understanding	Requirements gathering		Usability Evaluation	
Object of the study	Life, culture	Work	Leisure time	Tools, systems	Products, services
Suitable methods	Ethnographic studies: observations, interviews, questionnaires	Contextual Inquiry	Probes, use diaries	Usability tests	
				Contextual walkthrough	Informal walkthrough

The usability evaluation methods also differ in the way that they resemble the real use context and the authentic use situation. Observation and ethnographical studies are usually conducted as longitudinal studies observing the participants for a longer period of time instead of a session lasting about an hour or two. They get a “glimpse of the real world”. Contextual inquiry lasts only one working day or less, and the participants sometimes try to reveal only their “better side” to the observers leaving some “boring routines” behind. Contextual walkthrough and informal walkthrough in the real use context have this same constraint, and they concentrate on the use of one system instead of the work (or a hobby) as a whole. Evaluations with predefined tasks or in laboratory conditions miss most of the factors that real context and authentic tasks include. Figure 4 shows some methods according to their analogy to the real use context.

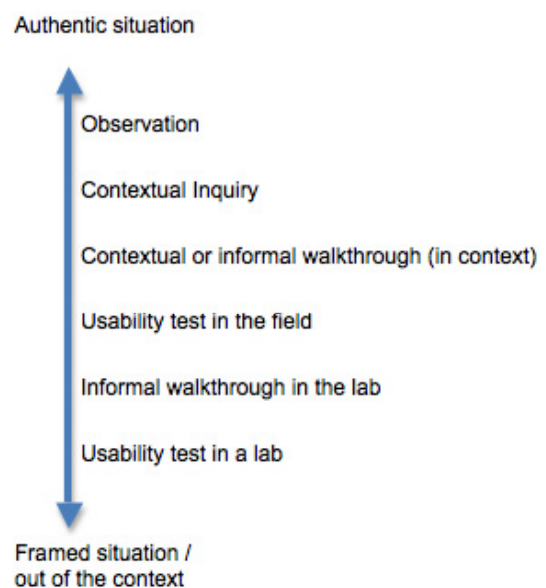


Figure 4. Methods according to the level that the research situation resembles the authentic use situation.

Other researchers have made similar attempts in developing the evaluation methods. For example, Savioja, Norros and Salo have created an evaluation method called Contextual Assessment of Systems Usability (CASU) that aims to evaluate both the information content and the presentation of the information in the systems [21]. They use real contexts in evaluating complex systems in nuclear power plant main control rooms as in contextual walkthrough, and they leave certain degree of freedom to the definition of the tasks in their experiments as in informal walkthrough, but they still use simulated process scenarios instead of authentic ones for the experiment in studying new designs, as is certainly necessary for safety reasons.

EMPIRICAL DATA FROM RECENT USABILITY TESTS

To get a better idea of the methods, I shall present a few examples of how we have applied the methods in our recent studies on the course assignments last spring. This time, we applied informal walkthrough to evaluating a game played in a group on a touch screen table, and to a web application supporting the use of a heart rate monitor. The contextual walkthrough method was applied in evaluating a web service for getting a building permit.

Case 1: Informal Walkthrough to a Game

This spring, we used informal walkthrough method to evaluate the usability of a multi-user videogame called Shepherd. The game is built on MERL DiamondTouch table, which is a touch-and-gesture-activated screen for small group collaboration (see Figure 5). As the concept of the multi-user touch-table is quite new to our research group, we were interested in getting ideas for its use in other games as well as in getting ideas for improving the present version of the Shepherd game. Therefore, we let the users explore the game quite informally on their own and interviewed them afterwards for their comments on the game and ideas for improvements and new games.



Figure 5. MERL DiamondTouch table with four users.

The Shepherd game is developed for four simultaneous players. The idea of the game is to gather one's own sheep to their stall. All the sheep are wandering in the fields at the beginning of the game, and each shepherd calls for his own sheep by whistling that is simulated by tapping on the table with a finger.

The setting of an informal walkthrough with no instructions on how to play the game revealed some problems in the interpretation

of the controls of the game. The users intuitively sat on the chairs that had the sensors required for controlling the table, and they also realized that they could control the table, but they did not immediately realize how to call for the sheep. Especially the kids first tried grabbing by hands instead of tapping with a finger on the table. The users also had difficulties in understanding the range of the "whistle", since the call had no auditory or visual feedback, and only calls made near one's own sheep had any effect. If the game had been demonstrated or even just explained to the users, this information about wrong interpretations would probably have been lost. The scenario of having the game available in a trade show worked well, and the test moderator acting as a janitor ready to help in technical problems was quite a natural role in the situation.

The informal setting also revealed many positive features in the system. For example, both the children and adult users enjoyed especially the learning phase of the new game and finding its features. The kids enjoyed the game itself, whereas the adult users liked more the concept of the multiuser touch table and its features than the game. Playing the game and trying out the new concept awoke numerous ideas for next versions of the evaluated game, as well as ideas for quite new games, just as we had hoped for.

Case 2: Informal Walkthrough to a Web Application

To evaluate the usability of a web application supporting the use of a heart rate monitor, we conducted informal walkthroughs in real use contexts. The development team was interested in the problems that the users might have in getting the application and in transferring the data from the monitor to the application. Analyzing the data was left to little attention, because it had been studied in previous assignments.

To motivate the users to transfer and analyze the data, we gave the users the monitors about a week before the tests, so that they could use them in their own exercises and collect data of their own. To make the situation as natural as possible and as easy for the users to attend as possible, we conducted the tests at the users' homes. In the test sessions, the users were asked to transfer the data they had collected with the monitor to the evaluated application, and to study the results. The setting was not thereby totally informal, but it did not include any instructions of the required phases or subtasks as in traditional usability tests. As only the goal of the data transfer was given to the users, the setting was almost as in an informal walkthrough.

The tests revealed serious gulfs between the subtasks. The data transfer and analysis required four subtasks: loading the application, installing it, transferring the data and analyzing it. The users had quite little problems inside these subtasks, but they did not get enough feedback or instructions to proceed from one subtask to another. If the tasks had been given in pieces, as in traditional test settings, this problem would probably not have been found.

The real use context – in this case the users' homes – also revealed problems in the software requirements of the application. The application required that the computer had programs that all users, especially non-technical computer users, do not have on their home computers. Still, the application was directed especially for these non-technical users. Some functions were also dependent on the type of the operating system and the browser. In laboratory conditions at least the operating system is fixed, but sometimes the user can select at least the browser to be used.

Case 3: Contextual Walkthrough to a Web Service

Contextual walkthrough method was applied to evaluate the usability of a web service for getting a building permit. Contextual walkthrough was used only to study the use of the expert users, and the novice users' actions were studied in traditional settings with pre-defined test tasks.

The test sessions of the contextual walkthrough method were conducted in the real context, i.e., at the users' work places. They were asked to use the evaluated web service while we were there and to use it for their own projects.

Observing and interviewing the expert users helped the evaluators to understand the use of the system, and to create credible tasks for tests with novice casual users. Observing the expert users also revealed, that all the terms in the system were not familiar even to the experts used to the terminology of the traditional application forms. Finding the right place for a piece of information that needed to be modified was also difficult even for the expert users. In addition, the expert users wanted more information about the application after it was sent to the officials: how does it proceed in the bureaucracy, how long does it approximately take to get feedback of the results of the application, and does the system support tracking the process.

The real context made the test sessions more natural to the users: they could use their own computers with their own browsers. As they also could do their normal real work instead of faked projects, they could at the same time assess the usefulness of the system and compare it to the traditional way of filling in paper forms and delivering them to the officials in office hours.

VALIDATION OF RESULTS

The informal and contextual walkthrough methods do not provide quantitative and comparable results, but they give lots of qualitative data about the user experience, the users' expectations and interpretations. In this way they resemble ethnographic studies, and thereby also include only a few participants, usually about five users and a pilot user testing the test settings.

To validate and to ensure the reliability of our results, we always use several methods in our evaluations. This triangulation usually includes one or two inspection methods (at least a heuristic evaluation, and quite often a cognitive walkthrough for a selected part of the system as well), and interviews and questionnaires with the users. When the results of all these methods indicate the same problems and the same strengths in the system, the results and our suggestions for improvements lie on a pretty safe ground.

ASSESSMENT OF USE IN UBIQUITOUS ENVIRONMENTS

Scholtz and Consolvo have developed a framework for evaluating ubiquitous computing applications that they call the framework of Ubiquitous Computing Evaluation Areas (UEAs) [22]. The areas and their metrics (metrics in parenthesis) include:

1. Attention (focus and overhead)
2. Adoption (rate, value, cost, availability and flexibility)
3. Trust (privacy, awareness and control)

4. Conceptual Models (predictability of application behavior, awareness of application capabilities, and vocabulary awareness)
5. Interaction (effectiveness, efficiency, user satisfaction, distraction, interaction transparency, scalability, and collaborative interaction)
6. Invisibility (intelligibility, control, accuracy and customization)
7. Impact and Side Effects (utility, behavior changes, social acceptance, and environment change)
8. Appeal (fun, aesthetics and status)
9. Application Robustness (robustness, performance speed, and volatility).

As we have not yet applied informal or contextual walkthrough methods to ubiquitous systems, I can only present my considerations of how the methods would find information about these areas on the basis of the results in our previous studies with the methods. It seems, that most of the areas would be covered by using real use context with users that already have some experience with the evaluated system or similar systems, as in the case of contextual walkthrough. Thereby, the users already have some basis for making interpretations of the usefulness of the system and the side effects it may have.

If the system is and its concept are quite new and thereby have only novice users, some of the areas, such as adoption, trust, impact and side effects and even the invisibility would probably require multiple sessions for contextual walkthrough, between which the users should be let to use the system on their own. This way they would have a chance to assess the usefulness of the system in their own work or life.

In informal walkthrough sessions, we have got information about the users' attention, conceptual models, interaction and appeal quite easily along with other information about the users' behavior. The invisibility, i.e., the integration of the system into the user environment, has not yet been studied but seems possible if the method is applied in the real use context.

DISCUSSION

Although the methods presented in this paper are no longer new [16], [18], they have again become topical due to changes in the type of evaluated systems, their use and user population. This subsection discusses some of the demands that have been presented in the literature, and the way that informal and contextual walkthrough methods try to respond to these challenges.

Room for Various Interpretations?

Users make various interpretations of the systems based on their previous knowledge and experience. Different users make different interpretations and the interpretations they make may change in time. Sengers and Gaver [24] give three levels of users' interpretations:

1. What does this do; how do I operate it?
2. What good is it for me; how can I use it in my tasks?
3. What value does it give to my life and social status?

Sengers and Gaver document how design and evaluation strategies shift when we abandon the presumption that a specific, authoritative interpretation of the systems we build is necessary, possible or desirable. For evaluation, they suggest

ethnographical studies, dynamic feedback by giving the collected information back to the users to interpret, and longitudinal studies for understanding users' changes in interpretations especially on the high level. They remind the readers that short-term studies and methods only catch a single snapshot of the many interpretations that users may develop across time in using the system.

Although informal walkthrough and contextual walkthrough methods imply short-term studies, especially contextual walkthrough can be used to collect data and user experiences of systems after they have already been used for a while for real tasks, and the value of the systems has been tested in real settings.

Contextual walkthrough is more intended for the evaluation of the second level interpretations instead of the basic level interpretations about how to operate and control the system. Also informal walkthrough method seeks to get information about what the users want to do with the system: which features they look for and try out, and which ones they notice but leave untouched.

Positive Findings Neglected?

Traditional usability evaluation methods focus on negative (e.g. [6]) although the impact of the results improves if the positive results are also presented. Molich [17], for example, suggests that one positive finding should be presented for three negatives to keep the designers motivated to make the suggested improvements.

In our experiments, informal walkthrough has proven to be an effective method in gathering positive findings. As the users get relaxed and are able to explore the system on their own, they have time to make interpretations on various levels, including the level of usefulness. The informal walkthrough also leaves room for the user experience and users' arguments, reflections or intuitions about a design, as Greenberg and Buxton [6] claim for.

Context of Use: a Problem or a Possibility?

The International Standard ISO 9241-11 defines usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [9] (see Figure 6). Thereby, the context of use has great relevance to the usability of a system.

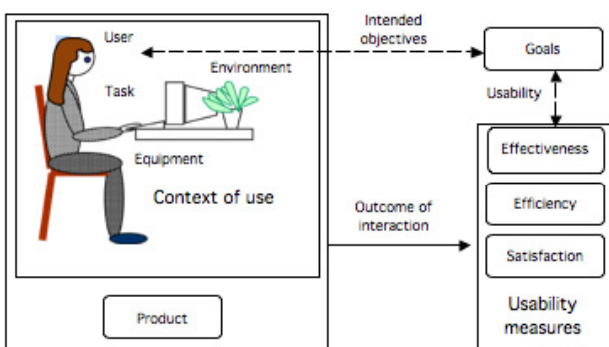


Figure 6. Usability framework according to ISO 9241-11.

As the systems have shifted from office systems for technical users to systems intended for everyday life in various use contexts and by very heterogeneous users, the evaluation methods must also change in their nature. Scott, for example,

states that we need methods that can anticipate and account for unexpected and continuously changing contexts of use [23]. She is also worried about the problem of selecting the contexts of use for studies.

On the basis of the study by Kim et al. [11] the use contexts of Mobile Internet, on the other hand, are of a concentrated type rather than being widely diverse. They studied the use contexts of Mobile Internet by categorizing the contexts by eight parameters including the reason to use the system, emotional status, one or two hands used, moving or staying still while use, visual and auditory distractions, number of people around and the level of interaction with others.

Their results showed that just 2 types of use contexts out of the 256 possibilities covered over 20% of the 1552 reported use sessions, and just 14 contexts covered over 50% of the sessions. The most common use context (14,6% of the sessions) was for a hedonic goal, in joyful mood, only one hand used, legs not moving, in calm and quiet environment, having only few people around and involving only low interaction with others. The second most common context was otherwise the same but the goal was utilitarian. They also noticed that 39 of the 256 possible contexts (38,7%) were never used.

The study by Kim et al. cannot be generalized to all systems, but still shows that some contexts are probably more common than others with all the systems, and are thereby worth more attention than others both in design and in evaluation. This also implies that for example contextual walkthrough can be used to cover a substantial portion of the problems affecting the usability of a system, if the most common use contexts are discovered and tested.

CONCLUSIONS

The line between contextual and informal walkthrough is indeterminate. Informal walkthrough can be carried out both in the laboratory and in the fields, whereas the contextual walkthrough requires the real context, as its name implies. In both the methods, the test tasks emerge from the use context, the real work or the scenario presented to the user.

One of the most significant differences in the methods is in the attitude towards the studied system: contextual walkthrough is usually applied to systems for professional use in certain working environment, whereas informal walkthrough is usually applied to free time applications emphasizing the user satisfaction and appeal.

Especially informal walkthrough method leaves room for the user experience and various interpretations, and contextual walkthrough, on the other hand, can be used in such a situation that the users have already have time to test the system and assess its usefulness in their own work. Both the methods can be used in real use context to discover the problems and the strengths that the real use context brings to the use of the system.

REFERENCES

- [1] Barkhuus, L. and Rode, J. A. 2007. From mice to men – 24 years of evaluation in CHI. In: Extended Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. San Jose, CA, USA, April 28 – May 3, 2007. CHI'07. ACM Press, New York, NY. <http://www.itu.dk/~barkhuus/barkhuus-altchi.pdf>
- [2] Beyer, H. and Holtzblatt, K. 1998. Contextual design: Defining customer-centered systems. Morgan Kaufmann Publishers, Inc. Pp. 41–78.

- [3] Card, S. K., Moran, T. P. and Newell, A. 1983. *The psychology of human-computer interaction*. Lawrence Erlbaum Associates, Hillsdale, N.J.
- [4] Gomoll, K. 1990. Some techniques for observing users. In: *The art of human-computer interface design*. Laurel, B. (Ed.). Addison-Wesley Inc. Pp. 85–90.
- [5] Gould, J. D. 1996. How to design usable systems. In: *Readings in Human Computer Interaction: Towards the Year 2000*. Baecker, R., Grudin, J., Buxton, W. and Greenberg, S. (Eds.). Morgan-Kaufmann. Pp. 93–121.
- [6] Greenberg, S. and Buxton, B. 2008. Usability evaluation considered harmful (some of the time). In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Florence, Italy, April 5–10, 2008. CHI'08. ACM Press, New York, NY. Pp. 111–120. DOI = <http://doi.acm.org/10.1145/1357054.1357074>.
- [7] Hansen, M. 1991. Ten steps to usability testing. In: *Proceedings of the conference on 1991 ACM ninth annual international conference on systems documentation*. Pp. 135–139.
- [8] Holtzblatt, K. and Beyer, H. 1993. Making customer-centered design work for teams. *Communications of the ACM*, 36(10), 92–103. DOI = <http://doi.acm.org/10.1145/163430.164050>.
- [9] ISO 9241-11. 1998. Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.
- [10] ISO 13407. 1999. Human-centred design processes for interactive systems.
- [11] Kim, H., Kim, J., Lee, Y., Chae, M. and Choi, Y. 2002. An empirical study of the use contexts and usability problems in Mobile Internet. In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. HICSS'02. DOI = <http://doi.ieeecomputersociety.org/10.1109/HICSS.2002.994090>.
- [12] Koivunen, M., Nieminen, M. and Riihiahho, S. 1995. Launching the usability approach: experience at Helsinki University of Technology. *SIGCHI Bull.* 27(2), 54–60. DOI = <http://doi.acm.org/10.1145/202511.202526>.
- [13] Lepistö, A. and Ovaska, S. 2004. Usability evaluation involving participants with cognitive disabilities. In: *Proceedings of the Third Nordic Conference on Human-Computer interaction*. Tampere, Finland, October 23–27, 2004. NordiCHI'04, Vol. 82. ACM, New York, NY. Pp. 305–308. DOI = <http://doi.acm.org/10.1145/1028014.1028061>.
- [14] Lindgaard, G. and Chatratchart, J. 2007. Usability testing: What have we overlooked? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, April 28 – May 3, 2007. CHI'07. ACM Press, New York, NY, 1415–1424. DOI = <http://doi.acm.org/10.1145/1240624.1240839>.
- [15] Nielsen, J. 1993. *Usability Engineering*. Academic Press.
- [16] Nieminen, M. P. 1996. *Designing user interface concepts for multimedia services*. Master's thesis. Helsinki University of Technology.
- [17] Redish, J., Bias, R. G., Bailey, R., Molich, R., Dumas, J. and Spool, J. M. 2002. Usability in practice: Formative usability evaluations – Evolution and revolution. In: *Extended abstracts on Human Factors in Computing Systems*, Minneapolis, Minnesota, USA, April 20–25, 2002. CHI'02. ACM Press, New York, NY. Pp. 885–890. DOI = <http://doi.acm.org/10.1145/506443.506647>.
- [18] Riihiahho, S. 2000. *Experiences with usability evaluation methods*. Licentiate's thesis. Helsinki University of Technology. Available as a pdf file: http://www.soberit.hut.fi/~sri/Riihiahho_thesis.pdf.
- [19] Rohn, J. A., Spool, J., Ektare, M., Koyani, S., Muller, M. and Redish, J. 2002. Usability in practice: alternatives to formative evaluations – Evolution and revolution. In: *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. Minneapolis, Minnesota, USA, April 20–25, 2002. CHI'02. ACM, New York, NY. Pp. 891–897. DOI = <http://doi.acm.org/10.1145/506443.506648>.
- [20] Rosenbaum, S. 2000. Not Just a Hammer: When and how to employ multiple methods in usability programs. In: *Proceedings of the Usability Professionals' Association Conference 2000*, Ashville NC, August 2000. UPA2000. <http://www.teced.com/PDFs/Upa2000-sr.pdf>.
- [21] Savioja, P., Norros, L. and Salo, L. 2008. Evaluation of Systems Usability. In: *Proceedings of the European Conference on Cognitive Ergonomics*, Madeira, Portugal, September 16–19, 2008. ECCE'08.
- [22] Scholtz, J. and Consolvo, S. 2004. Toward a Framework for Evaluating Ubiquitous Computing Applications. *Pervasive Computing, IEEE*, 3(2), 82–88. DOI = <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1308574&isnumber=29040>.
- [23] Scott, K. M. 2009. Is usability obsolete? *Interactions*, 16(3), 6–11. DOI = <http://doi.acm.org/10.1145/1516016.1516018>.
- [24] Sengers, P. and Gaver, B. 2006. Staying open to interpretation: engaging multiple meanings in design and evaluation. In: *Proceedings of the 6th Conference on Designing interactive Systems*, University Park, PA, USA, June 26–28, 2006. DIS'06. ACM, New York, NY. Pp. 99–108. DOI = <http://doi.acm.org/10.1145/1142405.1142422>.
- [25] Wharton, C., Rieman, J., Lewis, C. and Polson, P. 1994. The cognitive walkthrough method: A practitioner's guide. In: *Usability inspection methods*. Nielsen, J. and Mack, R. L. (Eds.). John Wiley & Sons, Inc. Pp. 105–140.

What Explains Differences in Users' Inclination to Appropriate Technology for Unconventional Purposes? A Preliminary Analysis

Antti Salovaara¹, Sacha Helfenstein², Mikael Wahlström¹, Antti Oulasvirta¹

¹Helsinki Institute for Information technology
Helsinki University of Technology and
University of Helsinki

P.O. Box 9800, 02015 TKK, Finland

{antti.salovaara,mikael.wahlstrom,antti.oulasvirta}@hiit.fi

²Agora Human Technology Center
University of Jyväskylä
P.O. Box 35, 40014 Jyväskylä, Finland
sacha.helfenstein@jyu.fi

ABSTRACT

It is common to state that inventions of new purposes of use arise in social interaction with other technology users. Social aspects of appropriation have subsequently received a lot more attention than individual users' characteristics in appropriation research. To remedy this imbalance, this paper presents a preliminary analysis of a web survey that charted aspects of digital camera use and individuals' photography orientations and used them as predictors of digital camera appropriation. Gender, technology understanding and exchange of ideas with others proved tentatively the best predictors of appropriation.

Keywords

appropriation, web survey, user characteristics

ACM Classification Keywords

H.1.2 [Models and Principles] User/Machine Systems – *human factors*.

INTRODUCTION

During the past two decades, it has been increasingly recognized that information systems, like any tools, are not only used for purposes specifically designed for them, but that they are appropriated for purposes that can surprise their designers. The inventions of new purposes of use are called *appropriations*. The reasons for appropriations are both social and individual. The social processes contributing to appropriations have already attracted enthusiastic attention in the research community. The question who decides how technology should be used (e.g., at a workplace) has been addressed in science and technology studies [1] and management studies [5]. Another often-studied topic has been how appropriations are promoted and adopted by certain people at a workplace [3, 4].

However, an individual user's role in the invention process has been a lot less studied topic. In particular, research on the role of cognitive processes and individual user's characteristics has

been almost non-existent. In the few studies available it has been found that users with a learning-centered work orientation have been found more willing to learn new uses of technology even when it can be hard, than users with a performance-centered orientation who are willing to learn only if learning is easy [2]. In addition to the learning vs. performance orientation, in a mobile phone context, evidence has been found that users can also be segmented into different kinds of adopters whose inventions of use are different [7]. Cyclical perception-action models of appropriation processes have also been presented [6].

However, these studies have ignored the question *what individual characteristics are important for appropriation*. This paper addresses this question by focusing on one easily appropriable technology – digital cameras. In addition to compact and single-lens reflex (SLR) cameras, also mobile phones are nowadays often equipped with cameras that have a reasonable picture quality. The versatility to take photos with these technologies has provided a fruitful basis for ingenious uses, eight of which have been used as a measure of appropriation in this paper. The goal of this study was to identify user characteristics that could statistically predict such uses.

RESEARCH APPROACH AND MODEL

The research approach described in this paper is different from many of the previous studies on appropriation. Compared to the sociologically oriented studies in which the focus has been on charting the complexity of the phenomenon and aiming at a holistic picture, the approach here is focused, narrowly defined and aims for measurable results. Appropriation is here interpreted as *an invention of a new purpose of use, previously unknown to the user*.

Antecedent Factors of Appropriation

No theory has been presented that would attempt to list the possible antecedent individual user's characteristics that contribute to appropriation. Therefore, for the purpose of this study, a set of *tentative* antecedent constructs related to digital camera use were generated by researchers. These were:

- Setting up personal *Goals* for one's photography activities (e.g., personal projects or directions of improvement).
- *Reflection* of one's practices by evaluating one's shots.
- Having a comprehensive and correct *Mental Model* of how a camera works and what its functions are.
- *Curiosity* of trying new ways of photography.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

- Taking photos spontaneously and in *Ad Hoc* ways, in a spur of action, without always thinking before acting.
- Having a broad understanding of the surrounding technology *Ecology*; e.g. how photos can be edited or used in other media.
- *Awareness* that a digital camera is an easily appropriable tool and thus a potentially useful in many situations.
- *Social Construction*: learning new ways of use from others through teaching, observation or exchange of ideas. The purpose of this construct was to evaluate the importance of some of the previously studied factors of appropriation.

Because the constructs were not drawn from existing theories, the nature of this study was exploratory and proto-theoretical. It was also conceptually organized in a top-down manner. Each construct was divided into sub-constructs, and phrased as a statement that could be answered on a Likert scale (1 = totally disagree, 5 = totally agree). For instance the Mental Model construct was represented of the following sub-constructs:

- Learning camera's functions comprehensively ("I have familiarized myself with more or less all the functions in my camera or cameras").
- Understanding cause and effect ("I know how to tune the camera settings to capture photos with good quality").
- Knowing the good and bad aspects of one's cameras ("I know which are the most important strengths and weaknesses of my camera or cameras").

All in all, 35 Likert scale statements were generated to represent the eight constructs. The constructs and their wordings were improved iteratively by organizing two pilot studies. On the other hand, it was admitted that the set of constructs could not be complete, leaving a possibility that another set of constructs could be a better predictor of appropriation.

Measuring Appropriation

To evaluate the tentatively postulated antecedent constructs, a measure was needed for appropriation. To start with, the following eight uses were defined as signs of appropriation:

1. *Mirror*: pointing the camera toward oneself, in order to see e.g. how one's face looks like.
2. *Map*: taking a photo of a map, and using that photo as a replacement of a paper map.
3. *Note-taking device*: using the camera for note-taking when the content is very visual, e.g. when shopping clothes.
4. *Scanner*: storing printouts and texts as images with a camera.
5. *Memory card*: plugging the camera into a computer like an USB memory stick (does not work with all models equally).
6. *Lamp*: exploiting the camera as a light source.
7. *Instructing device*: using a sequence of photos to provide step-by-step instructions.
8. *Periscope*: inspecting places that are otherwise inaccessible to human vision.

This list of uses was also researcher-generated. During the two pilot studies, respondents were asked to provide their own suggestions for unconventional uses. However, the suggestions received could be subsumed in already listed uses, or their meanings were conceptually unclear. In the final questionnaire,

the set of uses was fixed to the eight uses listed above, to gather a homogenous dataset of answers.

The questions in the web survey about each appropriation were tree-structured. The starting question addressed the respondent's familiarity of using any digital camera in the given way (e.g., as a mirror). The scale ranged from 1 = "this use has never occurred to me before" and 2 = "I have know that this is possible, but I have never done so" to 6 = "This is one of my established uses for a digital camera". Later questions addressed the accuracy of the memory of the first time when the use was learned, and the actual person who did the invention in the situation.

From this eightfold set of tree-structured answers the measure for individual's appropriation was calculated using the familiarity variable. For each of the eight uses, two new dichotomized variables were created, one having value "0" if the respondent replied with an answer coded as 1 (see the values above), and "1" otherwise, the other having a value "0" if the respondent replied with an answer coded as 1–2, and "1" otherwise. Thus, one variable expressed whether a user was familiar with an appropriation or not, the other whether he or she had ever used it or not. Summing these binary variables over the eight uses yielded measures for the overall degree of appropriation (Total Degree of Appropriation Familiarity and Total Degree of Appropriation Employment), respectively, both ranging 0–8. The use of two measures was deemed important since the purely imagined use (i.e., answer coded as 2) was a common choice in the data, gathering on average 21% of all the familiarity answers. It was important to know whether its exclusion from the appropriation degree would change the results considerably.

THE STUDY

The Likert scale statements about the antecedent constructs as well as the tree-structured questions about different appropriations were part of a web survey in Finland between November 2008 and January 2009. The survey contained also other items, the most important ones from the point of this study being the demographic details (gender, age, education among others), camera use experience (expressed as years of film, digital and phone camera use) and camera use frequency (asked separately for each type of camera, ranging from daily to terminated use). Each respondent was also asked to assess whether she or he considered herself or himself as a beginner, novice, snapshot taker, amateur, expert/professional, or other kind of actor in photography. The questionnaire could be answered anonymously. Two pilot studies were organized before the actual survey.

Due to the mundaneness of a digital camera as a consumer technology, reaching a high number of responses was deemed more important than a strict probability sample of respondents. Invitations to participate were distributed to authors' social circles, camera-related web forums, and camera clubs. By buying keyword-based advertisements from a large Finnish search company, the survey was also visible at different pages in the web. Respondents could also invite their friends by providing their email addresses. These addresses were not archived in the database. A raffle of fifteen 20 EUR gift coupons was arranged between those respondents who had completed the survey.

The survey reached N = 2388 of complete answers from digital camera and/or phone camera users. The distribution of values in the self-reported expertise variable shows that the recruitment from camera clubs resulted in high number of answers from amateurs: novice (8.7%), snapshot taker (36.2%), amateur (47.9%), expert/professional (5.5%) and other (1.7%). Genders were equally represented in the data (males 53.3%).

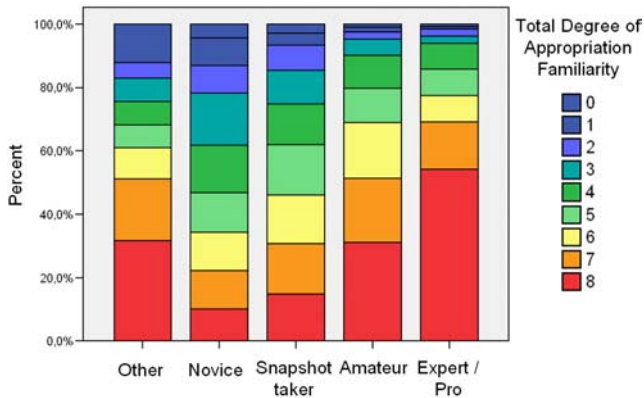


Figure 1. Distribution of how many of unconventional uses the respondents were familiar with, plotted separately by the different photography expertise levels.

Are Appropriations Invented Alone?

Two questions precede an analysis of individual users' characteristics: 1) whether the eight unconventional uses are rare enough to be informative of appropriation, and 2) whether they are learned individually and not only from others). In light of the data, the answer is clearly "yes" to both questions. Regarding the first question, as visible in Figure 1, the variance of values of Total Degree of Appropriation Familiarity is large, meaning that people exhibit appropriation to different extents.

To answer the second question, on average 54% of the respondents familiar with a use could remember the moment of learning "very well" or "partly". Among these people, inventing the use alone, without a help of others, was the most common context of invention (39%), learning from others being the next (21%). Appropriation by an individual is therefore common enough to warrant a study of its antecedent factors.

Scope of Analysis in This Paper

In the following analysis, the focus has been limited to snapshot takers only (N = . By this decision, the results are less susceptible to a possibility that active photographing turns out to be the underlying reason for appropriation. Such a biased result is less probable in a snapshot taker data. Due to this scope, the results are preliminary and will be extended in future work.

RESULTS

The reliability and validity of the antecedent constructs were evaluated for discriminant validity and internal consistency. For discriminant validity, a confirmatory factor analysis for the Likert statements was carried out. Based on the results, some of the initial constructs were combined into larger ones, and *Awareness* was dropped because of poor factor loadings. These changes yielded the following constructs for the actual analysis:

- Setting up personal *Goals* (unchanged).
- *Social construction* (one item dropped).
- *Technological knowledge* (Mental Model and Technology Ecology combined).
- *Exploration and learning*: (Curiosity with parts of Reflection and Ad Hoc photography style added): Users scoring high on this dimension display increased motivation to actively discover best use practices and learn from experiences.

- *Self-concept* (items in Reflection construct related to one's tendency to analyze oneself as a photographer): Users scoring high on this dimension display an increased sense of self-criticism regarding their photography and camera use skills.

To evaluate internal consistency of the constructs, Cronbach's alpha coefficients were calculated, attaining values between .76 (goals) and .89 (technological knowledge), which was interpreted as a strong support for the synthesized scales

How Appropriators and Non-Appropriators Differ

For each of the eight appropriations, analysis was carried out to see how users that (a) were familiar with an appropriation (i.e., whose dichotomized familiarity value was 1 or above) or (b) had ever employed an appropriation (i.e., familiarity value was 2 or above) differed from users that were not familiar (i.e., whose value was 0) or had never employed an appropriation (i.e., values 0–1), respectively. The intention was to reveal those basic user characteristics that can potentially distinguish appropriators from non-appropriators. U-tests, t-tests, and Chi-Square tests were used to assess the differences between those groups.

The analysis provided an overall proof that appropriators can be set apart from non-appropriators in basically all construct dimensions. In addition, people that were familiar with the eight uses also displayed differences regarding age, gender, as well as camera use history duration and use frequencies. In general, appropriators (i.e., those familiar with and using appropriations) had higher mean scores in all constructs, except Self-concept, which was associated to appropriation only for the Instructing device, Periscope, and Memory card use purposes.

Appropriation, as measured by familiarity level, was generally more common among men, except for the Mirror and the Lamp appropriations, both of which seem more general purpose or more valuable to women as well. Also, familiarity was on average more wide-spread among younger users.; however, not always in terms of the actual employment into use. In fact, users familiar with the Instructor appropriation proved on average older than others. Finally, appropriators seemed to be on average more active digital and phone camera users, and also had longer experiences with using digital camera devices.

Basic User Characteristics and Constructs as Predictors of Appropriation

In order to test for a statistical significance of user characteristics as predictors of the total degrees of appropriation (regarding both familiarity and employment), hierarchical, stepwise multivariate regression analyses were used. The purpose was to identify user characteristics that can well explain the overall variability in appropriation among users.

In a first block, the fundamental person variables (gender and age) were entered as regressors, followed by the five improved constructs (i.e., Goals, Social construction, Technological knowledge, Exploration and learning, and Self-concept), and finally by the remaining basic user characteristics (durations and frequencies of use for film, digital, and phone cameras). The research interest was to find out 1) which predictors are yielded as significant, and 2) how substantially each of the three regressor blocks increases the strength of the prediction model.

Since the analysis presented in the previous section had shown that basically all user characteristics were related to one or another appropriation, a decision was made to enter all variables into the initial regression model specification. For the Total Degree of Appropriation Familiarity, SPSS developed through

hierarchical stepwise regression a prediction model in five steps. The resulting predictors were given as gender, age, Technological knowledge, Social construction, and frequency of phone camera use. Calculating the regression with these predictors only, yielded a model with $R^2 = .24$, $F(5,691) = 44.02$, $p < .001$, with gender ($b = .23$, $t(691) = 6.49$, $p < .001$), and Representation ($b = .22$, $t(691) = 5.79$, $p < .001$) the greatest predictors.

Although the prediction model improved significantly (statistically speaking) with the addition of each of the three groups of user characteristics (i.e., person variables, constructs, use parameters), Technological knowledge and Social construction constructs (R^2 Change = .11, $F(2,692) = 47.78$, $p < .001$) had a substantial and clearly stronger effect on advancing the model strength than for instance use frequency (R^2 Change = .02, $F(1,691) = 19.14$, $p < .001$). And, when compared to gender and age in separate regression models, Technological knowledge ($R^2 = .18$) and Social construction ($R^2 = .11$) proved better predictors of users' Total Degree of Appropriation Familiarity.

For the Total Degree of Appropriation Employment, on the other hand, the regression analysis yielded a prediction model including gender, representation, social construction, phone camera use frequency, and duration of digital camera use history. The resulting model is described with $R^2 = .26$, $F(5,671) = 47.65$, $p = .000$, with representation as the strongest predictor, $b = .24$, $t(671) = 6.42$, $p < .001$. Again, a substantial predictive relevance was attested for the construct scales of Technological knowledge and social construction. In fact, the regression model strength appeared primarily attributable to these two constructs, R^2 Change = .14, $F(2,673) = 58.82$, $p < .001$, when compared to the model including gender, R^2 Change = .08, and use frequency and duration history, R^2 Change = .05.

DISCUSSION

The regression analyses point towards a tentative conclusion that Technological knowledge and Social construction are important factors contributing to appropriation, at least in a digital camera use context. That is, to generalize, acquiring a good understanding of how digital devices work, which functions they have, and on the other hand, exchanging ideas about their use with others, seem to be aspects that should be particularly supported when attempting to design easily appropriable technologies.

Many questions remain for future work, however. Why Exploration and learning was not rendered as significant predictor warrants future research, as well as whether the findings hold also among the more experienced users such as those who considered themselves amateurs or professionals.

ACKNOWLEDGMENTS

AS acknowledges financial support from UCIT Graduate School and Academy of Finland. SH acknowledges financial support from SoTech Platform project funded by the Technology Industries of Finland Centennial Foundation. The authors thank Lingyi Ma, Erkki Komulainen, and Andreas Dünser for ideas.

REFERENCES

- [1] Eglash, R. 2004. Appropriating technology: an introduction. In: *Appropriating Technology: Vernacular Science and Social Power*. Eglash, R., Croissant, J. L., Di Chiro, G. and Fouché, R. (Eds.). University of Minneapolis Press, Minneapolis, MN. Pp. vii–xxi.
- [2] Loraas, T. and Diaz, M. C. 2009. Learning new uses of technology: situational goal orientation matters. *Int J Hum-Comput St*, 67(1), 50–61.
- [3] Mackay, W. E. Patterns of sharing customizable software. In: *Proc. CSCW'90*. ACM Press, New York, NY. Pp. 209–221.
- [4] MacLean, A., Carter, K., Lövsstrand, L. and Moran, T. User-tailorable systems: pressing the issues with buttons. In: *Proc. CHI'90*. ACM Press, New York, NY. Pp. 175–182.
- [5] Orlikowski, W. J. 1992. The duality of technology: rethinking the concept of technology in organizations. *Organ Sci*, 3(2), 398–427.
- [6] Salovaara, A. 2008. Inventing new uses for tools: a cognitive foundation for studies on appropriation. *Human Technology*, 4(2), 209–228.
- [7] Wirth, W., von Pape, T. and Karnowski, V. 2008. An integrative model of mobile phone appropriation. *J Comput-Mediat Comm*, 13(3), 593–561.

Developing Practical Tools for User Experience Evaluation – A Case from Mobile News Journalism

Heli Väättäjä¹
¹Tampere University of
Technology
Korkeakoulunkatu 1
33720 Tampere
heli.vaataja@tut.fi

Tiina Koponen¹
¹Tampere University of
Technology
Korkeakoulunkatu 1
33720 Tampere
tiina.koponen@tut.fi

Virpi Roto^{1, 2}
²Nokia Research Center
P.O. Box 407
00045 Nokia Group, Finland
virpi.roto@nokia.com

ABSTRACT

We present a questionnaire called Attrak-Work to support the evaluation of user experience of mobile systems in the context of mobile news journalism. We discuss theoretical background of the questionnaire and describe the development process including the field study within which the questionnaire was developed. The presented questionnaire assesses user's perception of the pragmatic (usability and task and goal achievement) and hedonic (stimulation and identification) qualities and an overall judgment of appeal. We used the questionnaire as part of a field study to corroborate and expand the findings of observations and interviews. We found the Attrak-Work questionnaire a useful tool to be used in this manner especially for the evaluation of the hedonic qualities.

Keywords

user experience, evaluation, work, mobile phone, journalism

ACM Classification Keywords

H.5.2 [User Interfaces]: Theory and methods, Evaluation/methodology; H.5.1 [Multimedia Information Systems]: Evaluation/methodology.

General Terms

Measurement, Theory, Human Factors, Design.

INTRODUCTION

There are many well-established ways to evaluate the usability of interactive systems including questionnaires (e.g. [17]), but evaluating experiential aspects such as fun, meaning, or beauty is a much less covered topic. Focusing on user experience and evaluating the experiential aspects helps for example maturing industry sectors to differentiate from competition and gain a loyal customer base. Without means to evaluate user experience, it is impossible to manage experience related aspects. As the need for systematic user experience evaluation is high both in

industry and academia, user experience evaluation has gained increasing attention in Human-Computer Interaction (HCI) [6], [9], [11].

It is still unclear what the appropriate methods and metrics are for assessing user experience. This is partly due to the fact that there is still not an agreed definition for user experience, although standardization work is ongoing. There are for example methods for assessing person's momentary emotions or emotional trajectories [1] during interaction, which provide interesting information for content developers such as game narration designers or movie directors.

Not all products are designed to trigger a specific emotion, however, but to provide valuable and meaningful experiences in a broader sense [2]. For example, a mobile journalist might aim at generating a certain story experience for the audience, but the text and image capturing and editing tools that s/he uses are not targeted to raise specific emotions. A work tool is often primarily seen as a means to an end, having instrumental value. Not surprisingly, the evaluation of user experience with questionnaires in mobile work context has concentrated mainly on usability aspects with a few exceptions reaching beyond it [10], [15], [21].

The field work of news journalists and photographers has always been highly mobile. Advances in mobile technology, with converged devices, interoperability and fast mobile and broadband network connections enables journalists and photographers to use mobile tools for news reporting from the field. These tools can be used for capturing of photos or videos, creating stories, and submitting or even publishing them directly from the field. Instead of a van full of equipment, light tools such as a laptop or even a mobile phone can be used for producing the stories. Journalists and photographers do not necessarily return to the newsroom to deliver their stories but for example email them to the newsroom or even publish them right from the field [20], [22]. The ad-hoc and timely nature of mobile reporting brings a new flavor to journalism.

We conducted a field study with a mixed methods research design [18] to explore user experience with a mobile journalism system. Nineteen participants used a multipart mobile system based on a mobile multimedia phone for submitting and publishing news items to an online publication on two project days. One of the goals in the study was to create a questionnaire for assessment of users' perceptions of mobile system qualities and overall judgment of mobile systems used in the context of mobile news journalism. Our intention was to use the questionnaire for corroboration and expansion of the results from the qualitative data and to build a tool to be used in the evaluation of user experience in our future studies in the context of mobile news journalism.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

We chose Hassenzahl's model of user experience [5] as a starting point for the questionnaire development. We started by analyzing the collected observation and interview data with Hassenzahl's model as a guiding theory. Based on the findings we refined the model and developed the Attrak-Work questionnaire. Questionnaire was conducted at the end of the field study as one part of an online survey targeted at the participants of the study.

In this paper we present the theoretical background and the phases of the questionnaire development as well as exemplary results of using the Attrak-Work questionnaire. We also discuss critically many of the limitations in the development of the Attrak-Work questionnaire, which reflect the challenges of fitting questionnaire development into a relatively short period of time.

THEORETICAL BACKGROUND

Questionnaires are frequently used in various types of user studies in HCI. A variety of questionnaires have been developed for evaluating users' emotions. The affect grid [14] assesses emotional states with a 9 x 9-matrix that is surrounded by eight adjectives describing different emotions. A Semantic Differential Scale [13] is a type of a scale with which users can rate the system based on bipolar word couples. For example Mehrabian and Russell [12] have used the Semantic Differential Scale with 18 adjective pairs for emotion assessment against valence, arousal and dominance. Instead of words pictures have been used in emotion assessment questionnaires to avoid language difficulties [3], [8].

Emotion assessment is not the only way to evaluate user experience. When we aim to improve a system or want to gain an understanding of the user experience, we are also interested in users' perceptions of the product's qualities and their overall evaluative judgments of it. To be able to reach beyond studying the instrumental aspects, practical tools that support the assessment of user experience are needed. One approach is to include hedonic aspects in the measurement, like in the HED/UT scale [16], [19] or the AttrakDiff questionnaires [6], [7]. These questionnaires aim to assess users' perceptions of the product or system qualities.

AttrakDiff questionnaires are based on the user experience model presented by Hassenzahl [5], which is illustrated in Figure 1. We chose this model as the basis for the development of the tool for user experience evaluation in the context of mobile news journalism. Hassenzahl's model enables a relevant approach to studying aspects of user experience in work context, since the theory covers not only pragmatic (utilitarian) aspects, but also hedonic (non-utilitarian) aspects. For professionals not only the functional aspects of the used technology are important, but also how it relates to being stimulating, supports and enables creativity and, on the other hand, what kind of symbolic value it possesses.

Hassenzahl's framework is based on the assumption that product character can be described by two attribute groups, namely pragmatic and hedonic attributes [5]. Each person constructs her own personal version of the product character based on the product features and on her personal standards and expectations. *Pragmatic quality* is *instrumental* and related to the product's *usability and utility* when the product is used for tasks. On the contrary, *hedonic quality* is related to the user's self, such as autonomy, competence, relatedness to others, or security [4], [5].

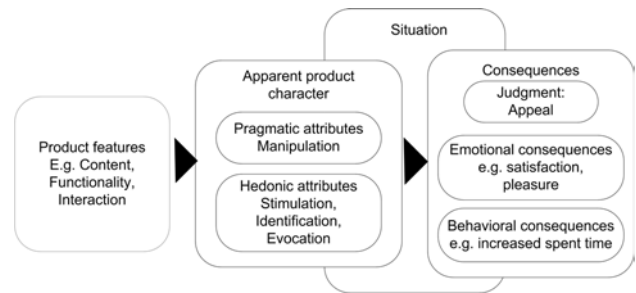


Figure 1. Key elements of Hassenzahl's model of user experience from the user perspective. Source: [5]

Hedonic quality focuses on aspects of stimulation, identification, and evocation [5]. *Stimulation* is related to personal development, that is, to curiosity, personal growth, development of skills and proliferation of knowledge. *Identification* addresses the expression of self and the user's personal values to relevant others through objects and is therefore social. An example of this in the context of journalism is a photographer's systems camera and the big camera case(s) he carries with him, which serve as symbols of his profession. *Evocation* refers to the product's ability to provoke memories such as important past events or relationships.

According to Hassenzahl, the subjective perception of the product character leads to consequences such as judgments about the product's appeal, goodness and beauty [4], [5], as well as emotional and behavioral consequences. As examples of emotional consequences Hassenzahl discusses satisfaction and pleasure [4], [5]. Based on the model, Hassenzahl presents two versions of AttrakDiff questionnaires, for assessing the attractiveness of products [6], [7]. The first version, AttrakDiff includes two attribute groups, that is, one group for pragmatic and another for hedonic, as well as one group for the judgment of appeal [7]. The second version, AttrakDiff2, separates the hedonic attribute group into two groups, one for stimulation and the other for identification [4], [6]. In addition, evaluative constructs such as goodness and beauty have been included in subsequent studies [4]. AttrakDiff questionnaires use a Semantic Differential Scale to assess the pragmatic and hedonic attributes as well as items in judgment of appeal and the evaluative constructs.

Although AttrakDiff questionnaires have been used by several researchers in studying user experience, we decided to use the original model as a basis for questionnaire development instead of using the AttrakDiff questionnaires. This decision was made, since when we used Hassenzahl's model as a guiding theory in analysis of the data, our findings on hedonic aspects differed considerably in their representation from the attributes in AttrakDiff questionnaires. Therefore, we used our findings in the development of the questionnaire. In the following sections, we first present the study and continue by describing the development of the questionnaire within it.

STUDY

We used the case study approach [23], which was carried out with a mixed methods research design [18]. Data was collected during a field trial by qualitative (semi-structured interviews, observations) and quantitative (questionnaires) methods. Questionnaires were used in the research design to corroborate and expand the results obtained from observations and interviews. Data was collected before, during, and after the usage of mobile system.

Figure 2 illustrates the data collection methods used in different phases of the study.

Pre-usage interviews	Pre-usage questionnaire	Observations - planning sessions - project days - wrap up sessions	Post-usage interviews	Post-usage questionnaire
----------------------	-------------------------	---	-----------------------	--------------------------

Figure 2. Data collection of the study.

Study was made in conjunction with a graduate level university course on web publishing in the Department of Journalism and Mass Communication at the University of Tampere, Finland. Data collection centered around two project days, when the students produced short news stories and videos for a web publication using a mobile journalism system based on mobile multimedia phones. The mobile system consisted of a mobile multimedia phone (Nokia N82), a wireless Bluetooth keyboard (Nokia SU-8W) and a mature prototype of a mobile journalism software application running on the mobile phone. Application enabled the creation of news stories with text and multimedia items (photos, audio and video clips) on a mobile phone and submitting of these stories to the publication platform.

During the two project days graduate students worked as journalists and photographers creating news stories to the publication from the field. University course was chosen as the context of the field study, since the mobile system was a prototype. It was not feasible to set up a trial in a news organization at this phase, since potential problems encountered may had disturbed the work of the professionals considerably. Researchers did not influence the decisions on the type of stories or how or for what purposes the participants used the system. The publishing process therefore was similar to the one which is used in real news organizations with editorial meetings (here planning and wrap up sessions) and with the roles of editorial staff (online producer and art director) included. See for example [20], [22] for more details on the study.

Participants

Participants of the study were graduate students of journalism and visual journalism who were taking a project course on web publishing. All students had practical experience in journalistic work either full time (1–15 years, median: 1 year) or part-time (0–4 years, median: 2 years). Most of them were working as freelancers in parallel to their studies. The students of visual journalism had used mobile phones (Nokia N93) earlier in their studies for video capturing and editing, whereas the students of journalism had no prior experiences of using mobile phones in their studies. From here on we refer to the students on journalism as journalists and to the students of visual journalism as photographers.

The number of participants in the study was nineteen (10 journalists, 9 photographers). The number of respondents to the online survey, which included the Attrak-Work questionnaire, was fifteen (8 journalists, 7 photographers). All interviewees and respondents were given as a compensation a ticket to movies.

Data Collection

As can be seen from Figure 2, data was collected using multiple methods at various points during the study. The study centered around two separate project days in the spring 2008, which were five weeks apart. Pre-usage interviews were arranged three days before the training of the mobile journalism system for two groups separately, one for two students of visual journalism and one for three students of journalism. The goal

was to familiarize with the field of the study, participants’ usage of mobile phones, prior experiences of using mobile phones in news journalism, expectations as well as attitudes. Results were used in the development of a pre-usage questionnaire, which was conducted as a paper questionnaire right before the training. All 19 students answered the questionnaire which was used to collect background information of the users, their prior usage and experiences of mobile devices and services, and expectations and attitudes towards upcoming project.

Observations were made on two project days, on two planning sessions a week before each project day, and on two wrap-up sessions a week after each project day. Three researchers were involved in the field trial during the first part of the study and four researchers in the second part of the study. A total of 85 hours of observations were made both in the mobile context (journalists and photographers working) as well as in the “newsroom”. Researchers made hand-written notes and took photographs when observing the work. Within three days of the first project day eleven participants and during the second project day four participants were interviewed. Interviews were semi-structured lasting from 60 to 90 minutes. All interviews were recorded. Interviews covered various user experience related themes as well as users’ perceptions on the suitability of mobile phones in journalism.

The post-usage questionnaire was conducted as web survey with a deadline for completing it within ten days of the second project day. Participants of the study were sent an email asking to complete the survey two days after the second project day. A reminder was sent by email two days before the deadline. The questionnaire consisted of several parts, of which one was the Attrak-Work questionnaire which began from the second question in the survey.

Analysis

The data from the observations was written into electrical form for further analysis and the interviews were transcribed. Interview and observation data were analyzed by content analysis. Specifically for the case of developing the questionnaire, data analysis was guided by theory to identify themes and attributes related to pragmatic and hedonic aspects (stimulation and identification) of the mobile system usage. For this article we used Cronbach’s alfa for testing the internal consistency reliability of the scales and nonparametric Mann-Whitney U test for identifying statistically significant differences between the user groups on their perceptions of the system’s qualities and overall judgment of appeal.

DESIGNING THE QUESTIONNAIRE

The process of developing the user experience questionnaire for mobile news journalism included several phases: first, gaining an understanding of the factors affecting user experience with exploratory, qualitative methods. We used Hassenzahl’s model of pragmatic and hedonic product qualities as a guiding theory in the analysis phase. We then built a framework for the instrument development from our findings and based on earlier theories [5], [11]. Finally, we developed the questionnaire based on our framework and on the findings from qualitative data. This section gives an overview of these phases.

Findings from the Qualitative Data

In this sub-section we present a short overview of the findings related to the pragmatic and hedonic aspects of using the system based on the observations and interviews.

Pragmatic Quality

Themes that were emphasized by several participants regarding pragmatic aspects of the mobile journalism system were for example ease of use, learnability, reliability, intuitiveness of use, performance and effectiveness. Support for the task, work process-related themes such as effect on working, and, in particular, reaching higher level goals of news journalism were discussed. On the mobile system level, these themes addressed the features and functionalities of the used mobile system as an entity or its sub-components, such as the keypad, camera or the mobile journalism application. Users described the usability-related aspects for example with words like *easy, intuitive, cumbersome, unreliable and fluent*. Furthermore, the themes related to carrying out the tasks or achieving goals covered for example effect on working and on the speed of publishing, support for working and the efficiency of the system.

Hedonic Quality – Stimulation

In addition to pragmatic aspects, participants mentioned several aspects related to the hedonic qualities of the system and its usage. Participants described the usage of the system as *interesting, (un)motivating, spontaneous, liberating, enchaining, exciting, frustrating and restricting*. These aspects were clearly related to the user's own self and his or her experience of using the system for capturing the material and for making the publication. Journalists took a very practical stand to using the mobile system, whereas photographers were more negative and reserved towards the system. Participants also emphasized that technology is essential for photographers to do their job. Due to the limitations in the technical capabilities of the mobile phone, photographers expressed that it restricts or even enchains their expression and creativity, and they found it non-motivating not to be able to achieve what could be achieved with "proper" tools. On the other hand, some photographers commented that using a simple device with limited capabilities was also in some sense liberating for them. However, both journalists and photographers expressed that using the mobile phone for capturing videos gave them new possibilities for news making and it was therefore found interesting and motivating for the specific purpose.

Hedonic Quality – Identification

Themes of hedonic identification that were emphasized in this study were related to communicating profession and status. For photographers, the systems camera, besides being a practical means and an important enabler of their job, is a symbol of their profession in a social context. It *communicates professionalism* both to the interviewees and other outsiders, including other professionals. Photographers and journalists also talked about the *reactions of outsiders* to using the mobile phone for multimedia capture. Participants described the reactions of outsiders to vary from neutral to surprise and disbelief. References to outsiders addressed two different groups, that is, the people they were interviewing and shooting photos and videos of, and other outsiders, either ordinary people or other professionals that were present in the usage situation.

When using mobile phones, in this study both photographers and journalists expressed that interviewees who were laymen and may had never been interviewed or photographed before by the media were more at ease with the small and everyday like device than with a systems camera. Participants felt that interviewees were also less reluctant to be interviewed and photographed. Participants therefore *reflected on outsiders' reactions and comments* as well as *anticipated reactions and attitudes towards mobile phone users*.

Evaluative Judgments – Appeal

As exemplified above, we found both pragmatic (utilitarian) and hedonic (non-utilitarian) themes and qualities related to the use of the studied mobile journalism system in mobile news journalism. The perceptions of these qualities are subjective, and they are related to a person's overall judgments of the mobile system. Based on the findings from qualitative data, there is a difference between the two user groups of the study, journalists and photographers, regarding the perceived hedonic qualities and appeal. Based on the observation and interview data, journalists were more positive than photographers towards the mobile system. On the individual level, there are, however, large differences in users' perceptions and overall judgments.

Description of the Attrak-Work Questionnaire

As described in Section: "Theoretical Background", we chose Hassenzahl's model as the starting point in the development of the questionnaire. We first created a model from our findings and earlier theories for the questionnaire development. The created model is presented in Figure 3, and it presents two groups of user perceptions of product characteristics, that is, the perceptions of pragmatic and hedonic qualities. Mahlke [11] refers to these as components of user experience and uses the terms utilitarian and non-utilitarian instead of pragmatic and hedonic. In our model, the user's perceptions of the pragmatic and hedonic qualities affect the overall judgment of the system, which in the Attrak-Work questionnaire is measured as appeal [7].

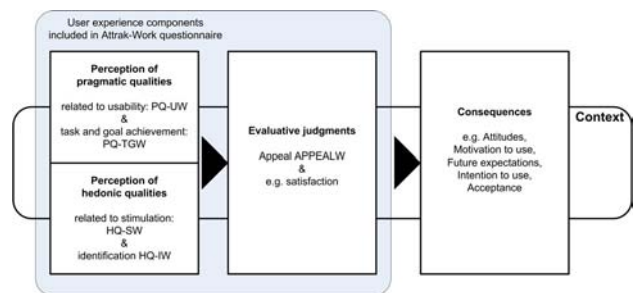


Figure 3. Model of user experience components in the development of the Attrak-Work questionnaire.

The evaluative judgments of the system are separated from the other consequences since judgments are related to the used system directly. As discussed earlier, in our view the perceptions of the pragmatic and hedonic qualities, the evaluative judgments of the system as well as the other consequences are context-dependent and relative to the usage situations. Contextual dimensions [22], their elements and the actual usage situation affect the user's perception of the pragmatic and hedonic qualities and his or her overall evaluation of the used system.

The pragmatic attribute group in our elaborated questionnaire covers usability, and the hedonic attribute groups cover stimulation and identification. We also included a second group of pragmatic attributes related to task and goal achievement, since this is an important aspect affecting the user's judgment when the system is used as a work tool. Appeal was included as a fifth theme for assessing an overall evaluative judgment of the studied mobile system. We selected the Semantic Differential scale for assessing a rating for attributes. Each of the attribute groups contain seven or eight pairs of words or short statements (items) presenting opposites of qualities on a bipolar scale. We used a five anchor scale for the

rating of the items, ranging from -2 to 2 when we implemented the questionnaire as part of the online survey.

It should be noted that the presented model is not intended to be a comprehensive model of user experience including all the aspects related to the phenomenon. It is a simplified model including components we used in the Attrak-Work questionnaire for measuring components of user experience. We did not for example include emotions in this model, although they could be included. However, the model presented in Figure 3 includes examples of themes related to the consequences of user experience that were found in our study and that have also been discussed in earlier literature.

In the following sub-sections we describe each attribute group in the Attrak-Work questionnaire. All attribute groups have been modified based on our findings from the observation and interview data. They therefore differ from the original items presented in AttrakDiff questionnaires, but also some of the original items are directly included in the Attrak-Work questionnaire. The items were created based on our findings, and they reflect how participants talked about the system and how they described its usage.

Pragmatic Quality – Usability (PQ-UW)

As a basis for the users' assessment of the pragmatic quality of the mobile journalism system we used the attribute group used in AttrakDiff questionnaires [6], [7]. There are altogether seven items in this group, of which two are directly from the original AttrakDiff (PQ-UW-1=PQ_2, PQ-UW-4=PQ_6). In addition, two items are related but not completely identical to the items in AttrakDiff (PQ-UW-2~PQ_7, PQ-UW-3~PQ_4). The selected items, modifications and new items reflect the findings from the qualitative data. For example, a new item on reliability was included, since reliability was strongly emphasized by the participants as one basic usability-related aspect that was essentially important in the work context.

- PQ-UW-1 Monimutkainen-Yksinkertainen,
Complicated-Simple
- PQ-UW-2 Vaikea-Helppo, *Difficult-Easy*
- PQ-UW-3 Hankala-Vaivaton, *Challenging-Effortless*
- PQ-UW-4 Hämmentävä-Selkeä, *Confusing-Clear*
- PQ-UW-5 Epälooginen-Looginen, *Illogical-Logical*
- PQ-UW-6 Epäluotettava-Luotettava, *Unreliable-Reliable*
- PQ-UW-7 Arvailua vaativa-Intuitiivinen,
Needs guessing-Intuitive

Task and Goal Achievement (PQ-TGW)

We created a separate attribute group related to task and goal achievement, since this is an important aspect affecting the appraisal of the system in the work context. Whereas the first pragmatic attribute group is related to usability, this second pragmatic attribute group concentrates on the effect and support of the product or system on working. The items in this group were created based on the themes that were found in the qualitative data.

- PQ-TGW-1 Työskentelyä hankaloitava-työskentelyä
helpottava, *Makes work harder-makes work
easier*
- PQ-TGW-2 Tehoton-Tehokas, *Inefficient-Efficient*
- PQ-TGW-3 Kompromisseihin pakottava- Tavoitteita tukeva,
Forces compromise-Supports goals

- PQ-TGW-4 Hidastaa julkaisua kentältä-Nopeuttaa julkaisua
kentältä, *Speeds up publishing from the field-
Slows down publishing from the field*
- PQ-TGW-5 Laattaa alentava-Laattaa edistävä,
Lowers quality-Enhances quality
- PQ-TGW-6 Työprosessia estävä- Työprosessia tukeva,
Obstructs the workflow-Supports the workflow
- PQ-TGW-7 Työskentelyä hidastava-Työskentelyä
nopeuttava, *Speeds up work-Slows down work*

Hedonic Quality – Stimulation (HQ-SW)

For assessing stimulation as an aspect of the hedonic quality of the system, we used the presented model as the starting point for developing the items. In our study, participants described several different types of stimulation-related aspects, as described in Section: "Findings from the Qualitative Data" that seemed relevant to be assessed with a questionnaire. However, these themes were not covered in the AttrakDiff 2 questionnaire, and we therefore redesigned the items to fit the context of the study. The seven created items are presented in the following:

- HQ-SW-1 Kahlitseva-Inspiroiva, *Restricting-Inspiring*
- HQ-SW-2 Turhauttava-Innostava, *Frustrating-Exciting*
- HQ-SW-3 Lannistava-Motivoiva, *Discouraging-Motivating*
- HQ-SW-4 Oppimista estävä-Oppimista stimuloiva,
Stimulates learning-Prevents learning
- HQ-SW-5 Luovuutta rajoittava-Luovuuden mahdollistava,
Limits creativity-Enables creativity
- HQ-SW-6 Kehittymistä rajoittava- Haasteita tarjoava,
Restricts development-Offers challenges
- HQ-SW-7 Ammatillista kunnianhimoa rajoittava-
Ammatillisen kunnianhimon mahdollistava,
*Constricts professional ambition- Enables
professional ambition*

Hedonic Quality – Identification (HQ-IW)

Similarly to the case of stimulation, we used the presented model as a theoretical background when developing the items for identification as the second group of hedonic quality. During the study, participants expressed various issues related to this theme. One item is identical to an item in the AttrakDiff2 questionnaire (HQ-IW-1=HQI-3). The other items have been created based on the themes found in the qualitative data.

- HQ-IW-1 Harrastelijamainen-Professionaalinen,
Professional-Amateurish
- HQ-IW-2 Epäuskottava-Uskottava, *Unconvincing-Credible*
- HQ-IW-3 Luottamusta vähentävä-Luottamusta herättävä,
Raises trust-Lowers trust
- HQ-IW-4 Haastateltavia tai kuvattavia epäilyttävä-
Haastateltavalle tai kuvattavalle kynnystä alentava,
*Increases suspicion in interviewees-Lowers the
threshold of interviewees*
- HQ-IW-5 Ammatillista imagoa laskeva-Ammatillista imagoa
kohottava,
*Lowers professional image-Promotes professional
image*
- HQ-IW-6 Ammattilaisten silmissä väheksyttävä-
Ammattilaisten silmissä arvostettu,
*Undervalued by professionals-Valued by
professionals*

HQ-IW-7 Vähentää työn arvostusta-Lisää työn arvostusta,
Lowers respect for the work-Enhances respect for the work

Appeal

Attribute group APPEALW in the Attrak-Work questionnaire comprises of aspects discussed by the participants that were related to the overall appeal of the system. Appeal is included also in the AttrakDiff questionnaire (2000), comprising of eight items. We included three of these items in the Attrak-Work questionnaire (APPEALW-1=APPEAL1, APPEALW-3=APPEAL2, APPEALW-4=APPEAL5). In addition, we included from the group of pragmatic quality in the AttrakDiff2 questionnaire the item of practicality PQ_3 (APPEALW-8), since in our view this is more an overall evaluation related to appeal. The other new items reflect the findings in the field study data as well. The eight items are as follows:

- APPEALW-1 Epämieluisa-Mieluisa, *Unpleasant-Pleasant*
- APPEALW-2 Yhdentekevä-Tärkeä, *Insignificant-Important*
- APPEALW-3 Huono-Hyvä, *Bad-Good*
- APPEALW-4 Vastenmielinen-Houkutteleva,
Unattractive- Attractive
- APPEALW-5 Vakava-Rento, *Serious-Relaxed*
- APPEALW-6 Tylsä-Kiinnostava, *Dull-Interesting*
- APPEALW-7 Hyödytön Hyödyllinen, *Useless- Useful*
- APPEALW-8 Epäkäytännöllinen-Käytännöllinen,
Impractical-Practical

Issues on Developing and Administering the Attrak-Work Questionnaire

The questionnaire was pre-tested by one researcher at two occasions during the questionnaire development. She was involved in the collection and analysis of the qualitative data, but she was not involved in the development of the questionnaire items. She was asked to point out items that were not clear, if they were not reflecting the findings, if the anchors for an item were not appropriate or if there were duplicates. The final questionnaire was pre-tested similarly by two other researchers, who were not involved in the study. Due to the tight schedule of the ongoing field study, which involved data collection, transcribing and analysis before creating the items for the Attrak-Work questionnaire, we were not able to involve outsiders, such as the participants of the study, or other professionals working in news journalism in the pre-testing.

For the questionnaire scale we selected the Semantic Differential Scale for assessing the participants' ratings on the items (attributes). We used in this study a five-point scale ranging from -2 to 2. The Semantic Differential Scale is known to be sensitive to selection of the anchors. Selecting a "wrong" pair has an effect on the responses and the reliability of the results. As described earlier, by pretesting we aimed to find these problems. However, we consider developing a second version of the questionnaire, which uses a Likert-scale instead of the semantic differential scale. Regarding the administration of the questionnaire as an online survey, each group of attributes was evaluated separately in the order presented in Section: "Description of the Attrak-Work Questionnaire". In the future studies we will group the pragmatic qualities into one group and mix the items, and do the same for the hedonic items as well.

EXEMPLARY RESULTS FROM ATTRAK-WORK QUESTIONNAIRE

In this section, we present and discuss an example of the results when the Attrak-Work questionnaire was used within our study. Our findings from the qualitative data indicate clearly, that the photographers perceived the mobile system more negatively than the journalists especially regarding the hedonic qualities. We were therefore interested in whether we could find a statistically significant difference in the perceptions of the pragmatic and hedonic qualities and overall judgment between the user groups by using the responses collected with the Attrak-Work questionnaire. However, we want to stress, that even if we cannot find a statistically significant evidence for an emerging theme or finding in the qualitative data, it does not mean that it is not important or it does not exist in real-life. Care must therefore be taken when interpreting the results and not to overweight the meaning of the questionnaire results in comparison to the qualitative data, which in this case study has the main emphasis.

For getting an overall evaluation of the mobile journalism system, we used the five attribute groups (PQ-UW, PQ-TGW, HQ-SW, HQ-IW, APPEALW) as scales. First, to gain a scale value for the perceived qualities and appeal for each respondent, we calculated the mean of the ratings for items (attributes) within an attribute group. We then tested the internal consistency reliability of the scales with Cronbach's alpha for all five attribute groups. We selected the items for the scales based on the corrected item-total correlation values (≥ 0.3) and Cronbach's alpha value (> 0.7). The scales PQ-TGW ($\alpha = 0.886$), HQ-SW ($\alpha = 0.870$), and HQ-IW ($\alpha = 0.845$) include the seven original items presented in the previous section. For scale PQ-UW we removed two of the original seven items, that is, PQ-UW-3 and PQ-UW-6 ($\alpha = 0.809$). In addition, for scale APPEALW we removed two of the eight original items, that is, APPEALW-2 and APPEALW-6 ($\alpha = 0.819$). We then recalculated the scale values.

To test if there is a statistically significant difference between the perceptions of the journalists and photographers on the perceived qualities and appeal we used the non-parametric Mann-Whitney U test. We first calculated the arithmetic mean of the scale values for both user groups. We then calculated the Mann-Whitney U test using the professional role (user group) as a grouping variable. The results of the significance test for each scale are shown in Table 1. The results show that for perceived hedonic quality of identification (HQ-IW), we found statistically significant difference between the user groups ($U = 8.5$, $p < 0.05$). For perceived pragmatic qualities PQ-UW and PQ-TGW as well as for overall judgment APPEALW, we did not find statistically significant difference. For perceived hedonic quality of stimulation HQ-SW ($U = 11.5$, $p < 0.1$), we found a trend, but it cannot be interpreted in strict statistical sense showing significance.

Table 1. Results of the Mann-Whitney U test for scales with professional role as a grouping variable.

	PQ-UW	PQ-TGW	HQ-SW	HQ-IW	APPEALW
Mann-Whitney U	22.0	18.0	11.5	8.5	13.5
Exact. Sig.	.536 ^a	.281 ^a	.054 ^a	.021 ^a	.189 ^a

a. [2*(1-tailed Sig.)], not corrected for ties.

DISCUSSION

Developing a questionnaire is tricky. Phases of the development include for example selecting or developing a framework or theory or using earlier findings as the basis for development, operationalizing the chosen theory and concepts, preparing the questions, selecting an appropriate scale and pretesting the questionnaire. In addition, when developing a questionnaire one has to consider carefully the goal of the development and consider the context it is intended for.

The primary goal of the questionnaire development in our study was to develop a practical tool for assessment of user experience in a work context, specifically in mobile news journalism. We developed the questionnaire to corroborate and expand the findings from observation and interview data in a field study. In addition we aimed to develop a tool for our future evaluations of mobile systems in the context of mobile news journalism. In the field study graduate level students of journalism and visual journalism used a multipart mobile system during two project days to publish an online publication directly from the field. The system consisted of a mobile multimedia phone, a wireless Bluetooth keyboard and a mobile application developed for creating stories and submitting or publishing them directly.

When developing the Attrak-Work questionnaire, we used Hassenzahl's model of user experience (see Figure 1, [5]) as a guiding theory in the analysis of the observation and interview data, but also looked for other related themes. Based on the findings we refined Hassenzahl's model by including task and goal achievement as a second attribute group for the assessment of the pragmatic qualities. The created model (see Figure 3) separates the evaluative judgments of the product, such as appeal and satisfaction from the other consequences, such as acceptance, motivation to use and intention to use. The findings from the qualitative data were used in the development of the items for the five groups of attributes, that is, for 1) pragmatic quality – usability PQ-UW, 2) pragmatic quality – task and goal achievement PQ-TGW, 3) hedonic quality – stimulation HQ-SW, 4) hedonic quality – identification HQ-IW and 5) for overall judgment of appeal APPEALW.

The developed Attrak-Work questionnaire was used at the end of the field study to assess the perceived pragmatic and hedonic qualities and the overall judgment of appeal of the used mobile journalism system. Attrak-Work questionnaire was administered as part of an online survey targeted to the participants of the field study. The qualitative data shows a clear difference in the perceptions of the hedonic qualities related to stimulation and identification as well as appeal between the photographers and journalists in the study. As an exemplary result for using of the developed Attrak-Work questionnaire we found statistically significant difference for the hedonic quality identification HQ-IW ($U = 8.5, p < 0.05$).

We found the Attrak-Work questionnaire to be a useful tool in an exploratory study of user experience for several reasons. First, since it is often not possible in a field study to observe or even interview all the participants, it provides a way of accessing the perceptions of a wider set of participants. Second, using a questionnaire means that all the respondents answer the same questions, and therefore we are able to get views on each item from all respondents. Therefore, themes that may not come up in the majority of observations or interviews can still be included to the study. However, one improvement related to this could be to ask the respondents to weight or order the importance of the attributes to find out how important the less

discussed themes are compared to the ones that are discussed more. Third, we can use the tool in our further studies in the same field and also for example compare results from different case studies with different participants.

There are several limitations in the development and testing of the questionnaire. Pre-testing was done with researchers, since the study was fast-paced and there was no time to involve outsiders into the development process. The use of an expert panel consisting of outsiders who preferably work as photographers and journalists would give feedback on the wording of the items and help in finding the correct anchors for each item therefore improving the validity of the items. In addition, using external experts working in news journalism would give feedback from a wider audience and improve the validity of items. A recommendable way of obtaining feedback would also be to involve the participants of the study to comment the questionnaire, its themes and individual items after filling it.

Another issue that can be seen as a limitation to the validity of the questionnaire is that the items (attributes) in the Attrak-Work questionnaire were created based on the findings from the observation and interview data. They therefore reflect this particular case study and the subjective views of its participants on the mobile journalism system and its usage. However, similar themes and attributes have arisen in our other case studies, with differences in the emphasis of themes depending on the group of participants and their backgrounds. We therefore believe, that for the context of mobile news journalism, the created questionnaire reflects well especially the hedonic aspects related to the mobile system use.

Although the current version of the Attrak-Work questionnaire is context specific especially regarding the attributes for hedonic quality identification, the themes included that are reflected by the individual items can be used as guidance when generalizing or targeting the questionnaire to another field of mobile work. Furthermore, the questionnaire can be applied to also any other type of mobile work tool in the context of journalism, be it a systems camera, laptop, audio recorder, or even pen and paper.

As a conclusion, we found the questionnaire to support our goal of corroborating and expanding the findings of the qualitative data and especially useful for capturing the perceptions of the hedonic qualities. In the future studies we are considering choosing one of the validated usability questionnaires to assess an evaluation of the pragmatic qualities either as a reference for attribute groups in Attrak-Work. In addition, we will carefully analyze the attribute group for appeal to identify possible needs for changes.

ACKNOWLEDGMENTS

We thank Tero Jokela for his valuable contribution to the study and Piia Nurkka for her comments on the article. This work has been supported by the PALTI project (2007–2009) funded by the Ministry of Education, Finland.

REFERENCES

- [1] Benford, S., Giannachi, G., Koleva, B. and Rodden, T. 2009. From interaction to trajectories: designing coherent journeys through user experiences. In: Proceedings of the 27th international Conference on Human Factors in Computing Systems. CHI '09. ACM, NY. Pp. 709–718.
- [2] Cockton, G. 2008. Putting Value into E-valuation. In: Maturing Usability – Quality in Software, Interaction and

- Value. Law, E., Hvannberg, E. and Cockton, G. (Eds.). Springer-Verlag London Ltd. Pp. 287–317.
- [3] Desmet, P. 2003. Measuring emotion: Development and Application of an Instrument to measure emotional responses to products. In: *Funology: From Usability to Enjoyment*. Blythe, M. A., Overbeeke, C., Monk, A. F. and Wright, P. C. (Eds.). Kluwer, the Netherlands. Pp. 111–123.
- [4] Hassenzahl, M. 2004. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, Vol. 19. 319–349.
- [5] Hassenzahl, M. 2003. The thing and I: Understanding the relationship between users and product. In: *Funology: From usability to enjoyment*. Blythe, M. A., Overbeeke, K., Monk, A. F., Wright, P. C. (Eds.). Kluwer, The Netherlands. Pp. 31–42.
- [6] Hassenzahl, M., Burmester, M. and Koller, F. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: *Mensch & Computer 2003: Interaktion in Bewegung*. Ziegler, J. and Szwillus, G. (Eds.). B.G. Teubner, Stuttgart, Leipzig. Pp. 187–196.
- [7] Hassenzahl, M., Platz, A., Burmester, M. and Lehner, K. Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. *Proc. of the CHI 2000 Conference on Human Factors in Computing*. ACM, New York, NY. Pp. 201–208.
- [8] Lang, P. J. 1980. Behavioral treatment and bio-behavioral assessment: Computer applications. In: *Technology in Mental Health Care Delivery Systems*. Sidowski, J. B., Johnson, H. and Williams, T. A. (Eds.). Ablex, Norwood, NJ. Pp. 119–137.
- [9] Law, E.L.-C., Bevan, N., Christou, G., Springett, M. and Larudsottir, M. (Eds.). 2008. *Proc. of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, Institute of Research in Informatics of Toulouse (IRIT), Toulouse, France.
- [10] Lee, C-P. and Shim, J. P. An empirical study on user satisfaction with mobile business applications use and hedonism. *Journal of Information Technology Theory and Application (JITTA)*, 8, 3, 57–74.
- [11] Mahlke, S. and Thüning, M. 2007. Studying antecedents of emotional experiences in interactive contexts. In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '07. ACM, New York, NY. Pp. 915–918.
- [12] Mehrabian, A. and Russell, J. A. 1974. *An approach to environmental psychology*. MIT Press, Cambridge, MA.
- [13] Osgood, C. E., Suci, G. and Tannenbaum, P. 1957. *The Measurement of Meaning*. University of Illinois Press.
- [14] Russell, J. A., Weiss, A. and Mendelsohn, G. A. 1989. The Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57, 493–502.
- [15] Ryu, Y. S. 2005. *Development of Usability Questionnaires for Electronic Mobile Products and Decision Making Methods*. Doctoral Thesis. State University, Virginia. 2005.
- [16] Spangenberg, E. R., Voss, K. E. and Drowley, A. E. 1997. Measuring the hedonic and utilitarian dimensions of attitude: a generally applicable scale. *Advances in Consumer Research*, 24, 235–241.
- [17] SUMI homepage [<http://sumi.ucc.ie/whatis.html>].
- [18] Tashakkori, A. T. and Teddlie, C. 2008. Quality of Inferences in Mixed Methods Research: Calling for an Integrative Framework. In: *Advances in Mixed Methods Research*. Bergman, M. M. (Ed.). SAGE, Great Britain. Pp. 101–119.
- [19] Van der Heijden, H. and Sørensen, L. S. 2003. Measuring attitudes towards mobile information services: an empirical validation of the HED/UT scale. In: *Proceedings of the European Conference on Information Systems (ECIS)*, Naples, Italy.
- [20] Vääätäjä, H., Männistö, A., Vainio, T. and Jokela, T. 2009. Understanding User Experience to Support Learning for Mobile Journalist's Work. In: Guy, R. (Ed.). *The Evolution of Mobile Teaching and Learning*, Informing Science Press. Pp. 177–210. Available at: <http://books.google.com>.
- [21] Wakefield, R. L. and Whitten, D. 2006. Mobile Computing: a user study on hedonic/utilitarian mobile device usage. *European Journal of Information Systems*, 15, 292–300.
- [22] Wigelius, H. and Vääätäjä, H. 2009. Dimensions of Context Affecting User Experience in Mobile Work. In: *the Proc. of INTERACT 2009*. Uppsala, Sweden, Aug. 26–28, 2009. (In press.)
- [23] Yin, R. K. 2003. *Case Study Research – Design and Methods*, 3rd ed. Sage Publications, USA.

Session 8: Analysis and Design of Tools for Process Control

Good Practice Regarding Alarm System Design and Alarm Management Strategies for Complex Process Control Settings

Anna Thunberg
M Sc, Lic Eng, Ph D student
Chalmers University of Technology
Division Design and Human Factors
SE 412 96 Gothenburg, Sweden
anna.thunberg@chalmers.se

Anna-Lisa Osvalder
Associate Prof, Ph D
Chalmers University of Technology
Division Design and Human Factors
SE 412 96 Gothenburg, Sweden
alos@chalmers.se

ABSTRACT

This paper summarises good practice regarding alarm system design and management. The results have been derived from a number of studies from different process control domains. A discussion is made concerning the need of studying good practice rather than focusing on solving existing alarm system problems, in order to design safe, efficient and resilient alarm systems. Factors contributing to successful interaction with the alarm systems are described and principle examples of new alarm system presentations are shown. The results can facilitate for control room designers to take the operator's perspective into account in alarm design projects, and thereby increase the operators' ability to handle unanticipated events which in turn improve process safety and efficiency.

Keywords

alarm systems, process control, safety

ACM Classification Keywords

H.5.2 [Graphical User Interfaces, User-centered Design].

INTRODUCTION

Alarm systems play an important role in maintaining safe and efficient operation of process control settings. The alarm systems are essential parts of the operator interfaces and should provide vital support for the operators in order to handle process variations and avoid or manage incidents and disturbances. Well-designed alarm systems facilitate safe operation, whereas deficiencies can contribute to incidents. Furthermore, poor performance of alarm systems can result in financial losses, environmental consequences and hazards to people [3, 25].

Research regarding alarm systems has mainly been divided into two approaches; improving the human-alarm system interaction,

and including technical methods, typically for reducing the frequency of alarms [9]. Improvement to the human-system interaction can be made by addressing the design of the alarm system interface, develop adequate training and provide alarm handling procedures [8, 21]. For example, Papin and Quellien (2006) [24], emphasize the relationship between the design of the human-system interface and the operator performance and reliability. Reduction of the number of alarms can be made by e.g. including filtering or suppression techniques [3, 26]. Both these approaches use existing alarm system problems as a starting point with attempt to reduce the effects of the problems. However, to design usable alarm systems, designers should not only avoid well-known problems. It is important to understand which alarm design and management factors that facilitate for operators to detect, understand and manage changes in the operational environment and unanticipated events [19]. Also, it is of great interest to realise how the alarm system can assist the operator in taking corrective measures to process anomalies whilst staying within the safe envelope [6].

The need of studying success factors is an emerging awareness, partly due to the introduction of resilience engineering. Good practice and lessons learned have been available for many years, but they have primarily been derived from existing alarm system problems. By applying the approach of resilience engineering, the focus is shifted towards understanding successful management rather than solving alarm system design and management problems. Resilience engineering represents a new way of thinking about safety proactively, e.g. by trying to enhance the ability of organisations to create robust and flexible processes in operational contexts, to monitor risks, and balance safety, production and economic requirements [18, 30]. A failure does not stand for a breakdown or malfunction of a normal system in resilience engineering. Rather, a failure represents an inability of the system to adequately adapt to cope with the real world complexity. Thus, success is the ability of the system to monitor the changing risk profile and make necessary adjustments so that the system can sustain operational. The adjustments can be made either prior to or following changes and disturbances. Further, resilience engineering could facilitate for operators to manage unexpected situations.

A resilient system must have the abilities to:

- respond, quickly and efficiently, to regular disturbances and threats,
- continuously monitor for irregular disturbances and threats, and to revise the basis for the monitoring when needed,
- anticipate future changes in the environment that may affect the system's ability to function, and the willingness to prepare against these changes even if the outcome is uncertain.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

By using the resilience engineering approach together with more recent developed user interface design approaches, for example ecological interface design [4, 14], the possibility for the operators to safely and efficiently manage situations characterised by novelty and change has dramatically improved. To facilitate for alarm system designers to pro-actively design well-functioning alarm systems, it is beneficial to use experiences from different industry sectors regarding alarm handling success factors.

OBJECTIVE AND GOAL

The objective of this paper was to summarise results from a number of studies concerning alarm systems in process control domains, performed during the 2000's by the Human-Machine System research group at Chalmers University of Technology, Gothenburg, Sweden. All studies have been carried out in order to understand the operators' interaction with existing alarm systems, the hazardous situations they need to manage and how the environment and organisation affect the operators' work. The goal with the summary is to identify factors contributing to successful interaction with the alarm systems, as well as good practice regarding alarm system design and management. The results can facilitate for control room designers to take the operator's perspective into account in alarm design projects, and thereby increase the operators' ability to handle unanticipated events, which in turn improve process safety and efficiency.

METHOD

The industrial sectors studied were nuclear power, pulp and paper, oil refining and heat and power. Furthermore medical care and aviation were also included. All studies have been performed at Swedish facilities and with Swedish users. In all settings, experienced users with more than 5 years of practice participated. The primary methods used to gain an understanding of the operators' work were in-situ observations in control rooms and simulators, and interviews with operators, suppliers and system designers. Furthermore, hierarchical task analysis (HTA) [13], applied cognitive task analysis (ACTA) [20] and scenario analysis [11] were used to describe the operators' work. The operators' mental workload was assessed by NASA-TLX [7]. Risk analysis was performed to predict and assess human error using PHEA [2] and HEART [5]. Information

regarding the individual studies used to analyse different alarm systems is presented in Table 1.

Table 1. Summary of the studies included in the analysis.

Industry sector	Studies			References
	Number of cases	Subjects	Methods	
Nuclear power	7	operators & system designers	observations interviews task analyses NASA-TLX HEART ACTA	[12, 23, 27–29]
Pulp and paper	4	operators & system designers	observations interviews	[12, 22]
Oil refining	3	operators & system designers	observations interviews	[10, 12, 22]
Heat and power	4	operators	observations interviews	[1, 22]
Medical care	3	nurses patients	observations interviews PHEA ACTA	[2, 12, 15–17]
Aviation	3	pilots	observations interviews	[12, 22]

The results from the different industry sectors in Table 1 were accumulated and compiled in a standardised format, derived from the system model presented in Figure 1. The description of the system, its goal and characteristics regarding the technical alarm system, the user interface of the alarm system, and the organisation and environment were compiled for each sector. Further, assessments regarding the operators' interaction with the alarm system were performed. From the results, advantages and success factors were identified. After summarizing the results from each domain, an overall analysis was performed to specify critical alarm design criteria independent of industry sector.

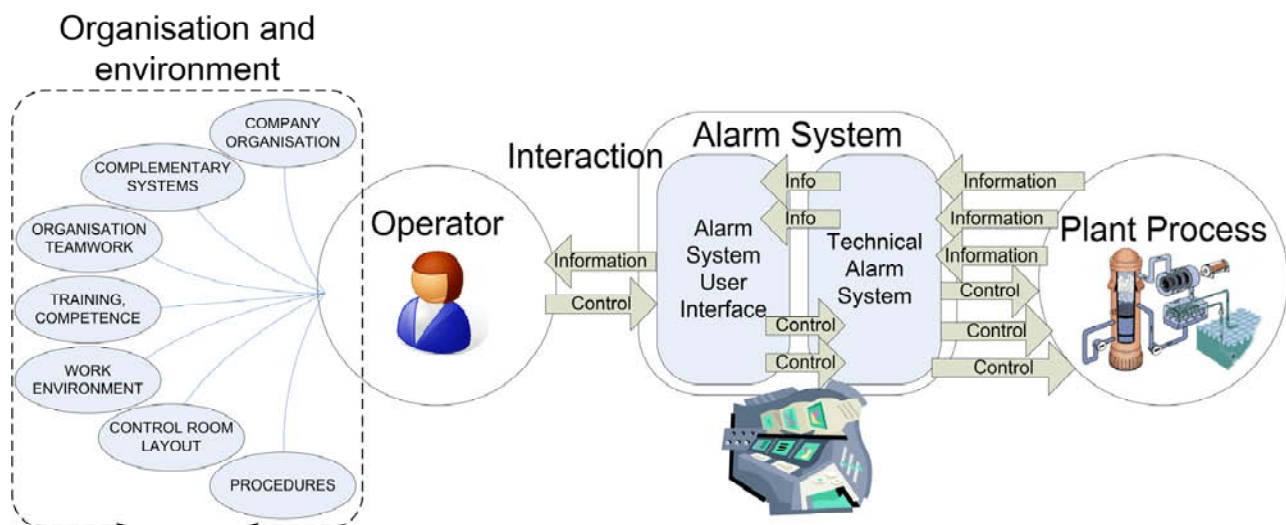


Figure 1. System model of factors affecting alarm management.

RESULTS AND ANALYSIS

The references presented in Table 1 demonstrate deeper descriptions regarding characteristics of the different industry sectors control room settings and the operators' interaction with the alarm systems.

The overall results show that in order to obtain successful management of alarms, the alarm system user interface, the technical alarm system, the organisation and the environment need all to be attended to. How well the design of the user interface, the technical alarm system, the organisation and the environment meets the conditions set by the process and operators define how successful the interaction is.

The following results are considered as general for all sectors.

Alarm System User Interface

The results from the studies show that depending on the operational mode (e.g., full-power, start-up and shutdown) and the work situation (e.g., normal operation or disturbance), the operator's role and the operational goal change and thus the operator's need of information varies. For example, in full-power operation the operators try to optimise the process and needs detailed process information, whereas in large disturbances the operators want to ensure safe shut-down.

A simplification to illustrate the differences is to use the abstraction hierarchy and part-whole decomposition (Figure 2) [4]. In steady-state full-power operation, the operator is interested in detailed information regarding individual components, e.g. the capacity and the component's relationship to other objects. This means that the operator needs information regarding the physical function on a component level (lower, right corner in Figure 2).

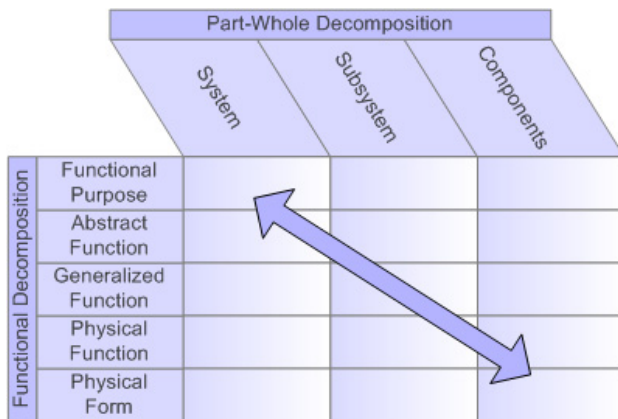


Figure 2. Combination of the abstraction and part-whole hierarchies [4].

In disturbances, the operator's main interest is to control overall process status on a system or subsystem level. However, this simplification is only partly true. To be able to ensure e.g. a safe shut-down, the operator needs to check the status of individual components as well, for example if valves

are closed or not. The important conclusion is that the classical user alarm interface, which often is focused on detailed information on component level, needs to be complemented with other information representing the status of the functional purpose, the causal relationships and, in addition, make the information relevant on system and subsystem level.

Consequently, the role of the operator is very important for the operator's behaviour and the processing of information. To meet the operator's varying needs, the alarm system should provide complementary information on different abstraction levels and levels of detail. Thereby the operator can choose relevant information in a specific work situation. By adding information on system level and with higher degrees of abstraction, the operator is also more likely to make sensible decisions regarding the prioritisation of the work, since the overall process status is more easily perceived and understood. A typical problem in e.g. oil refining is that the operators could get stuck on specific alarms and problems, and therefore miss the overall picture.

A key factor for successful performance is that the operators get continuous information and feedback of the system's status. The operators need feedback regarding the results of their measures, automatic sequences and they need information regarding critical process data. For example, safety-critical alarms should be spatially dedicated and continuously visible, which provide pattern recognition. Given this, the operator can quickly detect variations in the system's status. Further, visual aids to enhance detection of process changes should be implemented.

The operators included in the studies try to manage the process by being aware and pro-active. They actively monitor key parameters to be able to early detect deviations. To facilitate for evaluation of parameter values, the value should be presented together with the set-points and/or alarm limits.

When designing an alarm system user interface, special concern should to be taken to ensure that the design is consistent with other equipment and control systems within the control room concerning phrasing, abbreviations etc.

To facilitate interpretation of alarms, the alarm should be presented within a clear frame of reference, e.g. the alarm limits. Some type of indication of the parameter's trend is also helpful and enhances correct interpretation of the alarm and helps the operators with prioritising their work. Further, a user interface that guides the operator's initial response contributes to successful alarm handling.

Figures 3–4 present examples of how the alarm system user interface can be improved to facilitate for the operators to detect anomalies, to detect and handle individual alarms, and to identify and take corrective measures in disturbances. Figure 3 illustrates an example from a modernised system of a limited subsystem of a nuclear power plant, i.e. the modernised alarm system should function together with the existing alarm system in the control room. Figure 4 presents a generic example of improvements in the visual design of alarms in heat and power industries.

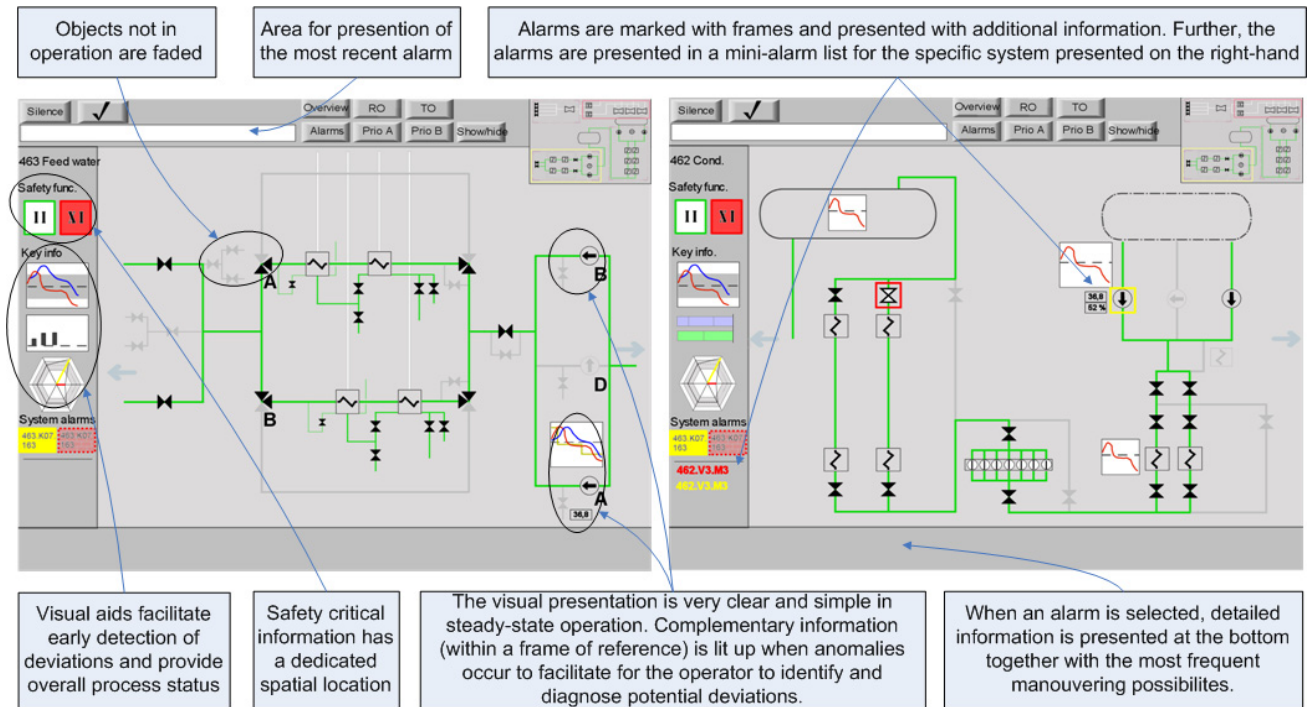


Figure 3. Improvements of an existing alarm system user interface (nuclear power plant).

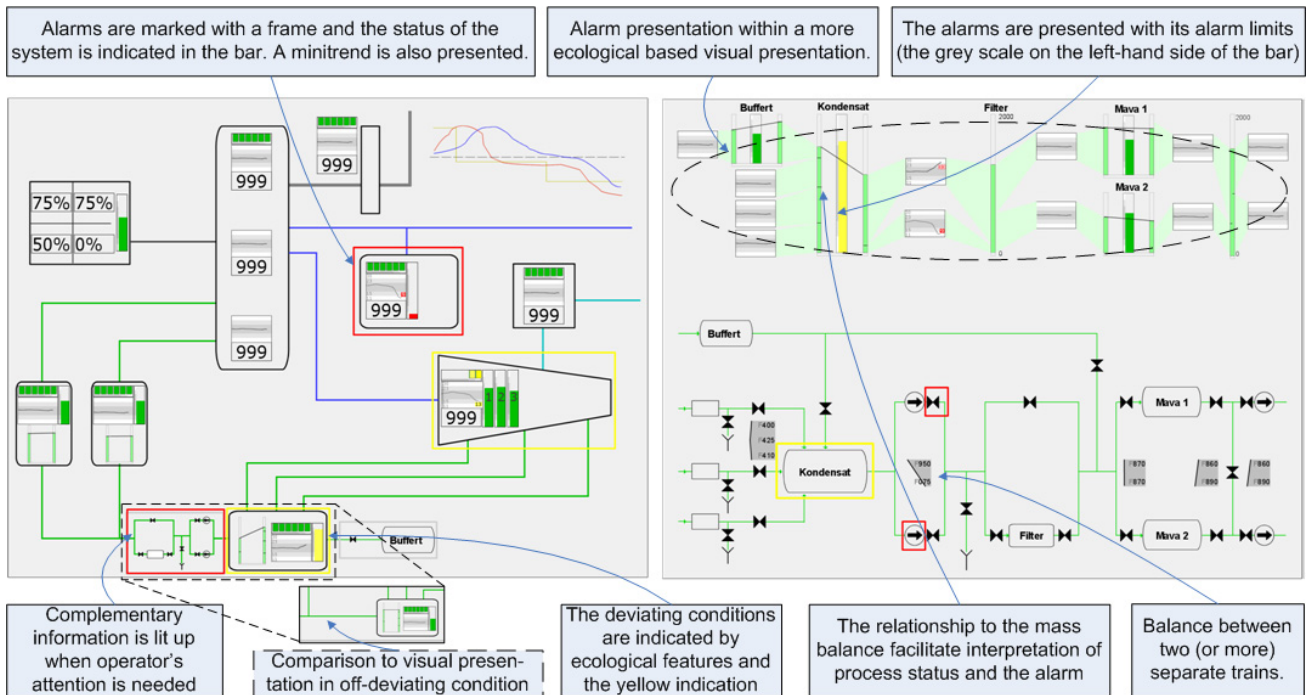


Figure 4. Principle example of new alarm presentation for heat and power application.

Technical Alarm System

The results indicate that the alarm systems today primarily are used in steady-state full-power operation and in minor disturbances, due to too high amount of alarms in off-normal operational modes and large disturbances. To improve the use of the alarm systems and make them useful also in off-normal operation, technical functions such as reduction of alarms and prioritisation of alarms should be evaluated.

To evaluate each alarm that is implemented and ensure that it requires a response from the operator is one of the most efficient strategies to keep the alarm rate low. With computerised alarm systems it is very easy and cheap to implement alarms, and therefore the number of alarms within the systems has increased dramatically in many industries. An alarm system in aviation was found to be the most technically efficient. It included a very low rate of alarms in total, about 200 alarms which can be compared to process industries which could

present up to as much 20 000 alarms. Aviation use mode-dependent prioritisation and suppression and alarm flooding is very seldom possible. However it is impossible to compare the situation in an airplane cockpit with the control room in a large process plant, but the idea of careful selection of alarms and prioritisation is very beneficial independent of industry, and one of the most excellent strategies when designing well-functioning alarm systems [3].

Proper prioritisation of alarms enhances the operator's ability to focus the attention correctly. This is especially important for less experienced operators and applicable to situations which can be foreseen. In nuclear power, the experience of an operator had no significant effect on the decision making process. Instead, the level of experience affected which type of information the operator perceived. The experienced operators were more likely to use multiple sources of information to verify information and they also had a good ability to foresee possible outcomes of different situations. Thus, the experienced operator has advantages in assessing situations correctly. On the other hand, if the less experienced operator is supported in the perception of information by checklists and procedures, he/she is very likely to reach the same decision. This indicates that a main task for the alarm system is to reduce distracting stimuli and to guide the operator towards the important information by using checklists in deviations and implement techniques for alarm prioritisation and alarm suppression.

Resilience engineering and ecological interface design both emphasize the need of supporting the operators in situations characterised by novelty and change. It is important that the operators can handle unanticipated events, but it is also significant that frequent tasks and anticipated high-risk situations can be managed efficiently. Thus, special attention should be given to classic task analyses and other methods that focus on evaluating operator awareness, operator reliability and task performance for anticipated tasks and events. By analysing the operative goal in the specific mode and/or work situation and the operator's need of information and support, an alarm system with a more effective design can be developed. Further, alarm prioritisation and suppression specially adapted to different operational modes, e.g. start-up or shut-down, can be very efficient to reduce the amount of irrelevant alarms and distracting stimuli.

An operator who feels in control and understands the process will be more successful in his/her work. One of the largest advantages with computer based technology was, according to operators, the higher degree of integration of information and links between process objects, procedures and parameter settings. For example, instructions integrated in the interface were experienced as a good support.

In conclusion, factors to attend to regarding the technical alarm system are primarily; proper prioritisation of alarms, alarm system adapted to the operator's differing roles dependent on system status and operating aim, and reduction of distracting stimuli, i.e. possibilities to suppress/eclipse multiple alarms or block irrelevant alarms, typically from equipment under service.

Organisation and Environment

One of the most common mentioned resilience enhancing factor and success factor identified in the studies was the ability of flexibility within the control room crew. Clear responsibilities but knowledge about each other's tasks and assignments facilitated collaboration and the ability to support each other. For example, in the cockpit, the pilots can divide

the work between them in a way that is suitable for the specific situation, since both pilots usually share the same type of education and knowledge. The operators in nuclear power, oil refining, pulp and paper, and heat and power can regulate the workload within the shift team to maintain high overall performance. The operators regulate their workload by prioritising tasks according to available time and resources. They use external cues to decrease the workload. The shift supervisor in nuclear power has an important role in maintaining a manageable workload level for all operators in the team, by distributing the tasks between them. For future alarm systems, it is important to implement support functions for the operators to create and use external cues e.g., possibility to define personalised alarms and write comments in connection to objects and/or alarms. Furthermore, external cues can reduce the need for the operator to recall information when moving from one operation to another in a task sequence. In addition, training should be provided that helps the operators to support each other in different operating modes and process situations.

One important aspect to consider is that the work performed in the studied control rooms is not the result of the performance of an individual operator, but the result of the operators working together. A central issue for the alarm system is therefore to facilitate for natural collaboration in the control room team. For example, a main shared overview display should be provided to support the operators to have a shared understanding of the situation. Clearly, an important factor for the management is to assess and appoint a proper staffing level.

Another key aspect in successful development of alarm systems is that the organisation advocates a culture of improvement. A newly commissioned alarm system is not fully developed, but needs improvements. Further, to identify the need of improvement, there must be a plan for recurrent review of the alarm system with alarm system measurements and targets.

The usefulness of procedures can be debated depending on industry sector. Some operators assert that procedures are useful for almost all types of tasks, whereas other operators say that procedures cannot be developed for all tasks and process variations. The more stable a process is, the more easily can procedures be developed and the procedures are also more likely to be correct. In such a context it can be beneficial to have procedures for many tasks. If the plant process is characterised by process variations, it is not likely that the operators will use the procedures since they will be difficult to apply to the different situations. However, procedures for anticipated disturbances are useful since they facilitate for the operator to manage the situation and reduce the mental workload, which can be very important in a dangerous situation. Thus, the operator is also more likely to detect if anything out of the normal disturbance procedure occurs.

Training is a factor that should not be underestimated. With proper training, the operator has sufficient knowledge and proficiency for dealing with various work situations. Thorough training gives the operator the needed skills, knowledge about the alarm system and also provides the operator with handling alternatives.

The environmental design of the studied control rooms was satisfying. The operators worked in a good thermal climate, with sufficient lighting and physical ergonomics. However, these factors are also important to consider during the design of the control room since they can have an impact on the operator's ability to detect, read, listen to and respond to alarms.

To conclude, organisational and environmental factors that contribute to successful alarm handling are; possibility to

regulate and distribute workload, continuous monitoring and improvement of alarm system performance, emergency operating procedures in stressful situations, and proper staffing levels in order to handle anticipated transients.

DISCUSSION

The factors presented in the result section are examples of good practice identified as important for all industrial sectors included in the studies. The results indicate that there are several success factors and good practice that could be considered as generic. However, the number of case studies in the included industry sectors varied as well as the extent of the interviewed operators. Process industries, and especially nuclear power control, have been investigated more thoroughly than sectors as aviation and medical technology.

The applicability of the success factors should be considered as generic, i.e. they are applicable independent of type of industry sector and type of alarm system development project (ranging from a minor modernisation with an add-on system to the complete development of a new control room with control and alarm system). However, adaptations are of course necessary and the importance of the factors can vary between industrial sector and project. For example, in minor modernisation projects it can be more important to be consistent with the other control and alarm equipment in the control room, than to follow specific design guidelines.

One of the most important factors when developing alarm systems is to consider the whole system and the various working situations that might occur. If only the design of the alarm system user interface is focused, the overall system performance is likely to decrease, since the consistency between the technical alarm system, the organisation and the environment will be insufficient.

Furthermore, since the different sectors have realised their alarm system design and alarm management strategies differently and with varying success, the need of exchanging experiences and knowledge is prominent. However, rather than to address alarm system problems in the knowledge exchange, it is most important to exchange good practice in order to obtain useful, efficient and safe alarm handling.

The combination of the approaches of resilience engineering, ecological interface design (EID) and classical design approaches and methods are most useful to provide operators with safe and efficient alarm systems. If only the approaches of resilience engineering and EID are used, the operators might be able to handle new and unanticipated events, and they will easily detect anomalies and predict how they affect overall system performance. However, frequent tasks might be inefficient to perform and interaction with objects can be illogical. The aim of the EID approach is to identify how component resources map functionality to purposes. For example, task analysis can provide information regarding which activities that are needed and facilitate for users to achieve particular goals [4].

The alarm design ideas (Figures 3–4) proposed in this paper are only extracts from more comprehensive design suggestions and only parts of the screens are shown. One of the aims in the design projects where the screen layouts have been developed [22, 28, 29] has been to address the operators' acceptance to new design proposals. The conclusion regarding this issue is that new designs are more easily accepted if they are complemented with a design the operators' are familiar with. In that case, they can easily understand and accept the potential of the new design features. Therefore, the suggested screens partly have been based on traditional mimic layouts.

CONCLUSIONS

The following conclusions have been found about operator performance and design of alarm systems:

- The operator's various roles in different operational modes are very important for information processing and thus the alarm system interface design. Future alarm systems should be adaptive and/or present information on multiple abstraction and detail levels. The design should take the modes of operation, the role of the operator, and the aim of the operator into consideration.
- Consider the entire system and design the technical alarm system, the alarm system user interface, and the organisational and environmental aspects in connection to each other.
- Workload regulation is very important to maintain high overall performance. The operators regulate their workload by prioritising tasks according to available time and resources. They use external cues to decrease the workload. The supervisor has an important role in maintaining a manageable workload level for all operators. A situation with high temporal demands does not necessarily imply high mental demands. The use of emergency procedures is one example of how the mental workload can be decreased.
- Prioritisation of alarms is one of the most efficient technical functions to enhance alarm handling.
- Fast detection and interpretation of anomalies and alarms are important to improve process availability and safety. Pattern recognition is efficient for safety-critical alarms and visual aids (e.g. mass balances and object information in a frame of reference) enhance detection of process changes.
- Ensuring that every alarm requires a response is an efficient strategy to keep the number of alarms down.
- The operator's decision-making is very much influenced by the perceived information and not so much dependent on the operator's level of expertise. The alarm handling can be made more efficient if the operator is supported by:
 - Emergency operating procedures
 - Alarm prioritisation
 - Reduction of distracting stimuli
 - Suppression of irrelevant alarms.
- Improve the alarm system's ability to guide the operator's initial response to the deviation. Action lists should be considered as a future feature of alarm systems.
- Combining design approaches focused on managing new and unanticipated events with traditional design approaches focused on task efficiency, will probably provide the best overall result and alarm management performance.

ACKNOWLEDGMENTS

The participation of all operators in the various industry sectors has been highly appreciated. The financial support given by Värmeforsk (Thermal Engineering Research Institute), the Swedish nuclear power inspectorate (now: the Swedish Radiation Safety Authority) and the Swedish nuclear power stations has made these studies possible. Finally, the work of our research colleagues Lars-Ola Bligård and Jonas Andersson at Chalmers is highly acknowledged. They have made valuable contributions to the designs of the user interfaces.

REFERENCES

- [1] Ahlén, F. 2009. Control Room User Interfaces – proposals for screen images used for visualization of plant processes Chalmers University of Technology, Gothenburg.
- [2] Bligård, L.-O. 2007. Prediction of Medical Device Usability Problems and Use Errors. Chalmers University of Technology, Göteborg.
- [3] Bransby, M. L. and Jenkinson, J. 1998. The management of alarm systems: HSE Books.
- [4] Burns, C. M. and Hajdukiewicz, J. R. 2004. Ecological Interface Design. Boca Raton, Florida: CRC Press.
- [5] Embrey, D. 2004. Qualitative and Quantitative Evaluation of Human Error in Risk Assessment. In: Human Factors for Engineers (Eds.). C. Sandom and R. Harvey. London: Institution of Electrical Engineers.
- [6] Hale, A. and Heijer, T. 2006. Defining Resilience. In: Resilience engineering: concepts and precepts (Eds.). E. Hollnagel, D. D. Woods, N. Leveson. Aldershot: Ashgate.
- [7] Hart, S. G. and Staveland L. E. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Human Mental Workload (Ed.). PA Hancock, N Meshkati, Amsterdam: Elsevier. Pp. 239–250.
- [8] Hickling, E. M. 1994. Ergonomics and engineering aspects of designing an alarm system for a modern nuclear power plant. In: Human Factors in Alarm Design (Ed.). N Stanton, London: Taylor and Francis. Pp. 165–178.
- [9] Hwang, S.-L., Lin, J.-T., Liang, G.-F., Yau, Y.-J., Yenn, T.-C. and Hsu C.-C. 2008. Application control chart concepts of designing a pre-alarm system in the nuclear power plant control room. Nuclear Engineering and Design 238: 3522–3527.
- [10] Jensen, F. and Mattsson, S. 2004. Konceptdesign av process- och larmpresentation: användarcentrerad systemutveckling av översiktlig informationspresentation i processindustri. Master Thesis. Chalmers University of Technology, Gothenburg.
- [11] Johnsen, S. O., Bjørkli, C., Steiro, T., Fartum, H. and Haukenes, H. et al. 2008. CRIOP®: A scenario method for Crisis Intervention and Operability analysis.
- [12] Jönsson, A. and Osvalder A.-L. 2004. En jämförande studie av larmsystem i svenska kärnkraftverk och andra branscher. Rep. 29, Chalmers tekniska högskola, Institutionen för produkt- och produktionsutveckling, ISSN: 1651 0984, Göteborg (In Swedish).
- [13] Kirwan, B. and Ainsworth, L. K. 1992. A guide to task analysis. London: Taylor and Francis.
- [14] Lau, N., Jamieson, G. A., Skraaning, G. Jr. and Burns, C. M. 2008. Ecological interface design in the nuclear domain: An empirical evaluation of ecological displays for the secondary subsystems of a boiling water reactor plant simulator. IEEE Transactions on Nuclear Science 55: 3597–3610.
- [15] Liljegren, E. 2004. Increasing the usability of medical technology: methodological considerations for evaluation. Ph D thesis. Chalmers University of Technology, Gothenburg.
- [16] Liu, Y. 2009. Usability Evaluation of Medical Technology: Investigating the Effect of User Background and Users' Expertise. Product and Production Development – Chalmers University of Technology, Göteborg.
- [17] Liu, Y., Dahlman, S. and Osvalder, A.-L. 2003. Evaluating alarm system design on medical equipment in the early design stage: A three-method based approach. Presented at 35th Nordic Ergonomics Society Conference, Reykjavik, Iceland.
- [18] Madni, A. M. and Jackson, S. 2009. Towards a conceptual framework for resilience engineering. IEEE Systems Journal 3: 181–191.
- [19] McDonald, N. 2006. Organisational Resilience and Industrial Risk. In: Resilience Engineering: Concepts and Precepts (Eds.). E. Hollnagel, D. D. Woods, N. Leveson. Aldershot: Ashgate.
- [20] Militello, L. G. and Hutton, R. J. B. 1998. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understand cognitive task demands. Ergonomics 41: 1618–4161.
- [21] O'Hara, J. M., Higgins, J. C., Persensky, J. J., Lewis, P. M. and Bongarra J. P. 2004. Human Factors Engineering Program Review Model. Washington: US Nuclear Regulatory Commission.
- [22] Osvalder, A.-L., Bligård, L.-O., Andersson, J. and Thunberg, A. 2009. Visualisation of plant status – New interface design proposals. Division Design and Human Factors, Chalmers University of Technology, Gothenburg.
- [23] Oxstrand, J. 2005. Utveckling av säkert och effektivt användargränssnitt – vilka faktorer påverkar risken för mänskligt fel? Chalmers University of Technology, Göteborg.
- [24] Papin, B. and Quellien, P. 2006. The operational complexity index: A new method for the global assessment of the human factor impact on the safety of advanced reactors concepts. Nuclear Engineering and Design 236: 1113–1121.
- [25] Smith, W. H., Howard, C. R. and Foord, A. G. 2003. Alarms Management – Priority, Floods, Tears or Gain?: 4-sight Consulting, http://www.4-sightconsulting.co.uk/Current_Papers/Alarms_Management/alarms_management.html.
- [26] Sørensen, A., Veland, Ø., Farbrot, J. E., Kaarstad, M. and Seim, L. Å. et al. 2002. Recommendations to alarm systems and lessons learned on alarm system implementation, Institutt for energiteknikk, OECD Halden Reactor Project.
- [27] Thunberg, A. 2006. A Cognitive Approach to the Design of Alarm Systems for Nuclear Power Plant Control Rooms. Lic. Thesis. Chalmers University of Technology, Göteborg.
- [28] Thunberg, A. and Osvalder, A.-L. 2007. Design of Alarm Systems in Different Complex Control Settings. In: Human Factors Issues in Complex System Performance (Ed.). D. deWaard, G. R. J. Hockey, P. Nickel, K. A. Brookhuis. The Netherlands, Maastricht: Shaker Publishing.
- [29] Thunberg, A. and Osvalder, A.-L. 2007. What constitutes a well-designed alarm system? In: Joint Conference between IEEE – Human Factors in Power Plants and Human Performance, Root Cause and Trending. Monterey, CA, USA.
- [30] Woods, D. D. 2006. Essential characteristics of resilience. In: Resilience Engineering: concepts and precepts (Eds.). E. Hollnagel, D. D. Woods, N. Leveson. Alderton: Ashgate.

Using Operating Procedures in NPP Process Control

Leena Salo
VTT Technical Research
Centre of Finland
P.O. Box 1000
02044 VTT, Finland
Leena.Salo@vtt.fi

Leena Norros
VTT Technical Research
Centre of Finland
P.O. Box 1000
02044 VTT, Finland
Leena.Norros@vtt.fi

Paula Savioja
VTT Technical Research
Centre of Finland
P.O. Box 1000
02044 VTT, Finland
Paula.Savioja@vtt.fi

ABSTRACT

This paper reports results of a qualitative analysis of procedure usage in NPP process control. The data was gathered in a test series that was conducted at the training simulator of the Fortum Loviisa nuclear power plant (NPP) in autumn 2008. The aim of the research is to construct understanding of the role of operating procedures in process control on a general level. In addition, the role of emergency operating procedures in structuring the activity in a specific accident scenario is studied. To demonstrate the methodology and to analyse in detail the role of procedures in one accident situation, the results concerning one run and one crew are presented.

Keywords

emergency operating procedures, process control, qualitative analysis

ACM Classification Keywords

H.1.2 [Models and Principles] User/Machine Systems – *Human factors*.

INTRODUCTION

The use of operating procedures is a basic issue when dealing with the work in complex safety-critical environments. Designing procedures for various operations is a challenging task because design cannot only rely on engineering knowledge. It also requires understanding of how the system is operated. Gaining of operational experience has therefore had a significant effect on the number of available procedures, their role in the daily work of the personnel, and also on the type of procedures. After the Three Mile Island accident in 1979 the nuclear community awakened to strictly require procedures, new event-based procedures were developed, and adherence to procedures was strongly insisted. At the same time it became evident that when operators face an unexpected situation procedures written to pre-defined disturbance classes or situations are not optimal. Hence, it would also be necessary to develop procedures that are focused on identifying signs of change from stable state. The symptom-based procedures emerged. These should offer

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

guidance to maintain or restore stability as a first goal. Later experiences of major accidents, especially Chernobyl, brought up another issue, that of safety culture. Appropriate procedures and adherence to them was thereafter one of the measures of good safety culture.

Both safety culture and procedures have dominantly been dealt with as organisational phenomena, which they certainly are. Actual practices of personnel's ways of using procedures, which both manifests the organisation's culture and creates it, has not been studied as broadly. The present study focuses on the usage of procedures.

The specific context of the study is the modernisations of the information and control systems (I&C) that are currently ongoing in both of the Finnish nuclear power plants. During these modernisations old analogue technology is replaced with digital automation and human-system interfaces (HSIs). In one of the plants, i.e. in the Loviisa plant, the modernisation started by replacing some of the original analogue human-system interfaces of control rod position indication and control systems, some preventive functions and nuclear water treatment systems with new digital HSIs. A couple of years before starting the modernisation of HSIs, the implementation of new symptom-based procedures for handling emergency situations took place. Our intention in the present study was to analyse at a full-scope training simulator the use of the new emergency operating procedures (EOPs).

Some recent international studies have also drawn our interest to the issue of procedure usage. These studies report of changes that take place in operator responsibilities and tasks in connection to implementing new procedures.

BACKGROUND

Earlier Findings

A large number of empirical analyses on the use of emergency operating procedures have been accomplished by Electricité de France in a series studies already in the 1990's [1]. Conceptions of operator activity and operators' application of procedures were drawn from more than a hundred tests in simulated nuclear power plant operation contexts. These studies provided a basic understanding of how procedure usage was, and how it probably should be understood in the NPP domain. Dien [1] discusses how the general agreement of the need of using procedures is interpreted by different stakeholders, i.e. regulators, designers, and operators, and whether operator use procedures as designers require or assume.

Different Conceptions of Procedures

From the point of view of the *regulator* procedures are a legal requirement and, consequently the nuclear regulator insists that the utility demonstrates the existence of appropriate procedures

and their competent use. The availability and correct use of operating procedures is an important factor contributing to plant safety. The *designers'* task is to create a system that operates effectively and safely in all operating situations. Designers need to anticipate possible problem situations and design means to handle them. From their point of view operators may endanger safety by committing errors. According to this view, errors can be avoided by following procedures that are pre-planned for difficult situations. Dien [1] maintains that the designers' basic assumptions of the internal characteristics of operator work activity are, however, inadequate and, hence, false expectations of the role of procedures and of the ways of their application arise. According to the designers' implicit model 1) an operator is "a guided" being who should follow instructions literally; 2) the situation in which operators use procedures is a series of chronological actions that must be carried out in a particular order; 3) operators are considered as "average operators", hence personal and cultural differences in work are devoted only minor attention. The message of Dien is that this view of operator activity is rather mechanistic and non-dynamic.

Operators themselves who have experience of the dynamics of the process feel the need for procedures as help in diagnosing problems and maintaining the process under control. According to Dien's [1] findings it is rare that operators withdraw from using procedures in accident situations. The author states, however, that non-use of procedures still takes place. According to his studies the major reason has been that operators have not been able to immediately connect the procedure to the plant situation. The problems in connecting procedures and the situation were due to either procedure design or previous operator errors. It was also found that almost in all test runs there were small deviations from the prescriptions of the procedures which, however, did not have effect on the quality of the end result. Sources for these deviations were needs to respond to events and detailed conditions that are not included in the procedure descriptions. Procedures describe situations in a rather straightforward way and do not include minor additional failures in typical event scenarios. Finally, procedures assume a particular, implicitly in-designed level of competence. Naming a required action may be described in very many levels of detail, and in reality, depending on operator competence, the needs for level of details are different.

The conclusions Dien [1] draws on his studies is that operators are not able to, and that it is not reasonable to require that they should, follow procedures literally. Instead operators are required in their procedure application to "make up their oversights", i.e. complete procedures to fit the demands of the situation. They must also compensate the static nature of a procedure, by which the author denotes the fact that the dynamics of the process situation are complex and do not result in exactly describable courses of events. These features correspond to what Dien calls "intelligent application of procedures". This notion refers to kind of compromise between the designer and operator view, i.e. a strict adherence to procedures is required as long as these are adapted to the situation. In other words operator competence is required to realise when the procedure descriptions divert from reality.

Psychological Demands of Acting with Procedures

Despite of the insight of procedure usage the issue actualised again at EdF when implementing computerised procedures to the N4 NPPs. According to Perin [2], in the connection of implementing computerised procedures to N4 EdF's own top managers expressed concerns of introduction of probably a too high level of guidance of operators [2, p. 217]. Later Filippi

and collaborators studied thoroughly the usage of procedures by the operators of N4 plants. Filippi [3] synthesised their findings by defining the psychological characteristics of procedure usage in incident and accident situations. Our interpretation, which may also be supported by other authors' work [4, 5], of these characteristics is as follows:

- *intelligent use of procedures* which is portrayed in a balanced combination of procedure guidance and operator competence especially in identifying the situation, whereby an ability to maintain control over own action is required
- *connecting the situation with the procedure* which requires the operator to create a context for interpretation, i.e. to make the procedures meaningful in the current situation and a relevant framework for action, which requires cognitive effort
- *maintaining coherence in acting* despite of interruptions that require changes in attention and maintaining unfinished tasks active for retrieval; the interruptions are often induced by other team members which requires anticipation and control of other persons' acting
- *organising efficient collaboration* and coordinating with other team members according to the prescriptions of the procedure which typically require parallel independent acting with occasional joint updates between the team at prescribed phases
- *developing trust in automation and procedures* is important for the development self-confidence and professional identity which have effect on operators' the ability to act in uncertain situations [6, 7].

Procedures as a Human-system Interface

Procedures are a relevant issue for the NPP operators' work also because they act in a role of an *interface* to the process. This aspect of procedures has gained relevance in I&C system digitalisations at NPPs, in which connection also the computerisation of procedures is considered. Studies in which computer-based procedures (CBP) are compared with the traditional paper-based procedures (PBP) are of interest to us because in these studies the role and functions of procedures in operator work are discussed. In many studies on procedure design the focus is more on specific issues of the format without connection to the functions of procedures.

Drawing on own work and a literature review O'Hara et al. provide good basis for comparing paper-based and computer-based procedures [5, pp. 4–2]. According to the authors the two types of procedures share many basic elements, both in format and in contents. O'Hara et al. maintain that, rather than the procedure format, a more significant difference between paper-based and computer-based procedures is the functionality. Originally procedures were designed to support planning of response to process deviations. While PBPs do support response planning, they are less helpful in such tasks as monitoring, situation assessment, and implementation of operations. Also, the computer medium gives procedure designers means to improve the layout and salience of information as well as to provide better support for procedure managing.

PBPs and CBSs have also been compared with regard to their support for team collaboration. In [8] implementing computerised procedures was found to have some effects on communication and collaboration. Because the new CBPs offered plant parameter data, the need for low-level communication between shift supervisors and board operators was reduced. Because the board operators did not have to serve as the "eyes" of the

supervisor their cognitive resources were freed somewhat. Although the reduction of the need for low-level communication can be seen as a positive thing it had also some negative consequences: there were occasions where operators failed to inform the supervisor of some significant events, and the common understanding of the situation among the crew members was weakened. Also, there were complaints about a narrower field of view provided by CBPs and difficulties in “looking ahead”, i.e. seeing next steps in advance.

Pirus [9] considers procedure-based control as an aspect of process automation. He studied operators’ reactions to different levels of automation of procedures in a situation where operators had a chance to select the level of automation of procedures they preferred. His observation was that more experienced operators accepted higher level of automation because they were able to anticipate and maintain understanding of the events that take place in the process. Less experienced operators preferred less automated procedures which enabled keeping themselves in the loop.

The idea of Pirus to consider procedures as a form of automation draws attention to the tight connection between operators’ activity, automated operations, and procedure-guided operations. All these forms of operation function together to provide an appropriate process control performance. Operators, procedures, and automation form of a system that can be comprehended as a Joint Cognitive System [see further on the concept in 10] that is expected to act in an intelligent and adaptive way. It becomes evident from Pirus’s results that a shared awareness of the system’s performance is assumed by the human operator to anticipate the behaviour of the process and trust the joint functioning.

On the basis of the literature we may raise three major issues to be guide further research and design in procedure development, paper-based or computerised. These are: finding an optimal level of guidance and a way of using procedures so that 1) operations are tuned to situation specific constraints, 2) a possibility to anticipate process phenomena is supported, and 3) communication among the operators is assured.

Theoretical Approach of the Study

In the previous sections we have already referred to studies that in a relevant way describe the psychological demands that procedure usage assumes. As indicated, we agree with these ideas. We feel, however, that in these studies procedures have still been considered as clearly separate from expert acting. Our idea is that expert acting in complex and changing environments rely fundamentally on routines which may be corporeal and internal structures or also expressed in the form of external artefacts. Hence, procedures are one type of routines. Further forms of routines are the behavioural schemes or habits operators have developed for coping with the demands of their work. If we take this position to procedures, we need not to confront procedure-based and expertise-based acting. All acting relies on routines, and the challenge is to understand how routines are incorporated in experts’ acting. Strong guidance by routines is not a negative feature as such, but it is necessary that actors comprehend how procedures relate to process phenomena and that also acting is constrained by the same features of the environment, the process, as the procedures.

Becker [11] has reviewed the role of organizational routines in a way that has relevance to our attempt to understand procedures as acting. The author distinguished several features that define acting with routines. The most salient feature is that routines are patterns or regularities that may be identified in

individual and collective behaviour. It is also evident that routines are recurrent. Even when routines are rarely if ever executed, e.g. routines for evacuation a building, they may be considered as such because they are rehearsed. Becker also states that essential to routines are that they are shared in a community, not just individual ways of acting. Problems may arise if people do not act according to expectations built on these shared routines. An important further observation is that routines are typically dealt with from an “ostensive” point of view, i.e. routines label or name a particular task e.g. emergency operating. A “performative” point of view is less usual. It relates to actual analysis of how routines are practiced, i.e. how emergency operations are accomplished. Finally Becker points out that routines change according to past experience and that changes take place incrementally and locally.

Becker’s analysis supports our idea that routines are an intrinsic part in the structuring of action. *Habit* is the notion that has been proposed to express the way the organism is organised to meet the changing and unexpected features of the environment. The concept was first proposed by the philosopher Charles Sanders Peirce as a fundamental principle of human thinking and action [12]. According to him human actors connect themselves to the possibilities of the environment by continuous action-perception cycles, during which the outcomes of action are observed. As a result, an initial state of doubt about the environment is turned into a state of belief about the environment. These beliefs are habits that embody meanings related to certain situation or objects. The fundamental role of habit is to enable interpreting the cues of the environment and anticipating the effects of own acting. Hence, habits are a corporeal mechanism of anticipation. They are repeated due to being meaningful and as such they also may express style or ethos of acting. Habit enables general conclusions but as corporeal phenomena they do not assume use of symbols.

Peirce also proposed that habits have a semiotic structure. Habit expresses the principle of human thought, which allows some thing to be in some way substituted or represented by another thing. Signs, the form of which is heavily dependent on the technological medium applied, are used to represent objects, and understanding of this relationship becomes evident in an action, thought, emotion, or another act or behaviour. As a structure that communicates meaning habit enables organising other actors’ behaviour, too. The semiotic structure of habit may be used in the analysis of communication processes that result in shared understanding of the states of the environment. Procedures are crystallised forms of habit and their function, as that of habits, is to draw attention to significant signs and connect them to interpretative actions.

Research Questions

Our first research question is targeted to understanding the role of procedures in operator work on a generic level. As cited earlier, according to Dien [1] different stakeholders have diverse conceptions of why procedures need to be used. In this research we are interested in how operators themselves see the role of procedures in process control.

1. What is the role of procedures in process control in general?

Secondly, we also study procedure usage in practice by analysing how the role of procedures in operator work comes apparent in a specific situation. Also, we want to find out if same kind of effects of implementing new procedures as reported earlier (e.g., [3, 8]) can be identified.

2. What is the role of EOPs in the construction of activity in a loss of coolant accident (LOCA)?

METHODS

Collection of Data

The data used in this study was gathered in a test series that was conducted at the training simulator of the Loviisa nuclear power plant (NPP) in autumn 2008. The plant is owned by Fortum and is of PWR type. The aim of the whole test series was to generate baseline data to be used in the forthcoming control room evaluations. This was done by measuring performance and by recording the ways of acting in the present control room before the current hybrid control room with analogue panels and desks and process computers will be changed into a fully digitalised environment. In together 12 crews participated in the test series by running three accident scenarios.

In this paper we concentrate on a small part of the test series data. The intention is to demonstrate the analysis methodology that was developed while carrying out an in-depth analysis of one run by one crew. The developed methodology will be used in the future for the analysis of the rest of the data. Because of the small amount of data used in the analysis so far the results cannot yet be used for the evaluation of procedures or HSIs, but requires the analysis of a larger sample of the runs. The data used in the analysis that is presented in this paper consists of:

- orientation interviews of the operators (all 12 crews)
- video recordings of the run: one overview video and three head-mounted camera videos showing what each of the three operators was looking at.

We have had in use also a description of the scenario given by a simulator trainer, the emergency operating procedures used in the scenario, and process computer logs.

Analysis of Data

Analysis of Interview Data

All the orientation interviews were transcribed and a qualitative analysis was carried out. We looked specifically at the questions concerning procedures, but noted also if procedures came up in answers to other questions. The different conceptions concerning the role of procedures were identified and collected, as well as the conceptions concerning good procedures.

Analysis of Video Data

In this study the performance of one crew in one emergency situation was analyzed from the point of view of procedure usage. As a first step of analysis the emergency situation was considered and a functional situation model was constructed. This model depicts the generic functions of nuclear power production in the light of the specific emergency situation. The model also has a temporal dimension and it describes the main operations the operators are supposed to conduct in the situation. In the model the operations are connected to the functions. Thus the model describes the meaning of the operations. It makes visible what people do and for which operational purpose.

At the same time the video data was transcribed into a chronological worksheet. All the communications, operations, and sources of detection, movements etc. of the crew during the run were transcribed and written on to a worksheet in correct order. This sheet provides a course of action description

of the simulator run which is as complete as possible with the recording methods used.

Based on the functional situation model we selected some important *episodes* which were then carefully analysed from the course of action description. In this analysis a semiotic model of the structure of habit (Figure 1) was used. The model connects an object, its sign, and the interpretation. It can be used in analyzing activity on a micro level. We used the model in analyzing what signs of environment (the control room) the operators observed and how they interpreted these signs in their actions and communications. By carefully analyzing these action-perception cycles we were able to identify the objects to which the actions were directed, and thus reason about the meaning of activity.

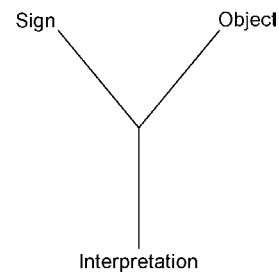


Figure 1. The semiotic model of the structure of habit.

RESULTS

Role of Procedures in Process Control

When the operators were asked about the role of procedures in process control different kinds of conceptions emerged. The answers could be classified into two qualitatively different classes. Below they have been labelled as “procedures as protection” and “procedures as basis of all process control”. The classes will be further elaborated below.

Procedures as Protection

When procedures are considered protection, operators refer to them as help and support in difficult situations. Answers in this class claim that procedures are good because they provide help when one does not remember something and that the procedures are used only when rare operations are conducted. Also the notion that procedures are not used in normal process control belongs to this class.

Characteristic for this class of conceptions was that procedure usage was contradicted with professional skill. This interpretation is supported by individual statements that claim e.g. that procedures are important for less experienced operators, and that procedures reduce operating errors and that they reduce operators’ responsibility.

Procedures as Basis of All Process Control

Another distinctively different class of conceptions was that of considering procedures as a basis of all process control. There were statements from the operators saying that all process control is conducted with the guidance of procedures and that procedures are used although they are not held in hand or looked at all the time.

Also the operators claimed that the role of procedures is to unify operating activity and that the use of operations guarantee the safety of operations.

It seems that the conceptions considering procedures as a basis of all process control claim that using procedures is something very fundamental to process control and that it is part of the professional skills to know the procedures and how to use them.

The Role of EOPs in the Construction of Activity in a Loss of Coolant Accident

For understanding the actual practice of using procedures we carried out an in-depth analysis of the performance of one crew in one emergency situation. The analysed run simulated a loss of coolant accident (LOCA) with an additional failure of a plant protection signal. In this kind of accident the use of emergency operating procedures is mandatory. A simplified scenario description is written below:

In the beginning of the scenario the plant is running on full power. Some minutes after the start of the scenario there will be a leak of 50kg/s in one of the six primary loops. The level of the pressurizer will drop quickly. The leak endangers the mass balance on the primary circuit side of the plant.

Signal PP16 will go off but fails to release a plant protection chain that is related to the automatic isolation of the containment. Very soon plant protection signal PP21 goes off and starts auxiliary pumps to compensate the loss of water in the primary circuit.

The operators should detect the abnormalities in the process state and perform appropriate actions to verify that there will not be further damages. PP16 signal is a criterion for taking the incident identification procedure II into use and PP2 for taking the accident identification procedure AI. After detecting that the process is not in a stable state and that shutting down the plant is required the operators can perform scram (and verify automatic shut down) on the basis of either shift supervisor’s decision or II, or, if they have not released scram before noticing the PP21 signal, on the basis of AI.

Identification of the accident situation is performed with AI, which then leads the operators to the AL primary circuit leak procedure. With AL the crew runs the plant towards a safe state by cooling the primary circuit and by manually releasing the PP16 signal that finalises the isolation of the containment.

The episodes presented in the following subsections were selected on the basis of the functional situation model of the scenario. For the purposes of this paper we aimed at choosing episodes that show the connection of procedures to crew’s situation awareness, and procedures’ role in structuring activity.

In the episode descriptions some abbreviations are used: shift supervisor (SS), reactor operator (RO), turbine operator (TO), CR (control room), PMS (process monitoring system, that is used for monitoring but not for committing operations), PP (plant protection system), and QDS (qualified display system, that is used for example for monitoring the position of control rods).

Episode 1

The first episode (Table 1) covers a time frame from the detection of the first signs of failure to the point of time when SS makes a decision of first actions.

Table 1. Communication and activities in Episode 1.

Time	Communication	Description of activity
00:00	Alarm sound	All operators are looking at the alarm/event lists on their PMS monitors.
00:02	SS: “Whoops!” RO: “Well!” TO: “What’s happening?”	SS opens up a primary circuit process display on his PMS monitor. RO stands up.
00:08	SS: “How come PP has not launched?”	SS wonders why plant protection signals have not launched yet. SS looks at the PP panel. RO and TO look at their PMSs.
00:09	RO: “II”.	RO notices the II criterion display on his monitor and notifies SS of it. SS does not respond at once (it is unclear if he has heard) but opens up a trend on his PMS. RO takes the II criterion procedure and TO the scram procedure.
00:15	SS: “The pressurizer level is dropping a lot”.	SS informs others of the change in a process parameter. RO turns gaze on a trend on his PMS monitor and then looks at the PP panel.
00:16	SS: “II”.	SS looks at the II criterion procedure and tells the operators that II should be taken into use.
00:22	SS: “A leak”.	SS is suspecting a leak. SS turns around to take the II procedure folder. RO is looking at the PP panel and TO the scram procedure.
00:35	SS: “Wait a second!”	SS opens up a display showing the containment. TO stands next to the SS’s table and waits for SS to give him the II procedure. RO is looking at the QDS.
00:45	SS: “Go on, scram the reactor”.	SS looks at RO and gives him a command to release scram. RO pushes the scram button and starts verifying the scram. TO returns to his place and starts verifying turbine scram.
00:48 ... 02:00	SS: “And verify scram”	TO and RO perform verification operations according to the scram procedure. SS places the II procedures on his table.

The data of the first episode shows that the operators are quickly able to create a good understanding of the situation and to take correct counter measures. This takes place without the direct guidance of the II procedure. Already from the first signs SS is able to anticipate the launching of plant protection chains which shows that early on he understands that a fairly serious event is going on. Information search and constructing understanding of the event is a collective effort, in which SS takes the leading role by announcing the initial diagnosis of a leak, and by commanding the first operations. It has to be noted that at this point of time the operators cannot be completely certain of the type of the event since the initial symptoms can refer to for instance a primary loop leak, a steam leak, or a primary-secondary leak.

The crew does not need the II procedure for making the diagnosis or for deciding about the first actions, but detecting the release of the II criterion and their knowledge of the contents of the II procedure (in which one of the first actions is to release scram) strengthen their diagnosis and confirm that the plant needs to be shut down. The crew has a common understanding that there is an “II situation” going on. In this episode the II procedure supports the creation of situational understanding although not any of the crew members has held it in hand yet.

We analysed the first episode by utilising the semiotic model of the structure of habit (see Figure 1). Table 2 shows the signs that were identified by the crew and the interpretations they made of these signs in their actions and communications. Based on the functional situation model of the scenario we identified the objects to which the actions were directed.

Table 2. Signs, interpretations, and objects in Episode 1.

Sign	Interpretation	Object
Alarms (e.g. ice condenser hatches)		
Single process parameters (e.g. pressurizer level)	Searching for more information	
Plant protection signal PP16 = II initiation criterion goes off, II criterion display opens	II procedure taken into use	Process situation is not stable
	Scram	There is an incident
		A leak, more cooling needed

Episode 2

During the second episode (Table 3) the crew should identify an “out-of-the-scenario” failure, i.e. a failure that is not caused by the loss of coolant accident. In this scenario there is an additional failure in the automatic release of one plant protection chain PP1. The episode lasts over 15 minutes starting from the activation of the PP1 signal and ending when the PP1 key is turned and the protection chain is manually launched.

Table 3. Communication and activities in Episode 2.

Time	Communication	Description of activity
00:09		SS looks at the PP panel, but the PP16 lamp is not burning yet. SS anticipates the activation of the PP16 signal.
00:12	RO: “II”.	RO notices the II criterion display emerging on his monitor (due to the activation of PP16). He notifies SS and takes the II criterion procedure.
00:15		RO turns to look at the PP panel. Now the activation of PP1 is visible and the lamp burns, but not the lamps below it that would indicate the automatic release of the protection chain. RO does not seem to detect the failure.
02:04		RO takes II into use and starts to perform checks. It is one of the first tasks in the II procedures of RO and SS to check the PP16 signal.
02:16	RO: “There are AP-pumps running”	RO notices PP21 signal (that starts AP auxiliary pumps) which is an initiation criterion of the AI procedure. The attention of RO and SS turns from II to AI.
03:07		SS goes to the PP panel to check the PP21 signal according to the AI procedure.
03:47		TO reads AI and looks at the PP panel. He is checking a group of PP70 signals according to the AI procedure.
06:43	RO: “Plant protection system checked”	RO informs SS that he has checked the plant protection system according to AI procedure (signal PP21, and some others).
09:54		SS double-checks the plant protection system.
15:52		RO walks to the PP panel reading the AL procedure. He immediately turns the PP11 key, then reads his procedure, turns key PP13, reads the procedure, and continues the same way with keys 14, 15, and 16.

Already in the beginning of the episode SS anticipates that the PP16 plant protection signal should soon go off. When this happens RO looks towards the PP panel but fails to notice that while the PP16 lamp is burning, the lamps below it are not burning: therefore, the automation has not released the plant protection chain and the containment isolation is not fully completed. This is a failure in the automation system. PP16 signal is an initiation criterion for II emergency operating procedure. The activation of PP16 is a sign for the crew that a fairly severe incident is going on and that II procedure should be taken into use. After seeing the PP16 signal the operators should also verify the functioning of the automatic plant protection chain, but this meaning is not communicated. From this point forward the crew’s understanding of the situation is slightly impaired. The crew seems to assume that automation is working as it should although this is not the case. In several occasions the crew checks various PP signals, also PP16, but fails to notice the unreleased chain. When RO turns the PP11-16 keys according to the AL procedure he does not verify the realization of the operations by looking at the signal lamps. At this point he actually releases the protection chain but seems not be aware of it. RO turns the keys because the AL procedure says so, but does not think of the meaning of the operation.

Although checking the release of PP16 signal is an action point both in the II criterion release procedure and in the II emergency operating procedure, the crew does not detect the unreleased PP16 plant protection chain. In the AL procedure the wording “PP11, PP13, PP14, PP15 and PP16 manual release keys are turned to horizontal position ON” does not seem to communicate the meaning of the operation: making certain that the containment is completely isolated by securing the functioning of automation.

The sign the crew detected and the interpretation and object related to it are marked with black colour in Table 4. It seems that the crew assumes the automatic plant protection is functioning (object), but they do not think of what the plant protection is supposed to do in this situation. Also it seems that they turn the PP keys only because the procedure says so (interpretation). Grey colour marks the sign the crew did not detect, the interpretations the crew might have made and the objects they might have pursued on the basis of this sign.

Table 4. Signs, interpretations, and objects in Episode 2.

Sign	Interpretation	Object
PP16 signal goes off		
PP16 signal goes off but the protection chain not launched	Turning of PP keys when following the procedure	Automatic plant protection
	Manual backup of automation following the procedure	Automatic containment isolation
	Manual release of the protection chain	A leak, containment isolation needed

Episode 3

In the third episode (Table 5) the crew should detect that the situation is deteriorating; the leak is so large that the auxiliary pumps start to compensate the loss of water. The situation turns from an incident to an accident.

Table 5. Communication and activities in Episode 3.

Time	Communication	Description of activity
02:16	RO: "There are AP-pumps running"	While following II, RO notices that PP2 lamp in the desk is blinking. PP21 signal starts pumps that feed more water to the primary circuit. RO turns towards the left side panel and then towards SS. RO informs SS.
02:19	SS: "AP?"	SS asks confirmation.
02:20	RO: "Yes"	RO answers and looks at the II procedure.
02:21	SS: "Is it PP21?"	SS asks RO if PP21 has gone off.
02:22	RO: "Yes"	RO confirms.
02:23	SS: "OK. Let's take AI into use"	SS tells the operators that AI procedures will be taken into use. PP21 is an initiation criterion of the AI procedure. TO is talking on the phone and is not listening.
02:25		SS takes the AI procedure folder from behind his table.
02:44	SS: "Hey!"	SS tries to get the attention of TO. TO walks towards SS's table.
02:45	SS: "PP21"	SS tells TO why the AI procedure was taken into use.

In the beginning of this episode the crew is performing checks according to the II procedure. When RO is returning from the front panel to his place he notices that the PP21 signal lamp in the desk is blinking. Immediately he turns to look at the left side panel where the AP pump symbols are displayed. Checking the PP21 signal or the pumps are not required in the II procedure which is an indication that RO is performing also other monitoring tasks than only those written in the procedure.

The going off of PP21 signal, which is an initiation criterion of the AI procedure, is a clear sign to which the crew reacts by swiftly taking the AI procedure into use. RO and SS are making the interpretation of the situation together, and SS also informs TO of the reasons of the decision. Having understood the situation the crew concentrates on performing the procedure. Starting of AP pumps (emergency auxiliary water) should tell the operators of the imbalance of mass in the primary circuit: the leak in the primary circuit is so large that as a consequence more cooling is needed. However, at this point of time the operators discuss very little about the nature of the event. Neither do they say aloud that the event has grown more severe i.e. turned from an incident into an accident.

It is likely that at this point of time (about 3 minutes from the start of the scenario) the crew's diagnosis of the situation is already quite accurate. In the AI procedure the diagnosis phase comes considerably later. At about 11 minutes SS reaches the end of AI and has thus completed the identification of the emergency situation: the situation is not a primary-secondary leak or a steam leak, but a leak in the primary circuit (LOCA).

Table 6 presents again with black colour the signs, interpretations, and objects which came true in the episode. Grey colour marks signs, interpretations, and objects that could have realised. The data shows that right after detecting PP21 the crew focused strongly on performing the procedure. They did not bring up the goal of AI (making a definite identification of the emergency situation) or consider the severity of the situation.

Table 6. Signs, interpretations, and objects in Episode 3.

Sign	Interpretation	Object
AI criterion display opens up		
Plant protection signal PP21 goes off	AI procedure taken into use	
Auxiliary systems (AP pumps) start	Verification of the diagnosis of the emergency situation	"AI situation" More severe situation, a large leak

Episode 4

In episode 4 (Table 7) the operators come to the end of AI which leads to AL. The episode illustrates how operators perform actions steps and make decisions based on the procedure.

Table 7. Communication and activities in Episode 5.

Time	Communication	Description of activity
10:31	SS: "And five minutes have passed"	SS checks an action step in AI. 5 minutes have passed since scram.
10:32	RO: "Yes it has"	RO confirms
10:36	SS: "And PP31-37 have not come through."	SS is holding the AI procedure
10:41		SS walks to the PP panel to check the signals again.
10:48	SS: "Have not, pressure is under yes"	SS reads aloud the action steps: PP31-37 signals have not gone off and primary circuit pressure is under 110 bar. He has asked RO about the pressure earlier.
10:50	SS: "Let's take AL into use."	SS gives a command to take AL into use.

SS reads aloud the action steps in the AI procedure and reaches the end of the AI procedure. He does not ask the operators if they have come into the same conclusion about where they should proceed next (from AI to AL). Reaching the end of AI means that the emergency situation has now been diagnosed: it is a loss of coolant accident. SS does not say aloud this conclusion, nor does he say what the next objectives are. This is maybe because the operators have already in the beginning come into the same conclusion of the situation and the SS sees it unnecessary to repeat it. At this point of the scenario procedure usage seems to be quite mechanical. Each operator performs independently the actions required by the procedure and communication is limited almost only to those points where the procedure literally requires it. Table 8 presents the signs, interpretations, and objects in the episode.

Table 8. Signs, interpretations, and objects in Episode 4.

Sign	Interpretation	Object
Actions steps lead to AL	AL taken into use	"AL situation"
Parameter values indicate a large leak in the primary circuit	Cooling and containment isolation	Stabilization of the plant to a safe state

Conclusions Based on the Episodes

In the beginning of the scenario, in episode 1, the crew was able to form an initial diagnosis and carry out correct actions without the direct help of procedures. Instead, the ability to make the identification was based on their knowledge and experience. Noticing the opening of the II criterion display and knowing the first action in the II procedure is releasing scram probably strengthened the operators' own reasoning about the nature of the situation, its severity, and the required actions. The operators acted according to the II procedure even though they did not have the procedure in their hands yet. Episode 1 shows the procedures supported decision making and that taking procedures into use did not interrupt the course of activities.

After taking the II procedure into use, in episode 3, the crew detected signs (PP21 signal and the starting of auxiliary pumps) the II procedure does not directly refer to. This indicates that at this point of time the operators' behaviour was well adapted to the situation and that in addition to performing procedures the operators were able to concentrate also on overall process status observation. It seemed that the operators' had already quite an accurate understanding of the situation, except the PP16 signal they had not noticed even though checking it is an action step in II (see episode 2). The crew had probably been able identify the type of the emergency situation before they had reached the diagnosis phase in the AI procedure. It is shown that constructing understanding of the situation does not go completely hand in hand with the progression of the procedure. However, the AI procedure probably provided them support for example in locating the leak. Detecting the initiation criteria of the AI procedure however drew the operators' attention, and without discussing much the nature of the situation or the reasons for changing procedures they went forward on using AI.

In episode 4 and in the remainder of the data (not shown in the selected episodes), it seems that when using AI and AL the crew focused more and more on performing procedures i.e. the procedures became their object. Operators worked separately and communication took place when required by the procedures. At certain points of time the operators however considered together the situation. This took place for example when the operators had to wait for some process parameter to reach a certain limit before continuing. These kinds of stops in the flow of actions are important from the point of view of constructing common understanding, and they provide a possibility for synchronising activity. It would be important to make use of these situations to orient to the goals and next actions. In the case of the studied crew this was not always done for example when turning procedures' pages.

The first episode shows an example of "intelligent use" of procedures (the term "intelligent use" explained in 2.1.2). The operators' competence and procedures' guidance combined well: the operators identified the process changes by themselves and the procedure supported them. The operators were able to maintain control over their own actions: they themselves dictated the course of actions, not solely the procedure. Also reactor operator's behaviour in the second episode shows proof of intelligent procedure use: his actions were not limited to those required in the procedure, but he continued observing the overall state of the process and detected the change in the situation.

In the first episode the operators were able to connect the situation (PP16 going off and other signs of failure) correctly with the procedure (II). However, they failed to notice the additional failure in the PP16 signal which they might have detected if they would have verified the functioning of automation and thought of the endangered functions in a loss of coolant accident (one of which is containment isolation).

Episodes 1 and 3 show the shift supervisor's skills in organising efficient collaboration and procedures' positive effect on it. Detection of the initiation criteria of procedures supported the shift supervisor in deciding what to do, and the operators could start performing required measures quickly. Episode 4 and the remainder of the data bring up some deficiencies in collaboration and in the coherence of acting later in the scenario. The reactor operator and the turbine operator were a little faster in going through the procedures, and due to falling behind with double-checking the shift supervisor carried out some of his tasks quite hurriedly. Also, he was sometimes interrupted by the other operators with notifications of completed action steps, and was not always able to remember what the other operators had already told him.

DISCUSSION

According to the results of the orientation interview the operators had differing conceptions of the role of procedures in process control work. It should be studied if these lines of thought are shared also by other groups in the plant since the conceptions influence e.g. procedure development, training, and HSI design. We see that using procedures should not be contrasted with professional skills but rather mastering them should be seen as an inseparable part of craftsmanship. It should be carefully thought of how the role of procedures is considered in training.

According to the design philosophy of symptom-based procedures it is not required that the operators make an exact identification of the situation but that they carry out operations procedure developers have defined. Also the aim is to standardise the ways of acting of crews. Our results indicate that the studied crew indeed carried out all action steps. According to an expert evaluation the overall performance level of the crew was good indicating that the crew succeeded well with the symptom-based procedures. The results show that in the beginning of the event the crew was well adapted to the situation and that procedure usage fitted seamlessly to the course of operator activities, i.e. procedures' level of guidance was appropriate. However, the results showed also that cooperation deteriorated somewhat when the situation evolved and that even though the crew used procedures they did not detect all important information. Although goals are mentioned on each page of the procedures, it seems that the procedures could have directed operators' attention even better to the endangered plant functions and to the meaning of operations.

In this paper we present the result of the analysis of one crew's behaviour in a simulated emergency scenario. Based on this material it is not yet possible to evaluate the quality of the procedures or to make conclusions of the procedure usage of all crews. In the future, the same usage-based methodology will be used for analysing more runs and for evaluating the procedures and HSIs.

ACKNOWLEDGMENTS

We want to thank our colleagues at VTT Jari Laarni, Hanna Koskinen, Marja Liinasuo, Iina Aaltonen, and Hannu Karvonen who have participated in the planning, data collection, and analysis of the test series. We also want to thank our collaborators at Fortum Loviisa nuclear power plant and Fortum Nuclear Services.

REFERENCES

- [1] Dien, Y. 1998. Safety and application of procedures, or how do "they" have to use operating procedures in nuclear power plants. *Safety Science* 29(3), 179–187.
- [2] Perin, C. 2005. *Shouldering risks. The culture of control in the nuclear power industry.* Princeton University Press, Princeton, NJ.
- [3] Filippi, G. 2006. Studying computerized emergency operating procedures to evaluate the impact of strong procedure guidance on operators' work practices. In: *Proceedings of the American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human-Machine Interface Technologies.* Albuquerque, NM, November 12–16, 2006.
- [4] Roth, E. M., Mumaw, R. J. and Lewis, P. M. 1994. Empirical investigation of operator performance in cognitively demanding simulated emergencies. Technical Report No. NUREG/CR-6208. Nuclear Regulatory Commission, Washington, DC.
- [5] O'Hara, J. M., Higgins, J. C., Stubler, W. F. and Kramer, J. 2000. Computer-based procedures systems: Technical basis and human factors review guidance. Technical Report No. NUREG/CR-6634. Nuclear Regulatory Commission, Washington, DC.
- [6] Nuutinen, M. 2006. Expert identity in development of core-task oriented working practices for mastering demanding situations. Six empirical studies on operator work. Doctoral Thesis. University of Helsinki, Helsinki, Finland.
- [7] Throneburg, E. B. and Jones, J. M. Situation awareness and trust in computer-based procedures in nuclear power plant operations. In: *Proceedings of the American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human-Machine Interface Technologies.* Albuquerque, NM, November 12–16, 2006.
- [8] Roth, E. and O'Hara, J. 2002. Integrating digital and conventional human-system interfaces: Lessons learned from a control room modernization program. Technical Report No. NUREG/CR-6749. Nuclear Regulatory Commission, Washington, DC.
- [9] Pirus, D. 2004. How functional approach allows to set on-line flexible level of automation for NPP's computerized operation. In: *Proceedings of the American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human-Machine Interface Technologies.* Columbus, OH, September 19–22, 2004.
- [10] Woods, D. and Hollnagel, E. 2006. *Joint cognitive systems – patterns in cognitive systems engineering.* Taylor & Francis, Boca Raton, FL.
- [11] Becker, M. C. 2004. Organizational routines: a review of the literature. *Industrial and Corporate Change*, 13(4), 643–677.
- [12] Peirce, C. S. 1998. *Harvard Lectures on Pragmatism.* In: *The Peirce edition project (Ed.). The essential Peirce. Selected philosophical writings.* Indiana University Press, Bloomington and Indianapolis, 133–241.

Improving Safety Through the Design of a New Function: Design Process Findings and Post-release Analyses

Thomas Koester
FORCE Technology
Division for Maritime Industry
Hjortekaersvej 99
DK-2800 Kgs. Lyngby,
Denmark
tsk@force.dk

Nikolaj Hyll
FORCE Technology
Division for Maritime Industry
Hjortekaersvej 99
DK-2800 Kgs. Lyngby,
Denmark
nnh@force.dk

Jan Stage
Aalborg University
Department of Computer
Science
Selma Lagerlöfs Vej 300
DK-9220 Aalborg East,
Denmark
jans@cs.aau.dk

ABSTRACT

This paper illustrates how knowledge from field studies, domain knowledge and psychological knowledge was combined in a design process leading to the invention of a new function in a maritime communication device. The function compensates for cognitive problems and limitations related to perception, memory and attention which are relevant in the interaction between the user and the device and in the given context of use – the bridge on a ship. It is shown how this has a positive effect on safety by reducing the potentials for error.

Keywords

design process, cognitive psychology, perception, persistence, disruptions, attention, prospective memory, post-release analysis

ACM Classification Keywords

H.1.2 [Models and Principles] User/Machine Systems – *human factors*; H.5.2 [Information Interfaces and Presentation] User Interfaces – *user-centered design*.

INTRODUCTION

In 2005 a new maritime VHF-radio was launched to the market, and it is now used on board ships world wide. The radio had a brand new function: The possibility of replaying an incoming message in the case it was not heard the first time. This function was the result of a design process where knowledge from field studies, domain knowledge and psychological knowledge was combined. The idea with the new function was that it should improve usability and enhance maritime safety. This was to a wide extent based on the scenario that the user missed a call, did not hear all of it or was unable to record (on paper) or recall (from own memory) important information from the call. The new function was used as one of the key arguments in the marketing material for the radio.

During the design process two problems were identified: (1) Problems related to perception and (2) problems related to

persistence. In the post-release analysis one additional problem was identified: The problem about disruptions caused by incoming calls. This paper describes the findings in the design process and the post-release analysis. The whole process can be considered a learning process where knowledge about the device, the users, the work context and the use of the device gradually is building up. One very important consequence of this learning process was the invention and development of the replay function. The paper is focused around the findings and analyses regarding the replay function, but other improvements and inventions in the design such as the dual-display interface could have been subject to similar analyses. The objective of this paper has been to present how the outcome of detailed ethnographic studies of a work task in combination with psychological knowledge can be used in the design of a communication device that contributes to increased safety (decreased error potential) in that work task.

The Maritime Work Domain

Practically all merchant and navy vessels and many fishing vessels have at least one VHF-radio on the bridge of the ship. The VHF (Very High Frequency) is used for communication between ships and between ships and land based stations such as pilot stations, harbour authorities and Vessel Traffic Service (VTS) centres. VHF-radios can also be used for internal communication on board the ship e.g. between crew working on the bridge and the crew working on the deck of the ship. The VHF-radio is one – but very important – piece of technological equipment among others such as radar, electronic charts, echo sounder and gyro compass. It is used for communication about ships interaction (e.g. how to manoeuvre in situations when there is a danger of collision or how to interact with tugs), about weather and sea conditions, fishing banks, navigational problems, taking pilot on board, going in and out of harbour, passing bridges, negotiating passage of other vessels or superstructures, and a number of safety related issues such as emergency situations, dangers, coordination of search and rescue operations etc. It has also been used for social conversation and through land stations for telecommunication with people in land. This particular use has lately decreased dramatically: The VHF has for this purpose more or less been replaced by the mobile phone. There is an in intensity variable stream of calls on the VHF. Some of them are important or even urgent seen from the perspective of a specific ship. And some of them are more or less just “back ground noise”. Common to all is that they – except from the calls made by the crew themselves and the replies to these – appear uncontrollably, however sometimes expected or anticipated. The crew can mute the radio to prevent disruptions and disturbances by

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

incoming calls, but in practice this is normally not done because then there is the risk of missing an important call.

Apart from using the VHF for communication the crew on the bridge has a lot of duties to perform. They will navigate the vessel safely, avoid collision with other vessels, control the maintenance of their vessel, maintain and keep record of the cargo, passengers etc., fight fires when necessary, perform work such as fishing, under water construction work and administrative duties on board the ship.

The scene of the ship's bridge can be reproduced with high levels of realism in a maritime full mission simulator. The whole bridge layout including instruments and equipment is reproduced in a scale 1 to 1. The view out of the windows of the bridge is reproduced as computer generated graphics on big screens.

The crew on the bridge of the ship and their interaction with the available technology has been subject to studies in the combined paradigms of psychology, technology and safety. Examples are Lützhöft [19], Lützhöft and Dekker [20] and Schager [27] studying the crew's use of technology on the bridge and Porathe [26] studying one particular type of equipment – the electronic sea charts – and the psychological implications involved in the use of this type of equipment. Hutchins has in the book "Cognition in the wild" [13] studied cognitive processes related to the operation and navigation of ships, and Van Westrenen [31] has made intensive studies of the maritime pilot at work with special focus on the pilot's mental workload. There is in other words a tradition of safety related research in both maritime work and the interaction with maritime user interfaces, and parts of this tradition are based on psychological knowledge, theories and methods. This paper follows the tradition by looking at the design of the VHF and how the design relates to psychological factors and maritime safety.

THE DESIGN PROCESS

The process of designing the new maritime VHF-radio was based on the assignment formulated by the Danish producer: To develop a new product with the highest focus on usability and improvement of maritime safety. The design process was planned with inspiration from the ISO 13407 on "Human-centred design processes for interactive systems" [14]. The producer engaged a Danish design company with industrial designers and a Danish human factors specialist, a psychologist with knowledge and experience from maritime human factors. The designers and the human factors specialist formed the design team together with product developers, engineers, software programmers, salesmen etc. from the producer's organisation. The configuration of this team was therefore compatible with the objective in section 5.5 in ISO 13407 [14].



Figure 1. The design team made interviews with end users on board small and large fishing vessels, ferries and supply ships. Captain on board the shrimp trawler Nelly explains how he is using his VHF-radio.



Figure 2. The designer (right) interviewing the captain (left) on the fishing boat Shannon about the work context in general.



Figure 3. Captain on the supply-ship Skandi Fjord talks about tactile feed-back and shows how he is using existing VHF-equipment.

The design process included all the traditional components: (1) Data collection through ethnographic field observations and informal interviews with end users (see Figures 1, 2 and 3), technicians and salesmen (corresponding to section 7.2 and 7.3 in ISO 13407 [14]), (2) analysis of data from the data collection and development of a preliminary design (section 7.4.2), (3) development of a mock-up and a set of use scenarios for usability testing (section 7.4.3), and (4) the iterative process of usability testing using the mock-up, use scenarios and a verbal protocol (thinking aloud) (section 7.4.4) and adjustment of the design accordingly (section 7.4.5). The testing was made partially in a maritime full mission simulator. Further, in relation to (1) and (2) a design seminar for all participating parties was planned. The objective of the design seminar was to give the participants fundamental knowledge and insight in relevant theories and methods from the human factors and usability disciplines; a common platform for discussion and – through a ship simulator scenario where the design team members acted as navigators on the bridge (even though many of them did not have any maritime education) – a common experience with the maritime work domain and context of use of the VHF-radio.

Perceptual Problems and Problems Related to Persistence

The analysis of data from the data collection process indicated a number of safety problems related the use of the VHF and showed that there was error potentials related to the situation where an incoming message was not heard due to background noise, inadequate sound quality or language difficulties. This was mainly related to human senses and perception – and to the finding that the work environment in which the VHF is used can be rather noisy and the sound quality for different reasons can be poor. These problems are also described in [17]. Since

the message could be important or urgent e.g. a navigational warning or an emergency call from another vessel or it could just be relevant for the safe navigation of the ship for example messages from tugboats, pilot boat or other vessels in close encounter situations, these perceptual problems could in worst case lead to safety problems.

Further, situations were identified where the message was heard but there were later problems recalling (from human memory) or recording (on paper) information from the message. These problems are related to persistence as described in [17]: The information in the message is transient – it is only present while the message is broadcasted in the radio. After the broadcast has ended it only exists in the memory of the listener or perhaps on paper if it was written down. The problems related to recalling or recording information from the message is of course only significant if the information is relevant or important for the specific ship.

Even if the message is heard and the content stored in the memory of the listener it is not guaranteed. Information kept in the human memory is not permanent and the memory has limited capacity. The memory of messages broadcasted on the VHF can be described using Baddeley's theory of the phonological loop. Behind the theory of the phonological loop lies Baddeley's working memory paradigm for understanding memory. This paradigm is still strong within cognitive psychology. This memory system has three components: (1) A modality-free central executive resembling attention, (2) an articulatory loop (known as phonological loop) that holds information in speech-based form and (3) a visuo-spatial scratch pad (now called visuo-spatial sketchpad) that is specialized for spatial and/or visual encoding [3].

The phonological loop is a slave system for the central executive and must when receiving a spoken message on the VHF radio preserve the order of the words and the content of the message. If the auditory message is kept in memory long enough and if the amount of information in it is limited it can be recorded by writing it on paper. Notepads, pieces of paper, post-its etc. used for this purpose can be called cognitive artefacts [25, p. 17]. The use of cognitive artefacts affects human cognitive performance: By writing down information from the message the performance is improved because the information content is less vulnerable; it can be kept for a longer time without deterioration or loss of information – of course as long as the piece of paper remains intact and available.

The Replay Function

These two findings, the perceptual problems and the problems related to human memory and persistence, indicated that the user could benefit from a replay function where it was possible to replay a message in case it was not heard, only heard partially or if important information from the message could not be recalled later or was not recorded for example on paper. The replay function should compensate for the potential problems. It was therefore proposed as a new function by the human factors specialist and integrated in the design by the rest of the design team. The replay function was developed not on the basis of user input but on the basis of the combination of psychological knowledge, domain knowledge and knowledge from field studies. The philosophy behind the function was to increase maritime safety by reducing error potentials. The idea was that it in a very simple way was possible to replay the last incoming message but also that it was possible through a menu to replay older messages based on their time, channel number and duration. In total 90 seconds of incoming messages is stored in the memory in the radio. Since each message usually

is rather short – only a few seconds in duration – a number of messages are stored in the memory, and it will typically be possible to replay the last 10-20 messages. Old messages are overwritten by new from the principal of first-in first-out. The replay function can be used for two purposes compensating for the identified problems related to perception and persistence: (1) It can be used to repeat messages making it possible for the radio user to control environmental variables affecting the perception (for example move closer to the radio, adjust the volume, and eliminate background noises) and (2) it could act as non-human phonological loop and as an effective cognitive artefact and thereby enhance the performance of the radio user. It can therefore – if used – have a potential positive effect on safety.

Validating the Design

Part of the design process was a validation of the new design and the replay function. The aim was to identify possible adjustments and improvements of the design through a user test.



Figure 4. The primitive mock-up made of a print of the front design glued to a piece of cardboard. Changes in the display during interaction were illustrated through exchangeable display cards (the pile to the right in the picture).

The design team made a very simple cardboard mock-up (see Figure 4), and the use of the replay function was tested through realistic use scenarios (based on data from the data collection and general domain knowledge) in a maritime simulator using the verbal protocol of thinking aloud (see Figure 5). The method corresponds to the method used by Kjeldskov and Stage [16] when they tested a mobile device for communication. They used a more sophisticated mock-up, but they also used a ship simulator, the thinking aloud protocol and trained maritime officers (ibid. section 6.2). A total of six experienced captains participated in the evaluation of the VHF design, and the input from the sessions were analysed and considered in the adjustment of the design.



Figure 5. The mock-up is mounted on the bridge in a ship simulator. The captain standing to the left and the designer are discussing the design.

The response from the four test subjects was that they found the function very usable in situations where an incoming message was not heard or where information from a message was lost due to the lack of persistence. Other safety related issues and usability issues of the new design was tested in the simulator together with the replay function, and the feedback from the subjects indicated where to make adjustments in the design in the next iteration. The conclusion from the usability experiments was that the replay function had a safety enhancing potential, and it was therefore integrated in the final design launched on the market in 2005.

The marketing material presented the VHF-radio with emphasises on the replay function as a new and revolutionary function with great potentials for improvement of safety and usability: "As a world first, the SAILOR RT5022 VHF introduces the Replay function for improved communication and safety at sea. Push the Replay button and the SAILOR RT5022 VHF will replay the important messages you may just have missed." [28].

POST-RELEASE ANALYSIS

The story about the replay function could stop here with the conclusion that (1) the combination of psychological knowledge about perception and memory combined with observations from field studies and domain knowledge about the work environment in which the device is used had lead to (2) the invention of a replay function which was (3) validated through normal usability studies in a maritime simulator, and which had (4) potentials for enhancement of maritime safety. However this is not the end of story or the conclusion. We have now – when the device has been on the market for more than 3 years – made a pilot study like post-release analysis of the replay function. Post-release analysis is for example described in the previously mentioned ISO 13407 under the headline "long-term monitoring" [14]. Our analysis had two components: (1) An analysis based on concepts and theories from cognitive psychology and (2) interviews with end users of the VHF-radio. In the analysis we have focused on problems with the VHF-radio related to the *disruptive or disturbing nature of incoming calls*.

Disruptions and Disturbances on the Bridge

Prioritizing tasks (task management) correctly in a complex scenario can be critical to the safety. Inspired by the so called ANCS hierarchy [32] we can say that the priority of tasks in the safe operation of the vessel at sea is dependent on giving each task the right priority in the following order (1=highest priority): (1) Handling the vessel (including keeping it afloat), (2) navigation (getting safely to the desired destination) and (3) communication (coordination with the system outside the vessel). Analyses of aviation accidents and incidents have shown that safety is compromised when tasks of higher importance are superseded by those of lower importance (ibid.). This is exactly what happens when a VHF radio call is interrupting – and thereby disturbing – a primary task – where the primary task for example can be turning the ship or giving way to another vessel according to the rules.

It is worth noting that analyses in aviation often shows that (3) communication often interferes with (2) navigation which is primarily a visual task. Generally visual monitoring has been found to be extremely vulnerable to interruptions. Many studies of interruptions of pilots where an ongoing task is terminated by the arrival of a new task have been made [5, 7, 9, 22]. There seems to be a general tendency for auditory pre-

emption. One possible explanation for this is that auditory onsets are inherently (and biologically) attention-grabbing (newborn babies can turn their head in the correct direction of a sound). Another reason for the tendency for auditory pre-emption is the limitation of the working memory. This limitation is thought to prompt people to deal with an auditory message or communication immediately when it arrives. In order not to forget it (lose it from their working memory) they will keep their attention directed towards it until it is completed [18].

Interruptions can be described as "an external intrusion of a secondary, unplanned, and unexpected task, which leads to a discontinuity in task performance" [4]. Many studies have examined the effects of interruptions and found them to be disruptive to the performance of a primary task [11, 23, 24, 30]. No comprehensive theory of interrupted task performance currently exists but Altmann and Trafton [1] propose the goal-activation model stating that disruption is greater the longer the duration of the interruption and the less a person rehearses the primary task during the interruption. Further more, the timeline of an interruption can be seen as: primary task – alert for secondary task – starting on secondary task – end of secondary task – resume primary task. Interruption lag is the time between the alert for the secondary task till the officer starts to work on it and resumption lag is from the end of the secondary task till the primary task is resumed [29].

There are many sources of auditory disturbances on the bridge of a ship such as alarms, telephones, faxes, guests on the bridge and of cause calls on the VHF-radio. One way of reducing the disturbances from the VHF is to turn it off. But then there is the risk of missing important or urgent information. Another option is to listen to the messages – not at the time when they come in, but at a time when it is appropriate taking the other tasks performed on the bridge into consideration. The replay function makes this last option possible to some extent. When the crew on the bridge of the ship is disturbed or interrupted in a work task it can cause certain cognitive problems – and thereby compromise safety. Here we will look at the problems related to attention and prospective memory.

Attention

Knowledge of attention and performance limitation is of great importance. The literature tells us that attention is referring to selectivity of processing. Early in the history of psychology William James described attention as "...the taking possession in the mind [...] of one out of what seem several simultaneously possible objects... [...] Focalisation, concentration, of consciousness are its essence" [15, p. 403–404]. His point is that attention not necessarily has to be conscious or that the operator of a VHF radio can be more or less conscious of a certain stimulus. There is a difference between listening to or noticing a call on the VHF- radio and actually hearing what is said or part of what is said (understanding the message partly or fully). Baars [2] gives the example: "We listen in order to hear" (p. 364) – an important distinction between selecting an experience and actually being conscious of the selected event. Selecting a channel of the VHF for monitoring is not the same as being conscious of all that is being or has been transmitted on that channel!

Attention can be divided in *focused attention* and *divided attention*. Researchers studying people presented to two or more stimulus inputs at a time while instructing them to respond only to one (focused attention) and people presented to two or more stimulus instructed to attend to all stimulus inputs (divided attention) have found differences in individual processing limitations [10, p. 141-186]. Findings justify that attention as a construct is divided in focused attention and

divided attention. The focused attention of an individual can further be studied auditorily or visually while divided attention is dependent on task similarity, task difficulty and practice. Most studies have been made in the laboratory under reasonably controlled conditions (control of confounding variables) (ibid.).

To better understand how these constructs are at work when the VHF – radio is used in “a real world scenario” we have included the following example which is based on true experience.

Example 1: The officer of the watch is getting ready to meet with the pilot boat at the pilot embarkation position¹⁶. He is manually turning the vessel to head a certain course when at the pilot station. According to instructions the container vessel must also make a certain speed through the water when at the pilot station. The officer is therefore turning the rudder to get the right turn rate while reducing the speed of the vessel by reducing the engine revolutions. At this very moment he hears his vessels name spoken of the VHF radio. He hears nothing else than the name of the vessel meaning that there is a potentially important message for him. He decides to finish his manoeuvre and when the container vessel several minutes later is at the pilot station steering the right course and doing the right speed he uses the replay function on the VHF to find out who called him. He finds out it was the pilot and call him back.

In this example the experienced officer takes measures to buffer the primary task of manoeuvring his vessel safely from any negative effect of interruptions. The primary task is then (his choice) to manoeuvre the vessel safely so that it steers the right course and proceeds at the right speed when at the pilot station. Monitoring the VHF radio and responding to a call can be seen as a secondary task. The officer makes the decision to follow his manoeuvre safely through (primary task) before checking the VHF radio. In this example the officer knows his limitations and executes the primary task before attending to the secondary. This protects his execution of the primary task. The interruption never unleashes its error potential. This is an example of successful task management.

But what if the VHF radio interrupts the officer resulting in an unsafe and error prone situation? Let us consider another example.

Example 2: A container feeder vessel is approaching the Great Belt Traffic VTS area from the south. Great Belt VTS is established to assist, monitor and to some degree control the traffic through the Great Belt. Shortly before entering the traffic separation scheme the vessel has to call the VTS at a certain position and give information about ships name, call sign, destination and cargo, etc. As the container feeder approaches the calling point the officer of the watch is getting ready to call the VTS. Knowing this is a potentially safety critical part of the navigation in this area he is well prepared. He picks up his prewritten calling info. He calls the VTS on channel 11 and transmits his message. He leaves the VHF radio and checks his position to see how far away he is from the critical turning point. The vessel is now at the turning point and the officer has to start the turn. At this moment the officer receives a call from Great Belt VTS saying that the officer did not report the air draft of the vessel in his previous message (the total height of the ship; important for VTS to know because they want to make sure that the vessels are not too high to pass under the Great Belt Bridge). The officer replies that he will look up the air draft. He leaves the VHF radio and

searches for the laminated paper with the ship particulars in order to find the air draft. While doing this the ship passes the critical turning point.

In this case we will argue that the officer has his attention focused on the primary task which can be seen as the safe navigation through the waters including calling the VTS. The officer must monitor the VHF radio while performing this primary task. His divided attention registers the call from Great Belt VTS. He chooses to respond to this call and the time he spends on this secondary task is the taken from the primary task. As Trafton et al. described this period of time it includes interruption lag and resumption lag [29].

We have now seen two examples dealing with the attention of the navigator. In the first example attention was managed and kept on the primary task. He ignored the disruption by the VHF. In the second example the disruption from the VHF interrupted the performance of the primary task and resulted in a potential dangerous situation: The ship passing by a critical turning point. The replay function could in both examples enhance safety: In the first example, where the call is ignored, the replay could help the navigator repeat the call and thereby determine if it had importance or was urgent. The risk of loosing a potential important call is thereby reduced. In the second example the replay function could have been used in a similar way: The navigator could (having the replay function in mind) have kept focus on the primary task – commencing and executing the turn at the turning point – and postponed the processing of the incoming message a few minutes.

Prospective Memory

Prospective memory (PM) is the ability to recall a previously formed intention at a specific time or in response to a specific cue in the future, without being encouraged to recall the intention [21]. PM is often playing a central role in safety critical domains. The consequence of failure of PM can be a so called human error. PM is about forgetting to do something in the future [12]. Many historical examples of severe consequences of failure of PM are known. In 1991 an aircraft controller positioned a commuter aircraft at the end of a runway and later forgot to move it to another location. The aircraft was still there when a cargo transport aircraft got clearance to land on the runway. The inevitable collision took several lives. It is estimated that in five out of the 27 major U.S. airline accidents between 1987 and 2001, in which the NTSB found crew error to be a causal factor, inadvertent omission of a normal procedural step played a pivotal role [8, p. 909]. And such inadvertent omissions are a form of prospective memory failure (ibid.) Dieckmann et al. [6] has described how PM failures is a threat to patient safety in hospitals – another safety critical domain. PM is vulnerable to disturbances and disruptions. If the navigator is disturbed – for example by a call on the VHF-radio – there is a risk that an intention (e.g. turning the ship at a turning point) is forgotten or that it is executed in a wrong way or at a wrong time. The replay function can be used to eliminate the disturbing effect incoming calls normally can have on prospective memory. The navigator can safely ignore incoming calls – without the risk of loosing an important or urgent message – in the time between the intention and the execution of the intention and thereby reduce the risk of prospective memory failure. Incoming messages can then be replayed after the execution of the intended action.

¹⁶ A pilot is an experienced navigator with comprehensive local knowledge acting as a consultant to the ship’s crew.

Interviews with Users

The analysis has so far illustrated that the disruptive and disturbing nature of incoming calls can lead to two safety critical problems: (1) Attention related problems where the attention is drawn from the execution of a primary task to the call on the VHF and (2) prospective memory related problems where the call interferes with an intention of doing something important. We have asked 12 potential users both captains, pilots, harbour masters, navigation school teachers and salesmen questions about their experiences using the RT5022 VHF. Some of them had experience using the RT5022 and the replay function while others had general experience using VHF but no experience using the RT5022. The last group answered according to fictive scenarios where they – based on their experience from sea – evaluated the replay function. The method was semi-structured interviews, and we presented the respondents for the transcriptions in order to validate it. When asking the users, they describe scenarios where a primary task e.g. manoeuvring the vessel has first priority and messages on the VHF is considered later when they are not disturbing the performance of the primary task:

“On a tug, you often need to keep attention fully on the work task [manoeuvring], and you can not take your hands of the handles to find paper and pencil to write down important information and positions. ...The alternative to the replay function would in situations like these have been that you later on should call Lyngby Radio [land based radio operator] to have them repeat the message or wait having the information until it is broadcasted again. ...When you know that you have the replay function you can continue your primary task. You do not need to get distracted by the message on the VHF. You can just notice that it has been there, and you do not need to use energy trying to understand or remember it in a stressful situation. You only need to remember, that you should listen to it using the replay function later.” (Interview with former captain on a tug boat, January 21, 2009)

The captain in the example above describes how he can be so engaged in a work task (focused attention) that he is not even able to find paper and pencil to write information down. The reason for this is also that the manoeuvring of the vessel requires intense manual control keeping both of his hands busy. Use of paper as cognitive artefact is under these circumstances not possible. The replay function here serves the purpose as a cognitive artefact. When the captain knows that it is possible later to get access to the information in the message he does not need to pay attention to its actual content – just keep in mind that there has been a message. The statement from the tug boat captain is further supported by a navigation school teacher:

“The function [replay] is especially useful at reception of emergency, urgency or safety calls because you do not need to abandon all other work – for example if you are in the middle of turning the ship – to write down call name and position. And if you do not write it down it is gone. You can not get this information again unless you call and ask: Who called about this and that?” (Interview with former radio operator, now teacher in GMDSS and radio courses at a Danish Navigation School, January 23, 2009)

He is here talking about persistence, that the information is gone if not recorded and that the disturbing nature of incoming calls is reduced because it can be replayed at a later time. Another respondent describes how a combination of manual control manoeuvring the vessel and communication with the deck crew together can form the primary task that needs full attention. He also describe that it is possible to hear and detect

own ship’s name mentioned even though he is not paying full attention to the message.

“The captain who is busy bringing his 70000 tonnes ship to the quay is not always able to answer calls on the VHF. He is listening what the crew on deck is reporting on the handheld radio. But he is nevertheless able to detect that the ship’s name has been mentioned – then he press replay to hear who made the call.” (Interview with salesman and navigator, dealer of the VHF, January 23, 2009)

Sometimes the primary task can be communication using VHF. A second call on another radio channel at the time when communication already is going on one channel can disturb the ongoing communication:

“A classic situation is that you are busy communicating on one of the VHF-radios and someone calls on the other channel which the other radio is set on. With the replay function I can continue my communication and then later on listen to find out whom made the call on the other radio.” (Interview with pilot, January 22, 2009)

“We have had it [the VHF-radio] on the harbour office for long. ... The ships call the harbour often, and at some times on different channels at the same time. With the replay function no information is lost.” (Interview with harbour master, January 22, 2009)

Incoming calls on the VHF-radio is one of the major sources of disruptions and disturbances on the ship’s bridge. The attention and prospective memory of the navigator are two important cognitive functions related to safety. When attention and/or prospective memory are compromised due to disruptions from incoming calls it can eventually and possibly lead to serious safety problems. Evidence from the interviews with users shows how the replay function can be used to counteract for these problems. Knowing that the message can be repeated later using the replay function the navigator can keep full attention to the primary task e.g. manoeuvring the vessel. The replay function reduces the need for attention to incoming messages when this will disturb the execution of a primary task. Thereby the risk that the call interferes with prospective memory leading to PM failure is also reduced. We have not found examples in the interviews showing how the replay function directly is used to protect prospective memory, but we have found examples showing that it is used to protect focused attention on a primary task.

LESSONS LEARNED AND CONCLUSIONS

This paper describes a design process where knowledge is acquired at different stages and at different levels. Knowledge about the maritime work domain, maritime safety, psychological theories and methods, use of VHF-radios etc. was brought into the process through the competences among the members of the design team. Further knowledge about the VHF and context of use was acquired during the data collection process through the on board observations and interviews and in the simulator scenarios used in the design seminar for the design team. This knowledge resulted in the replay function as proposed by the human factors specialist in the design team. At this stage it was still unknown how the user’s response to the replay function would be. It was therefore tested in a usability test where also other design solutions were tested. The result was new knowledge about the user’s evaluation of the function. This knowledge was used in the adjustment of the design, and eventually the final design was produced and released on the market. In

overall the design process could be viewed as a learning process where knowledge is collected at all stages and gradually build up in the design team. The design team has learned about the use of maritime equipment, about problems and constraints in the maritime work domain and about the cognitive aspects of the use of VHF-radios. The team members have learned about technology, users and use contexts. This knowledge has been used within the design process and is further more anchored in the design team for reference in future projects.

Now – when the device has been on the market for more than 3 years – we can analyse the replay function in the device once again using new psychological concepts and theories in combination with collection of further domain knowledge and knowledge about the use of the device through interviews with users, and thereby acquire new knowledge about the user interface, the use context, the impact on safety etc. We could now take it one step further: We could make field observations on how the replay function is used in real life scenarios or even look for situations and examples where the use of the replay function played a key role preventing accidents. We could thereby acquire further knowledge which for example could contribute to further development of the function; in later versions or new products. The knowledge could also be valuable in the development of instructions for use, guidelines and procedures and work practices. The design process seems to be a never ending process. Going back to the marketing material describing the function as quoted earlier in the paper we could add that the replay function can be used both (1) in situations where a call was missed by accident (unable to perceive) and (2) in situations where another task requires full attention and the processing of incoming calls therefore with benefit can be postponed. The result of the analysis shows that the new replay function supports usability of the radio and maritime safety in general to a far wider extent than anticipated in the design process and reflected in the way the radio is presented in marketing material. The apparently successful invention of the replay function came from the combination of psychological knowledge, domain knowledge and field observations. It was not delivered by users in a user-centred design process. It was derived from knowledge about users and the context of use.

Through our analysis we have shown that the function counteracts for and solve several types of problems and thereby enhance safety which was the objective from the start of the project. Some of these problems was identified during the design process (perception and persistence) while others were identified in the post-release analysis presented here in this paper. Looking at the device in use and applying new psychological concepts not included in the original design process gives brings us new knowledge. And this demonstrates the benefits from doing post-release analysis – what in ISO 13407 is called long-term monitoring [14].

We have in this paper not considered potential negative effects of the new replay function. Our aim has been to demonstrate how the function from a theoretical perspective and from the perspective of the users could be considered an improvement of maritime safety. But there could also be potential negative effects on short or long term, and these should be evaluated as well: Perhaps the crew will be so engaged in their primary task that they will even forget that an important message needs to be replayed. Or maybe there will be so many messages in a short period of time, while they are engaged in a task, that when they will replay them, some of them had already been erased. Perhaps the attention paid to incoming messages will

decrease to a critical low level. Some of these problems can be solved through technology. Putting more memory in the device will increase the amount of available calls stored in memory and it will become less likely that an important message has been deleted by new messages. Potential negative effects of the replay function can be identified through field studies, observations and interviews, and they should be considered in the overall evaluation of the replay function. It should also be evaluated to what extent the negative effects are minor compared to the positive and safety enhancing effects. This could be subject to future studies and work.

ACKNOWLEDGMENTS

Our thanks to the crews onboard the ships we visited as part of the design project and to the respondents in the post release analysis.

REFERENCES

- [1] Altmann, E. M. and Traflet, J. G. 2004. Task Interruption: Resumption lag and the role of cues. In: 26th Annual Conference of the Cognitive Science Society.
- [2] Baars, B. J. 1997. Consciousness versus attention, perception and working memory. *Consciousness and Cognition*, 6, 363–371.
- [3] Baddeley, A. 2007. *Working Memory, Thought, and Action*. Oxford University Press, Oxford.
- [4] Brixey, J. J., Robinson, D. J., Johnson, C. W., Johnson, T. R., Turley, J. P and Zang, J. 2007. A concept analysis of the phenomenon interruption. *Advanced in Nursing Science*, 30(1), 26–42.
- [5] Damos, D. L. 1997. Using interruptions to identify task prioritization in Part 121 air carrier operations. In: 9th International Symposium on Aviation Psychology. Ohio State University, Columbus.
- [6] Dieckmann, P., Reddersen, S., Wehner, T. and Rall, M. 2006. Prospective memory failures as an unexplored threat to patient safety: results from a pilot study using patient simulators to investigate the missed execution of intentions. *Ergonomics*, 49(5–6), 526–543.
- [7] Dismukes, K. 2001. The challenge of managing interruptions, distractions, and deferred tasks. In: 11th International Symposium on Aviation Psychology. Ohio State University, Columbus.
- [8] Dismukes, K. 2006. Concurrent task management and prospective memory. Pilot error as a model for the vulnerability of experts. In: *Human Factors and Ergonomics Society 50th Annual Meeting*, 2006. Pp. 909–913.
- [9] Dornheim, M. A. 2000. Crew Distractions emerge as new safety focus. *Aviation Week and Space Technology*, 17, 58–65.
- [10] Eysenck, M. W. and Keane, M. T. 2005. *Cognitive Psychology – a student’s handbook*. 5th edition. Psychology Press. New York.
- [11] Gillie, T. and Broadbent, D. 1989. What makes interruptions disruptive? A study of length, similarity and complexity. *Psychological Research*, 50, 243–250.
- [12] Harris, J. E. and Wilkins, A. J. 1982. Remembering to do things: A theoretical framework and illustrative experiment. *Human Learning*, 1, 123–136.

- [13] Hutchins, E. 1995. *Cognition in the Wild*. MIT Press, Cambridge, MA.
- [14] ISO 13407. 1999. *Human-centred design processes for interactive systems (ISO 13407:1999)*. European Committee for Standardization, Brussels.
- [15] James, W. 1890 *Principles of Psychology*. Holt, New York.
- [16] Kjeldskov, J. and Stage, J. 2003. The Process of Developing a Mobile Device for Communication in a Safety-Critical Domain. In: 9th IFIP TC13 International Conference on Human Computer Interaction, Interact 2003.
- [17] Kjeldskov, J. and Stage, J. 2006. Exploring “Canned Communication” for Coordinating Distributed Mobile Work Activities. *Interacting with Computers*, 18(6), 1310–1335.
- [18] Latorella, K. A. 1996. Investigating interruptions: an example from the flightdeck. In: 40th Annual Meeting of the Human Factors and Ergonomics Society. Pp. 249–253.
- [19] Lützhöft, M. H. 2004. *The Technology Is Great When It Works. Maritime Technology and Human Integration on the Ship’s Bridge*. Dissertation no. 907. University of Linköping. Sweden.
- [20] Lützhöft, M. H. and Dekker, S. W. A. 2002. On Your Watch: Automation on the Bridge. *J. of Navigation*, 55(1), 83–96.
- [21] McDaniel, M. A. and Einstein, G.O. 2000. Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, 14, 127–144.
- [22] McFarlane, D. C. and Latorella, K. A. 2002. The score and importance of human interruption in human-computer interface design. *Human Computer Interaction*, 17(1), 1–61.
- [23] Miyata, Y. and Norman, D.A. 1986. Psychological issues in support of multiple activities. In: Norman, D.A. and Draper, S.W. (eds.) *User-Centered System Design: New Perspectives on Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- [24] Monk, C. A. 2004. The effects of frequent versus infrequent interruptions on primary task resumption. In: *Human Factors and Ergonomics Society 48th Annual Meeting*.
- [25] Norman, D. A. 1991 *Cognitive artifacts*. In: Carroll, J. M. *Designing Interaction: Psychology at the Human-Computer Interface*. Cambridge University Press. Pp. 17–38.
- [26] Porathe, T. 2006. *3-D Nautical Charts and Safe Navigation*, Dissertation no. 27, Mälardalen University Press, Västerås. Available at <http://www.diva-portal.org> Search for author “Porathe”.
- [27] Schager, B. 2008. When Technology Leads Us Astray: A Broadened View of Human Error. *J. of Navigation*, 61, 63–70.
- [28] Thrane & Thrane. SAILOR RT5022 VHF DSC brochure downloaded from <http://www.thrane.com> on January 21, 2009.
- [29] Trafton, J. G., Altmann, E. M and Brock, D. P. 2005. Huh, what was I doing? How people use environmental cues after an interruption. In: *Human Factors and Ergonomics Society 49th Annual Meeting*.
- [30] Trafton, J. G., Altmann, E. M., Brock, D. P and Mintz, F. E. 2003. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International J. of Human Computer Science*, 58(5), 583–603.
- [31] Van Westrenen, F. 1999. *The maritime pilot at work*, PhD Thesis TU-Delft, Netherlands.
- [32] Wickens, C. 2007. *Aviation*. In: Durso, F. T (Ed.). *Handbook of Applied Cognition*. 2nd edition, John Wiley and Sons, Ltd. Chichester.

Activity-driven Design Approach to the Novel Control Room Environments

Hanna Koskinen
VTT Technical Research
Centre of Finland
Vuorimiehentie 3, Espoo
P.O. Box 1000, 02044 VTT,
Espoo
hanna.koskinen@vtt.fi

Jari Laarni
VTT Technical Research
Centre of Finland
Vuorimiehentie 3, Espoo
P.O. Box 1000, 02044 VTT,
Espoo
jari.laarni@vtt.fi

Leena Norros
VTT Technical Research
Centre of Finland
Vuorimiehentie 3, Espoo
P.O. Box 1000, 02044 VTT,
Espoo
leena.norros@vtt.fi

ABSTRACT

This paper reports an ongoing work carried out in a project Tilava (Spacious). The aim of the project is four-fold: 1) to study how the control room as a space/ environment enables operators to understand and control the system under their supervision; 2) to examine the possibilities to improve the control space's ability to communicate the process behind by better acknowledging the control room's physical, social and virtual space qualities and how these qualities could be supported through the introduction of new technology; 3) to create a future control room concept and 4) develop methods for early expert user involvement in design and elicitation of user requirements.

Keywords

control room environments, control and monitoring activity, expert user involvement, user requirement elicitation, situation awareness

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *user-centered design, interaction styles, theory and methods.*

INTRODUCTION

The control room is a space / a place that is typically designed for monitoring and controlling of a particular process [5]. Thus, the control room can be seen as a comprehensive human system interface, which provides process information and from which the process can be controlled. The control room design has predominantly been focused on technically fulfilling control room's process monitoring and controlling functions (e.g. detecting and managing disturbances). Over the years new functionalities has however evolved and additional dimensions of control room functions has been introduced (e.g. information sharing within the whole organisation and data management). Thus, a larger variety tasks are conducted through and allocated for the control room in addition to the previous ones. [9, 10]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Diversification of control room functions follows the development of technology. Traditionally, control centers and industry control rooms are built by bringing the needed interface and communication equipments (e.g. control panels and consoles, radiophones) to a room space that is specifically dedicated for this purpose. In conventional control rooms based on analogical technology the monitoring and controlling activity was done by observing and manipulating the control panels and consoles. In this kind of setup the whole control room environment and its elements (e.g. structures, walls and objects) was taken advantage of and the entire room space constituted a comprehensive interface for monitoring and controlling the complex system.

There is an increasing use of new information and communication technologies (ICT) in control room environments. Technology development and the digitalization of control rooms have not only made it possible to collect but also more widely represent information from the process that is controlled. Consequently, the controlled system has extended due to the increased amount of information and the possible interplay between the different elements of the system. Nowadays it is possible to produce a more diversified and richer picture of the controlled system for operators. Overall, the effects of digitalization have been considered positive but at the same time these changes are challenging, and they have deep impacts on the control room and operators' work practices. For example they challenge operators' abilities to process the amount of data that is made available and the ways how they are used to navigate within the information system. It has also been argued that because of the digitalization of the old analogical systems operators more likely face a so called keyhole effect. The control room space has in a way vanished inside the display based control system and only a small part of it can be observed at a time. Instead of feeling of being inside the controlled system it is observed outside through a small hole. The interaction with the control room environment has come more passive and maintaining situation awareness has become a focal problem. One reason for these problems is that the control room space, the interface elements and the work activity by itself are designed separately. This makes the integration of the control room's functions and physical facilities more challenging. Also changing the functional roles and structures that were once designed into the control room environment might be difficult.

These open questions have lead designers and developers to seek ways to improve the information presentation and operators possibilities to better "grasp" and feel being control of the system. Developments of ICT based interface technologies has brought about the possibility to integrate and update the spatial qualities of earlier control room designs in new designs and in this way facilitate more intuitive representation and control of the process. Multimodal (e.g. integrated use of visual-, audio-,

speech- and tactile-based information) interfaces can be developed in order to support operators' interaction through various communicational channels. This makes it possible to mediate more qualitative information and so called "weak signals" without overburdening vision. We think that these kinds of environmental cues that human partly subconsciously perceive by acting in an environment might play an important role when creating awareness of the prevailing situation. These space related interaction qualities should not be undervalued in new control room designs.

This paper describes our attempts at applying activity driven design approach in the design of control room spaces in order to be able to better address all aspects of interaction in the design of control room environments. The paper is organized as follows. First, meaning of the situational understanding of an environment is discussed. Second, three space qualities present in control room environments are introduced and discussed. Third, three control room metaphors are created by combining these qualities in order to conduct case studies about the form and meaning of interaction that takes place in a control room. We end up with a discussion and implications of the study for the design of control room applications.

TECHNOLOGICAL SUPPORT FOR SITUATION AWARENESS

From the process control point of view it is important that the operators of the process have a feeling of being control of events and the situation, and, that they have a clear picture of the situation constraints and their own possibilities to take control actions. A term situation awareness (SA) is used for describing this generic operational purpose [2]. Situation awareness is defined as the degree to which human operators are aware about what is happening and how their own situation is related to the surrounding context. Many human actors take part of the process control and monitoring activity and therefore the establishment of a common ground of understanding is essential to support through control room designs.

Another concept that is used for describing the action possibilities of a control room is the concept of affordance that rises from the field of ecological psychology. The term was introduced by James Gibson and according to him the term affordance refers to the complementary and interaction between an organism and an ambient environment. It is used to describe those things that an environment offers or furnish to an animal through interaction [4]. For Gibson, a key to understanding behavior was an analysis of the ecology. In the context of complex process control the control room environment is perceived directly by means of what kind of possibilities it can offer to operators' actions.

With the technological solutions, for example the use of novel interface techniques can be influenced on what action possibilities are perceived in a control room space, in which case the ability of the control room to support and promote operators activity and understanding of the process is closely connected to the variety of the views that the room space offers for users to observe the process and availability of diversified space modalities.

CONTROL ROOM AS A SPACE

Three space qualities are introduced in order to better understand the design of control room environment. These three qualities are physical, virtual and social space quality. They can be used as a basis for analysing affordances embedded in spaces, in this

connection control room space. Next it is discussed how the three introduced space qualities can be interpreted from control room environment's point of view. The *physical space quality* of a control room is often thought to be related to the physical structure of the control room (i.e. walls, roof, objects), but it also includes aspects that supports operators' physical activities and control operations. The constant interplay between the physical control room environment and operators' physical actions affords some actions while hinders the others. Our assumption is that users' physical action possibilities is an essential element of maintaining situation awareness. In control room design it is important that solutions promote physical interaction as well as provides users rich alternatives for such interaction within the control room environment. It is supposed that this kind of active involvement with an environment supports memory (e.g. embodied memory) and teams' cooperation and communication (e.g. mediating the meaning of movement and operations) [1].

Virtual space quality deals with an idea that the control room is a representation of the complex object of an activity. Therefore the control room should provide the operators with a realistic and an understandable picture from the prevailing process state. Digitalization of the control rooms has a certain extent thought to bring operators apart from controlled process. By utilising technological solutions that supports and make use of the virtual space quality users can be provided with a new kind of views to the controlled entity. The object of control activity (e.g. power plant) could even be brought centre of a control room space for team's common review [3]. With the help of new technological solutions it is also possible to perceive and explore objects that earlier were out of reach of operators (e.g. reactor core in nuclear power plants).

The *social space quality* of the control room is created through the crew's cooperation and communication. The different actors in the control room as well as in the entire plant complement each other in order to achieve the goals of the activity. The control room environment can be considered as a framework in which the activity and cooperation takes place and through which the operational information is mediated. In the control room space the social space quality plays an important role in the process of creating shared understanding of the process state and in the coordination of the crew's activity and control operations. When the operational monitoring and control of the process is performed by using digital display based systems crew's old cooperation and collaboration strategies are challenged. The variety of work tasks that the crew is responsible for are done through operators' individual work stations. Therefore it is not so easy to be aware about the other crew members' activities within the control room space which may influence the crew's ability to maintain shared understanding of the control room activities and the effect of all separate operations as a whole. Also the role of the personnel that works in a main control room and the ones that are on the field is changing. Since operators are not necessarily working alongside at one room space, different kinds of computer-assisted teamwork and communication tools are needed in the control of complex systems. Overall, it has been paid too little attention to the communicative function of tools and physical spaces in the control room design [8].

ACTIVITY-DRIVEN DESIGN OF CONTROL ROOM ENVIRONMENTS

Tilava project has a four-fold aim: Firstly, to study how the control room as a space / an environment enables operators to understand and control the system under their supervision. In

other words how human operators create a picture and attain situation awareness of a certain process. Secondly, to examine the possibilities to improve the control space's ability to communicate the process behind by better acknowledging the control room's physical, social and virtual space qualities and how these qualities could be supported through the introduction of new technologies. These space qualities and possible supportive technological solutions are explored and tried out in a small scale design workshops together with users. Thirdly, the project aims to create a future control room concept, and fourthly develop methods for early expert user involvement in design and user requirement elicitation.

Research Questions

The questions that emerged at the beginning of the project were following: 1. How users distinguish and experience the challenge of maintaining situation awareness in different kind of industrial control centre settings? 2. How new interface elements and solutions that are brought to the control room space influences to the joint functioning of the human – environment system as a whole? And how they contribute to attaining the situation awareness? 3. In what degree and way the expert users should be involved in the design of new control room applications? And, how their operative knowledge could be best employed to the design process?

Research Approach and Methods

Traditionally, the control room and interface elements in it are considered from the point of view of a variety technical equipments that are needed for the process control and monitoring. The mainstream interface development has been technology driven. However, in this project the control room space is thought to be an integrated wholeness in which all components and human actors are interrelated to each other. The elements of the control room environment can merge differently depending on what activity it is strived for. Similarly, the tasks carried out and the ultimate goals of the work as well as the quality and performance demands that are imposed on the activity affect how well the control room space functions and support operators' situation awareness. [11] The control room space constitutes a kind of information ecology in which the human – machine interaction takes its place and form. [6] In Tilava project we are trying to find suitable combinations of control room elements for human – technology joint functioning.

Exploring the Space Qualities of the Control Room

Our assumption is that making broadly use of the space qualities (physical, virtual and social) in control room designs it is possible to develop control room settings that facilitates users to maintain situation awareness. In Tilava project we combined these three space qualities in diverse sets of pairs and build around them three case studies to be explored together with the expert users in design workshops. The idea in these workshops is to discuss with the users the meaning and essence of the three space qualities in control room environment and test new interface technologies in order to assess their potentiality to support such qualities. Our aim is to not only to study the applicability of some specific technological solutions in control rooms as such but also more generally the meaning and the role of different space modalities for situation awareness.

By combining these three space qualities in diverse sets of pairs we created three control room metaphors: Interactive-, Intuitive- and Boundless control room.

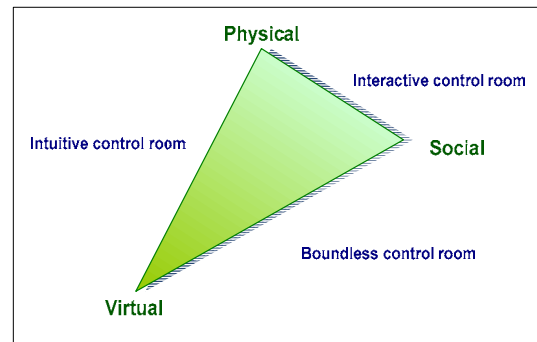


Figure 1. Space qualities and control room metaphors combined in diverse sets of space quality pairs.

Interactive control room metaphor is a combination of the physical and social space qualities. In Interactive control room case the focus is on possible use of different kind of interactive surfaces in control room environments as a crew's cooperative tool. For example, in a digitalized control room the control operations are done by using mouse control. Thus, the place of execution and the physical form of control operations (physical space quality) has changed when compared to the conventional analog control room setting where the control operations were done through the control panels and desks. Operators have reported difficulties in monitoring other operators' activities when there is no specified execution location or gestures involved (social space quality). This makes it also harder to reflect others activities in relation to own activity and realize what their control operations (gestures) are referring to. The control room space should be able to mediate, by choice, automatically information about the development of the situation and operators' activities within the control room. Our assumption is that through innovative use of interactive surface technologies (i.e. multi-touch displays, digital tangible objects and large screen displays) in control room environments, the control room's physical and social space quality can be promoted.

In *Intuitive control room* the physical and virtual space qualities are combined in order to explore how complex and sometimes even invisible object of control activity could be made understandable and perceivable for human operators. The challenge that arises is how information is presented to the operators. Control room space can be seen as a representation of a complex process behind. It should mediate an overview of the status of the process in understandable and intuitive way. Operators should be able to obtain a picture from the object of their supervision and follow the progress of the process even in some processes the area of supervision is extensive or the observed magnitudes might not be visible. In order to be able to better illustrate the relationship between pieces of data and the supervised object new kind of digital pictures representing the state and the nature of the process situation has to be developed.

In the last case study, *Boundless control room*, the combination of the virtual and social space qualities are explored. Controlled entities are becoming increasingly complex in size, dynamic and functionality with the advent of new technological advances. Nowadays, it is possible to collect and more widely represent information from the controlled process. However, in order to enable this amount of specified process information more local and detailed data from the process

needs to be collected. This challenges the cooperation and communication between the personnel in the main control room and the ones working in the field. It can be said that the control room in a way spreads over the whole supervised area and employs actively resources in the field. In this kind of setting it is essential that the control room functions as a mediating surface between the operative parties and helps them to maintain a shared understanding of the current status of the plant and its relation to potential future states. As a result, the control room space comes flexible and it can adapt its functions and structures according to the prevailing process situation and the needs of the human operators.

A FUTURE CONTROL ROOM CONCEPT

It is expected that based on the design workshops a concept level description of the future control room and the elements interplaying in it will be formed. This future concept of operation creates a ground for activity driven design of ICT-based tools and control room environments. It is important to understand the functioning of the control room environment as a whole in order to get a better idea of the technological solutions and the functional elements that the future control room designs should be built upon.

Our assumption is that by involving expert users and employing their context knowledge it is possible to define more detailed what the future control room should look like. In a research that aimed to study the use of large screen displays in a control room environment one of the outcomes of the study was an operation oriented large screen display framework to help developers to define the usage requirements for the coming large screen application. [7] The framework pursues to describe what the purpose of use of large screens is in control rooms in different usage situations, which user group large screens are for, how they are used and in which form the information should be presented on it. Similarly, the outcome of this project, the concept level description of future control room environments addresses the essential underlying structures and functionalities that need to be paid attention in the new designs.

DISCUSSION AND CONTINUATION OF THE PROJECT

An activity-driven, user-centred approach to design deepens the understanding of the operators – control room functioning as well as the tensions that create incitement for change within this system. We believe that a thorough discussion and description of the elements and form of activities within present control room environments is an essential first step to understand in what form the activity could take place in the future. It is also important to involve expert users to the design process as early as possible. Their operational knowledge and experience is an important resource for the design of tomorrow's control room environments. After the organized design workshops the continuation of the project is to analyse and try to generalize the meaning and essence of the spatial qualities of the control room spaces. Further, we believe that

the insight gained from the involvement of operators can also open the eyes of the designers to see the hidden interaction elements of control rooms.

ACKNOWLEDGMENTS

We would like to acknowledge many insights from the participants of the Tilava project and operators participating design workshops.

REFERENCES

- [1] Dourish, P. 2001. Where the action is. The foundations of embodied interaction. Cambridge, Mass. MIT Press.
- [2] Endsley, M. R. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32–64.
- [3] Flach, J., Hancock, P., Caird, J. and Vicente, K. J. 1995. Global perspectives on the ecology of human-machine systems. Hillsdale, New Jersey, Lawrence Erlbaum Associates.
- [4] Gibson, J. 1977. The Theory of Affordances. Perceiving, Acting and Knowing. In: *Toward and Ecological Psychology the Theory of Affordances*. Shaw, R. and Bransford, J. New York, Lawrence Erlbaum Associates. Pp. 67–82.
- [5] Hollnagel, E. and Woods D. 2005. Joint cognitive systems. *Foundations of Cognitive systems engineering*. Boca Raton, Taylor & Francis.
- [6] Kaasinen, E. and L. Norros (Eds.). 2007. Älykkäiden ympäristöjen suunnittelu. Kohti ekologista systeemijättelua. Helsinki, Teknologianinfo Teknova Oy.
- [7] Laarni, J., Norros L., Salo L. and Koskinen, H. 2008. Designing large screen displays for process monitoring and control. Enlarged Halden Programme Group meeting, Loen, Norway.
- [8] Norros, L. and Salo I. online. Design of joint systems a theoretical challenge for cognitive systems engineering. *Cognition Technology and Work*.
- [9] O'Hara, J., Pirus, D. and Beltracchi, L. 2003. Information display: Considerations for designing modern computer based display systems. Palo Alto, CA, Electric Power Research Institute.
- [10] Paunonen, H. 1997. Roles of informing process control systems. Tampere, Tampere University of Technology: 166.
- [11] Savioja, P. and Norros L. 2008. Systems usability promoting coretask oriented work practices. In: *Maturing Usability: Quality in Software, Interaction and Value Systems usability promoting coretask oriented work practices*. Law, E., Thora Hvannberg, E., Cockton, G. and Vanderdonck, J. London, Springer. Pp. 123–143.

Session 9: Activity Theory in Design

Cultivating and Enriching a Tool in Use – Case ‘Development Radar’

Hanna Toiviainen

Hannele Kerosuo

University of Helsinki
Center for Research on Activity, Development and Learning
Teollisuuskatu 23 B, P.O. Box 26
00014 University of Helsinki, Finland
hanna.toiviainen@helsinki.fi; hannele.kerosuo@helsinki.fi

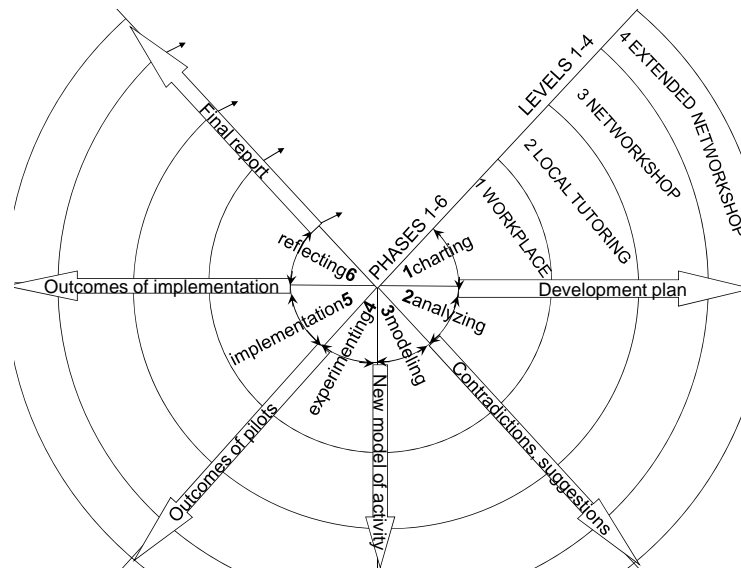


Figure 1. Development Radar.

ABSTRACT

The implementation of a tool is a crucial phase of the design process and sets the challenge to the users to learn. It is the process of transforming the artifacts to instruments of activity [2] through an expansive integration [7] that is supposed to expand the object of work [5]. We want to explore this regarding the tool ‘Development Radar’ that was designed to support learning in the network project on workplace development. This study is a continuation to the analysis, in which we looked into the co-configurative ways of designing the learning tool. After the successful implementation [14], how did the in-house developers perceive and modify the tool; did they manage to integrate it to their work and development practices by cultivating and enriching the tool in use [9]?

In the activity-theoretical terms, by cultivation and enrichment a new tool creates new mediation of the work activity. Mediation, however, has multiple dimensions and meanings.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Introduction of the tool for workplace development may mediate between the learning network and the single workplace of a participant, but the tool may also mediate the emerging identity of an in-house developer in one’s work community. There is variance among the participants of the learning network regarding the readiness to cultivate and enrich the tool in use; will this also tell something about the potential of learning and co-configuration on the workplace and network levels? The analysis is based on the longitudinal study of the regional learning network and the discursive data gathered in its workshops, which offers possibility to focus both on the collective and the individual actions of cultivating and enriching the radar tool.

Keywords

activity theory, developmental work research, tools, development radar, learning networks, multiple mediations

INTRODUCTION

Development Radar as such – hadn’t it that possibility to return without losing face, the whole experimentation sector would have to be taken away. Because experimentation is never complete. (In-house developer, 2008)

When carrying out learning-interventionist research in work life networks we often need to reconsider methods and practices approved in the development of single organizations.

This is, briefly, the background for designing the tool we call Development Radar. We researchers designed it to support our instruction activity mainly by redesigning concepts used in our learning network. Simultaneously, we offered the radar tool to the participants as the tool for learning work development methods in a network setting and managing their workplace projects [14]. This kind of interaction may lead to mutual learning that Pascal Béguin [1] terms ‘activity exchange,’ a dialogical process between users and designers mediated by temporary design outcomes, “whose use in action *reshapes, enriches, or shifts* the characteristics of the object currently being designed” ([1], p. 715, italics added). The opening citation by an in-house developer clearly demonstrates this by reshaping the characteristics of the tool or even the development activity itself, *a possibility to return without losing face*, with a message not necessarily crystallized by instructors: *experimentation is never complete*.

All this hints that, from the point of view of learning, there are many dimensions of design to be examined. Prost et al. [11] have elaborated on the approach mentioned by using dimensions of crystallization, plasticity and development as suggested by Béguin (2007, according to Prost et al., [11]). We want mainly to contribute to the last one, the development. As the authors point out, development perspective to design implies that the use of tool always results in re-inventiveness on given artifact. Moreover, the design strategy should simultaneously foster both the re-design of the tool and the examination and change of the users’ and designers’ activities. This is closely related to the notion on co-configuration [15; 3], by means of which we analyzed the invention phase of Development Radar [14].

Our exploratory concepts to analyze the actions of design-as-development are *cultivating and enriching* contributing to the process of co-configuration of the radar tool. Cultivating and enriching may lead to a redesigned and refined tool, but equally interesting for us is the way, in which the cultivating and enriching episodes cross-light and potentially expand the object of activity, the object being the in-house development work taking place on multiple levels (see levels in Figure 1). The relationship of a tool with the object of activity is that of mediation [16]; the introduction of a new tool therefore creates new mediation or re-mediation of activity [8]. Users’ uses of a tool may be expected or unexpected, seen from the perspective of designers, but in the developmental activities it is only through the implementation and use of a tool that the “expected” use may be revealed. The tool is cultivated to explore its use potential at its best leading to enrichment of the tool as well as enriched insight into development activity.

In this paper, we first discuss the methodological issues, in chapter 2 the educational approach Developmental Work Research (DWR) [6] applied in the case of the learning network, and in chapter 3 the data and the methods of analysis. Chapter 4 analyzes the discursive actions of cultivating and enriching in a chronological order. Finally the notions on analysis are discussed in chapter 5 and the results summarized in chapter 6.

EDUCATIONAL APPROACH

Developmental Work Research (DWR)

We agree with the discussion of this workshop of ECCE 2009 conference, that the capabilities required in the working life more and more involve mastering the future-oriented representations of change and work development. These capabilities are not only assigned to specialized designers, but also will be expected from so-called ordinary workers. This was one of the starting points, when the Forum of In-house

Development was designed as a part of the learning network of the South Savo region [13].

By applying the *Developmental Work Research* (DWR) [6], the aim was to invite workers from workplaces of the region to study the concepts of work development, learn together and from each other, and become “in-house developers” qualified to carry out projects in their work communities. One of the core concepts of DWR is the cycle of expansive learning that is redesigned in the center of Development Radar (see phases 1–6 in Figure 1). The cycle and other DWR-based concepts discussed in the data are briefly presented in order to make the analysis understandable.

The processes of DWR are carried out to trigger *cycles of expansive learning* based on the historical understanding of the change in collective work activity, its past, present, and projected future. The ultimate motive for change and, therefore, learning is in the collective need and effort to expand the object of present activity, to find the solution to the developmental contradictions experienced and analyzed in it. The steps or phases are: *charting and questioning* the present situation, *analyzing* the present activity and the developmental contradictions, *modeling* and analyzing the new activity as the solution to the present contradictions, *experimenting and piloting* with the new, *consolidating, generalizing, and reflecting* on the new activity. In the Forum’s work, we wanted to apply the dynamics of the cycle of expansive learning on two dimensions, a) by organizing the series of 1,5 years workshops along the steps of the cycle, and b) by guiding the in-house developers to carry out their workplace projects according to the steps of the cycle. The trajectory of the Forum’s work in this paper covers the phases from the analysis to the reflection.

The basic tool for analyzing activity is the conceptual *model of the activity system*, usually called “*the triangle*,” as appears in the discussion excerpts in Section: “Analysis”. It is composed of systemically interacting elements the object, subject, tools, rules, community, and the division of labor. The triangle model was most important for the in-house developers of the Forum. It was used to help the participants analyze and realize the focal topics of development in their workplaces.

In addition to the cycle and triangle models, a central concept discussed in the data of this article is *the mirror material*, derived from the DWR-based Change Laboratory approach [4]. Gathering data from the work activity, first to manifest the need for change and later to reflect on the experimentation of the new, was one of the recurring themes of the Forum and the participants’ “homework.” It is most demanding, often laborious requiring iterative rounds of data collecting to serve the intended developmental purposes. One source of mirror material is *a problem log* that the workers are asked to keep during a defined period of time to write down the disturbances experienced at work.

Glossary of Learning Network

Here are the explanations of the central concepts of the empirical research case. The term in parentheses is also used in the text. Capital initial letters emphasize the special use of the term.

Learning Network of South Savo (Learning Network) is a regional network and one of the projects financed by the Learning Network program of the Finnish Workplace Development Programme TYKES. Learning network of South Savo is a composition of five forums of work life development coordinated by a regional Anttolanhovi rehabilitation and research center.

Forum for In-house Development (Forum) is one of the forums. The first Forum from February 2007 to May 2008 (discussed in this paper) had participants from six workplaces, five work-life consultants as local tutors, two researchers as instructors, and the project coordinator. The main work took place in **Networkshops** of Forum. **Extended networkshop** invited the managers of workplaces to discuss the projects. The last session was called **Learning workshop**.

In-house Developer (Developer): Participating workplaces or organizations sent 1–4 representatives to Forum to learn methods of workplace development and become in-house developers capable to conduct developmental projects in their work communities and enhance boundary-crossing and networking.

Workplace project: The central part of Forum’s work was based on the development projects the in-house developers started in their work units.

Local tutoring is a mediating level between the workplace project and network. The workplace projects were grouped in pairs coached by three **local tutors**. One tutor took care of a group of in-house developers from two workplaces. The meetings were held in the intermediate phases of Networkshop meetings.

Development Radar (Radar, the radar tool) is the tool developed by the researchers to manage the multiple phases and levels of the Forum. The core is formed by the **cycle of expansive learning and development**. The data for analyzing its use is drawn from the meetings of Forum, based on the discourse and visual representations in the collaborative working called **Radar round** and **Learning Paths Gallery**.

METHODS AND DATA

Data is from the workshops of Forum. Table 1 shows all workshops of the analyzed period – including topics, tools and methods not explained in this paper. (The evaluation of the methods is beyond the focus of this article [12]). After the introduction of the Development Radar tool in August 2007 [14] we organized the “Radar rounds” whenever it served the tasks of the given session, in order to learn about its usability for the participants. This happened in two sessions, in October and February. In addition, the radar model was used in some of the participants’ presentations in September 2007, and April and May 2008, and briefly discussed in a couple of other situations during the analyzed period.

Table 1. Forum for In-house Development, meetings 5–11.

	<i>Date/Workshop</i>	<i>Main topic</i>	<i>Main tools, methods</i>
5	September 2007/ Networkshop 2	Modeling the objects of development	Models of DWR
6	October 2007/ Networkshop 3	Planning the mirror material collection	Development Radar
7	November 2007/ Networkshop 4	Planning the first workplace workshop	Sparring-group work
8	February 2008/ Networkshop 5	Modeling the present phase of project	Development Radar Implement. platform
9	March 2008/ Networkshop 6	Mirror material workshop	Video projector
10	April 2008/ Extended Networkshop	Presentations to workplace managers	Project portfolios
11	May 2008/ Learning workshop	“Me as in-house developer”	Project portfolios Learning groups Learning-paths gallery

The unit of observation is a discussion episode, in which Development Radar is the topic taken up, discussed, and closed.

Table 2. Data analyzed.

<i>Date/Workshop</i>	<i>Task and working method</i>	<i>Data/Radar episodes/excerpts</i>
October 2007/ Networkshop 3	“Radar round” (Project-based small-group discussion 10 min and shared round 5 min/project) (1) Phase of development plan? (2) Phase of mirror material? (3) Schedule of workplace workshops; what has been planned with management?	Radar round answers by all six participating workplace projects 8/17 discussion episodes
February 2008/ Networkshop 5	“Radar round” (Project-based small-group discussion 30 min and shared round 25 min/project) (1) Where are we going on the cycle of expansive learning? (2) What intermediate goals have been reached? (3) What kinds of developmental tensions have occurred?	Radar round answers by five out of six workplace projects 6/17 discussion episodes
May 2008/ Learning workshop	“Learning-paths gallery” (Small-group discussions, preparation of a Learning Path poster, gallery presentations) Analyze your own development in relation with the material and phases of the project - What has been challenging? - How has your thinking changed? - Do you see turning points? - How do you see your development as in-house developer from now on?	Gallery presentations by two in-house developers who used the radar tool 3/17 discussion episodes

We framed out brief episodes that did not reveal additional information of the use, other than confirming the fact that the radar tool was used by the in-house developers beyond our workshop assignments. In this way we ended up to 17 discussion episodes of the workshops in October 2007, February 2008, and May 2008, from which the excerpts in the following analysis are drawn (Table 2).

The tasks and methods in Table 2 give the context to the discussion episodes, whereas the space of this paper does not allow the detailed description of the pedagogic process. All Radar-related assignments addressed the workplace development projects that the in-house developers were expected to carry out while learning the methods in the workshops and in tutor groups. It is noteworthy that the introduction to the foundations of the work development and the implementation of the workplace projects were intertwined throughout the process, and the use of Development Radar was to support both. Evidently the way of dealing with Development Radar and the questions addressed varied qualitatively from one workshop to another. Starting by relatively simple questions concerning the schedule and the phase of the workplace project (October 2007), then combining the tool with the use of another tool and

addressing more complex questions (February 2008), and, finally, observing the voluntary use of the tool for the personal reflection (May 2008) will presumably produce the variety of observations on the cultivation and enrichment of Development Radar. In order to keep the developmental logic, the observations are narrated and analyzed chronologically.

Transcription conventions in the excerpts are: [text in brackets] = clarifying addition; (--) = talk has been cut out; (text in parentheses) = incompletely audible; underlined = stress by speaker; **bold text** = key expressions.

ANALYSIS

October 2007 Radar Round

The first Radar round was carried out as a warm-up of the session, primarily to test the usability and usefulness [10] of the new Development Radar tool. To give an idea, this is how one of the in-house developers answered (questions on Table 2).

Excerpt 1

Developer 1: [Question 1] We have proceeded so that on this radar [holding the Radar figure] we are in the modelling and piloting phase. Next Monday we will have this modelling finished and presented to our personnel. But simultaneously we have expanded our task a little bit, as we have this big organizational change (--). It causes us quite many breaks, sort of, so that we do not know who is supposed to take care of what. [Question 2] So last week 25.10., we decided to start to keep a problem log, so, in a sense, we went back to the collecting of mirror material. And they [the notes on the problem log] are discussed weekly in our short meetings and we will agree on further actions. Altogether, on the one hand we have proceeded, on the other hand returned back [gestures the movement on the cycle]. [Question 3] And about this management question, I had an opportunity on 8.10. to present this project to the management board even by showing one triangle [model]!

The episode above reveals many of the observations we made on the early use of Development Radar.

Cultivating

Cultivating: Calendar use

Using Development Radar for marking critical dates was typical to all project presentations. It was the usage that we researchers expected to take place and what we ourselves did when implementing the radar tool [14].

Cultivating: Following developmental steps

The steps of the expansive cycle in the middle of Development Radar (steps 1–6, Figure 1) were used to articulate the present phase of the workplace development project, partly in an evaluative manner. Some comments took up the development dynamics of the expansive cycle more strongly than others, as demonstrated in excerpt 1: “*On the one hand we have proceeded, on the other hand returned back.*”

Cultivating: Administrative use

The discourse in this phase of Forum’s work had a shade of organizational formality, which we identify as an administrative use of the development tool. The language of planning, not surprisingly considering the assignment, and slight target-group thinking characterizes this type of use. “*Next Monday we will have this modelling finished and presented to our personnel*” (excerpt 1).

Cultivating: Non-use

In one case out of six workplace projects we observed non-use. Regardless of the “Radar round” assignment, the “Development Project Portfolio” was preferred to Development Radar. It was a tool of a table format displaying the pool of developmental tasks, one row of the table for each, designed by the in-house developers of the given organization.

Enriching

The concluding discussion after the Radar round produced two episodes that opened up enriching options for using Development Radar in Forum’s work. This means that Development Radar was evaluated and reflected upon concerning its further usability.

Enriching: Pendulum between the steps

The pendulum as an enriching aspect grew out of an episode during the Radar round, reported as ‘Cultivating: Following developmental steps.’ The need of returning to a previous phase of the development cycle was discussed further.

Excerpt 2

*Developer 2: About that moving on and getting back, I somehow think that the more one is proceeding on this Radar, the bigger the risk, or wish, or desire to return backwards. Then **the pendulum just broadens**, or rather I am – I’m sure I may speak for [my colleague] as well – **we are strongly prepared to move back and forth.***

Enriching: Comparison between the workplace projects

Another question addressed more directly the usefulness of the Radar round and the usability of Development Radar. One of the in-house developers phrased this issue as the possibility to compare the proceeding of the workplace projects, which helps evaluate one’s own pulse of development.

Excerpt 3

*Developer 3: Well it [“Radar round”] is useful, **it gives a mirror to one’s own project**, that “oh, those guys are already on that phase, I wonder what they have done differently, or should we possibly act in a different way...” But, on the other hand, the understanding here is that, as we deal with different development topics, also the **rhythm of development differs**. So I find it most interesting **to hear where each of us is proceeding** – even though we seem to be the last! [Laugh, joking who else felt being the last.]*

February 2008 Radar Round

The second Radar round was held three and a half months later. In between there had been one Forum workshop in November and a lengthy pause in the turn of the year. Now the assignment and working with Development Radar tool was carried out more in depth (Table 2). Moreover, to provide sufficient space for presenting the workplace development project we gave two tools, Development Radar to model the process on different levels, and Implementation Platform. The latter one has been used in some other DWR projects meaning the expansive cycle with space for making notes on the achievements, outcomes and plans in each phase. In this tool, the steps of the cycle are exactly the same as in Development Radar, but the articulation is different – one of the numerous variants of the cycle typically designed for different purposes.

We did not give any detailed directions to use each of the tools. Of the representatives of five projects present in the workshop, four used both tools and moved easily from one to another. The fifth, the previous non-use case, used both the tools offered and their own Development Project Portfolio.

Cultivating

Compared to October, the discussion episodes reveal much more multifaceted image of the use of Developmental Radar. Development projects are discussed on two levels, the workplace and the local tutoring. The administrative use exclusively on the workplace level is not observable, but references to the workplace variants of Development Radar are heard. The non-use case continues the successful cultivation of own tool.

Cultivating: Calendar use

Development Radar was continuously used to put down the critical dates and events. Now the calendar notes moved between the levels of workplace and local tutoring. In the case of excerpt 4, the workplace level seemed even to include two emerging “levels,” departments A and B.

Excerpt 4

*Developer 4: On the **workplace level** we have met intensively (--). And the first **local tutoring** we had in the **turn of August and September** (--), and the **28th of February** we are going to meet again. (-- And then we will have a **critical situation in 20.2.** when we present this project to **Department B**. How we are going to start to work with **Department B** on this same issue that we have been working on with **Department A** (--).*

Cultivating: Following developmental steps

We may observe the similar way of reporting on moving along the phases of the workplace project as in October. This time it was explicitly asked (according to assignment, Table 2), whereas in October the comments arose more spontaneously.

Cultivating: Workplace variants

The participants had not brought any examples of Development Radars from their workplaces. Two references were made to different variants that may be paralleled to the calendar use, the expected cultivation of the radar tool by “filling in” different contents. Examples were a calendar with critical events, and a resource plan. The latter was addressed to the management board of the workplace presenting the resources needed for the workplace project. (Bringing the developmental tool to the workplace to be used in communicative tasks of various kinds may alternatively be interpreted as an enriching action, which needs further examination.)

Cultivating: Non-use

One workplace project was consistently cultivating and enriching its own tool, Development Project Portfolio. In the Radar round presentation, the in-house developer showed both Development Radar and Implementation Platform, but did not discuss them. In the discussion that followed, the participants of Forum were interested in the portfolio tool and strongly encouraged these in-house developers to keep using the tool natural to their organization and its working culture. Moreover, it was agreed that the portfolio tool will be sent to the Forum participants, in case others would find it usable (which was the case at least in one project). Excerpt 5 gives an idea of the cultivation of an alternative tool and the development logic involved.

Excerpt 5

*Developer 5: That first version was just a table with the title row, but now **this has started to grow**, as we write down, for example, in the last meeting, “Create the procedures and principles of supporting training,” we agreed that “the training (pool) will be printed out for managers’ information during this year before the next appraisal interview.” And here we have put down the responsible person. (-- And here we make notes and give instructions, and this is really a **good tool***

for me. I will then go and “clatter” [at the responsible person] when the deadline is at hand, asking what has been done.

Enriching

One of the enriching episodes concerned the pendulum between the phases of development, as discussed already in October’s workshop. Another addressed the new development activities arising in the workplaces. The third enriching episode discussed the possibility to use Development Radar for seeing the outcomes and the impact of the change. All enriching episodes were associated to the emergence of the pilot groups on the workplaces, and how the radar tool might better serve them, not just the in-house developers in the Forum’s work.

Enriching: Pendulum between the steps

Pendulum in October was discussed mostly in the “in-house developer driven” manner, as the need for returning, for example, to collect more mirror material to specify the development task (excerpt 1). In February, it focused on the tension between the development process of the in-house developers and the rest of the workplace community, especially the pilot groups, participating in the implementation of the project (excerpt 6).

Excerpt 6

*Developer 6: In some sense, we are having **the phases of two processes**; this group [of in-house developers] has already processed it and gone pretty far, and now, in a sense, **we have to return backwards when starting the pilots.***

Enriching: Cultivating multiple versions

This enriching episode grew out of the discussion on one of the workplace projects, where the piloting phase was to be distributed across two working units (see excerpt 4) having different methods and schedules of implementation. One of the in-house developers put forth an idea of cultivating many Development Radars, one for each group.

Excerpt 7

*Developer 4: ...it came to my mind right now, that, as we have these **multiple models** – Developer 3 [same workplace] has one filled-in model, I have one – so it came to my mind that we could have one Development Radar model for us in-house developers, we could have one radar to our coming pilot team [Department B]. (-- And one radar in Department A. Because we all are on very **different levels**. (-- Then we would have **two, three radar models to reflect on.***

This idea gained support from the participants. It was elaborated by the local tutor who guided the in-house developers in question. His comment links the emerging development activities to the enriching-as-pendulum; not only pendulum for the in-house developers, but pendulum as the asynchrony of development of participating groups (excerpt 8).

Excerpt 8

*Local tutor 1: Using **different points of the radar for different groups** would show this asynchrony, which is true. (It may be that) for the manager or management or for some outsider, the project looks very advanced in some respect, but then a **new group emerges, the pilot group**, for example – it’s anything but self evident that it will question the old activity. It may be happy with the old, and if the cycle of expansive learning won’t start, the whole community won’t get anywhere. **Those who are active will progress on their cycle, but the next one will always start from the beginning.***

Enriching: Making change impact visible

The presenter of this enriching aspect suggested adding a dimension into the radar tool that would display the outcomes of development. The play with words cannot be literally translated (kehittämis – kehittymis). Basically the present Development Radar was strongly associated with planning and implementing deliberate development actions – how the things are to be developed. Developing Radar would reveal the change and the outcomes of development on different dimensions – how things have developed and what is the impact of change.

Excerpt 9

Developer 7: It is actually quite interesting, I think, now that we have this Development Radar, we would put here Developing Radar, put something inside it, the change, for example, and see the change, how it works in different dimensions, and how the impact of change – somehow depict it here. So that those in-house working groups that are designing new services [in their workplace project] could recognize the change. To see the dimensions of talk(?), somehow, and the impact, on what level it moves, maybe to see the change.

Researcher 1: Feel free to modify that as far as you can, even though I know that it is a little bit inflexible.

Developer 7: I'm fascinated by the idea of making that dimension visible.

May 2008 Learning Paths Gallery

The last episodes are from the learning workshop, in which the participants were asked to make visual representations of their learning paths during the 1,5 years process. Two in-house developers, Developers 3 and 8, used Development Radar for this meaning making, one building on the idea of pendulum, the other emphasizing the developmental steps. Taking the radar tool into use for reflecting personal learning, as an in-house developer, represents simultaneously both cultivation and enrichment (Table 3).

Cultivation and enrichment: Pendulum between the steps

Developer 3 redesigned Development Radar by replacing the radar metaphor with a pendulum clock. A pendulum metaphor does not support the idea of levels designed in Development

Radar, but the levels of the in-house developer, work place, and the instruction of Networkshops (Figure 1) are identifiable in a new form.

Excerpt 10

*Developer 3: This picture is the picture of my in-house developer. It is from Radar, but it is a part of a pendulum clock. That pendulum clock is [our work organization]. And this machinery keeping the pendulum in motion is the theory instruction we have got. On the pendulum, there sits a happy in-house developer with a triangle [model] in hand, that she will never lose grip of. **The pendulum motion is not going one direction, but it is possible to come back, a little bit, and it's not a bad motion. And again forth and back.** That's how it has been. We have tried and come back and tried again. Of course we want to go ahead, but I'm sure that when the spring loosens a little bit, it will be wound up, and then again we start to move back and forth with some new project. But it is safe to sit there, smiling, and swing.*

Cultivation and enrichment: Following developmental steps

Developer 8 had a sketchy radar model by means of which he discussed the importance of seeing the phases of development activity. (In fact the phases are graphically weakly represented whereas the circles for the levels are visible.) Like Developer 3, also Developer 8 associated the use of the radar tool with the self-confidence resulting in good development outcomes.

Excerpt 11

*Developer 8: This depicts my in-house developer. When I joined the learning network of South Savo, I felt insecure. So much theoretical knowledge available, but somehow I couldn't structure and internalize it too well. (-) But then this Development Radar was extremely important for me, something that clarified my learning process. **From it I could clearly see what we had achieved, where we are now, and what should be addressed next.** That's when I had a light bulb moment, in a way; and what followed was our development workshop in February, that made me nervous in beforehand, whether it will succeed and how it will go. But then it really succeeded well, the results were good and we made really good [products developed in the workplace project], that were implemented in April.*

Table 3. Cultivating and enriching Development Radar – summary.

Use	October 2007	February 2008	May 2008
Cultivating	<ul style="list-style-type: none"> calendar use following developmental steps administrative use non-use 	<ul style="list-style-type: none"> calendar use following developmental steps non-use workplace variants 	<ul style="list-style-type: none"> following developmental steps
Enriching	<ul style="list-style-type: none"> pendulum between the steps comparison between the workplace projects 	<ul style="list-style-type: none"> pendulum between the steps cultivating multiple versions making change impact visible 	<ul style="list-style-type: none"> pendulum between the steps

DISCUSSION

The findings of the analysis on the cultivation and enrichment of the radar tool highlight both pedagogical and design issues. One question of pedagogical concern for us researchers as “learning-interventionists” runs: Why is it that our presentation of the cycle model as the basis for the development projects had not the effect comparable to the reception of Development Radar? The dilemma is that when the radar tool was taken into use, it was often used like the developmental cycle (moving on the workplace level only). “On Radar we are situated...” might often be replaced by “on the developmental cycle.” The effects of displaying the multiple levels may now be reflected on in the light of this analysis.

Our data (beyond this analysis) shows that the theoretical contribution of Forum was questioned by the participants in Networkshop of September 2007, between the introduction of Development Radar in August and the Radar round in October. It was difficult to combine the theory instruction with the workplace development projects; specifically, the phase of the instruction in Forum was not in synchrony with the phase of one’s workplace project. Development Radar may have directly responded to this dilemma by crystallizing [11] the in-house development activity. First, the recognizable levels gave perspective to the context of the in-house development and made the model “our tool” in the eyes of the in-house developers, whereas the cycle model represented the “difficult theory” offered by researchers. Secondly, by seeing the levels the in-house developers could realize the asynchrony of development; asynchrony that was not only necessary but acceptable and dynamic. This came out in the notion of pendulum as will be discussed in Conclusions. The third point resides in the change of the visual design of the cycle giving space to movement, which is analyzed in our previous paper [14].

Another question is related to the contentual design of the levels of the radar model. Why didn’t the in-house developers explicitly question the levels designed by the researchers, in view of the fact that they clearly expressed the need for enriching the levels of the workplace project? Lacking the workplace-level data (we hope to gather in the follow-up of the learning network) we cannot assess to what extent the in-house developers actually modified the radar tool to match with their needs, and whether Development Radar had plasticity [11] allowing to do so. It was not until carrying out this analysis that we realized the enriching initiatives of the participants, mainly emphasizing the pendulum between the steps, cultivating multiple versions, and making change impact visible. While

considering how we can elaborate on these, we are planning to carry out a new type of Radar round in the coming Networkshop of Forum. This time we are going to ask the in-house developers together with their appointed local tutors to examine the levels of Development Radar and re-design them according to the levels emerging in their workplace projects. The group of in-house developers is other than the group analyzed here, but some of the local tutors are previous in-house developer participants, which may strengthen the continuity of cultivation and enrichment of the radar tool.

CONCLUSION

Cultivating and enriching Development Radar tool in the learning process of in-house development gave the researchers and participants interesting knowledge of the qualities and potential of the tool that was designed to facilitate the development activity. The use of the tool also revealed new

features of the development itself and changed the participants’ perspective to their assignment. For the in-house developers, it seemed to be a commonly shared enriching insight that development of work activity does not always mean moving onward; progressing is also returning to the steps already left behind. Whether this insight formed a divide between the users and “non-users” of the tool remains to be evaluated through the follow-up study. For us researchers this movement may have been known on the academic level, but we gained insight in the significance of this dynamic and in the ways of mediating this to the students of workplace development.

Paying attention to the cultivating and enriching episodes revealed some asynchrony of the researchers’ expectations and the actions and outcomes produced by participants. The levels displayed on the radar were considered focal by the researcher-designers, whereas the in-house developers’ needs were mainly on the workplace level and gradually expanded to the level of local tutoring. The cultivation of a tool made the reality of development more visible and provided common ground to discuss the expectations and needs.

The findings of a longitudinal analysis suggest that both cultivating and enriching actions evolved towards diversifying perspectives on the in-house development activity. This arose primarily from the evolving workplace projects and their growing complexity, which was mediated by the radar tool and observable in the workshop discussions. The transition from the charting and analyzing phases by the in-house developers to the experimentation phase to be carried out by the pilot groups at workplaces (phases in Figure 1) was interestingly represented and articulated in the use of Development Radar, particularly in the enriching initiatives in the February workshop. We are going to discuss the findings in terms of *the multi-mediation of learning facilitated by (the use of) the tool*. These are tentative notions:

Mediating horizontal learning: Comparison of projects, knowledge of workplace variants and reversely even the non-use of the given tool are examples of the cultivation that mediated peer-to-peer and workplace-to-workplace learning typically associated to the network learning.

Mediating change: As already discussed, the notion on the pendulum and asynchrony of development and learning on different levels of activities provides powerful evidence of the process, in which the cultivation of a tool may lead to an enriched perspective on the dynamics and challenges of the in-house development.

Mediating emerging levels: Emerging new levels of activities gave impetus to the tool-enriching episodes indicating the progress and growing complexity of development in the work communities and networks, which the users wanted to “include” in the tool.

Mediating the in-house developer’s “changemaker” identity: An open-minded experimentation on development tools produces surprising outcomes, such as linking the in-house developers’ tool use with the feelings of safety, self-confidence and optimism amid turbulence on workplace projects.

Learning and development are fundamentally intertwined presupposing each other in the Developmental Work Research projects. In the empirical applications there is a risk of focusing on developmental outcomes and even on external time constraints of development while leaving the dynamics of and reflection on learning on the background. The co-configurative design and use of developmental tools may expand the dialogue of educational settings by making learning materially visible.

ACKNOWLEDGMENTS

Our thanks to project coordinator Tuula Syrjälä and other participants of the Forum for In-house Development of South Savo learning network, and the Finnish Workplace Development Programme TYKES (2004–2009) at the Finnish Funding Agency for Technology and Innovation TEKES.

REFERENCES

- [1] Béguin, P. 2003. Design as a mutual learning process between users and designers. *Interacting with Computers*, 15, 709–730.
- [2] Béguin, P. and Rabardel, P. 2000. Designing for Instrument-Mediated Activity. *Scandinavian Journal of Information Systems*, 12, 173–190.
- [3] Daniels, H., Leadbetter, J., Warmington, P., Edwards, A., Martin, D., Popova, A., Apostolov, A., Middleton, D. and Brown, S. 2007. Learning in and for Multi-Agency Working. *Oxford Review of Education*, 33(4), 521–538.
- [4] Engeström, Y. 2007. Putting Vygotsky to Work: The Change Laboratory as an Application of Double Stimulation. In: H. Daniels, M. Cole and J. V. Wertsch (Eds.). *The Cambridge companion to Vygotsky*. New York: Cambridge University Press. Pp. 363–
- [5] Engeström, Y. Engeström, R. and Vähäaho, T. 1999. When the Center Does not Hold: The Importance of Knotworking. In: S. Chaiklin, M. Hedegaard and U. J. Jensen (Eds.). *Activity Theory and Social Practice: Cultural-Historical Approaches*. Aarhus: Aarhus University Press.
- [6] Engeström, Y., Lompscher, J. and Rückriem, G. (Eds.). 2005. *Putting activity theory to work: Contributions from developmental work research*, Lehmanns Media, Berlin.
- [7] Hasu, M. 2001. *Critical Transition from Developers to Users. Activity-Theoretical Studies of Interaction and Learning in the Innovation Process*. Academic Dissertation. University of Helsinki, Department of Education. Espoo: Otamedia Oy.
- [8] Kaptelinin, V. and Kuutti, K. 1999. Cognitive Tools Reconsidered: From augmentation to mediation. In: J. P. Marsh, B. Gorayska and J. L. Mey (Eds.). *Humane Interfaces: Questions of method and practice in Cognitive Technology*. Elsevier. Pp. 145–160.
- [9] Kerosuo, H. and Engeström, Y. 2003. Boundary crossing and learning in creation of new work practice. *Journal of Workplace Learning*, 15(7/8), 345–351.
- [10] Landauer, T. 1995. *The Trouble with Computers: usefulness, usability, and productivity*, Cambridge: The MIT Press.
- [11] Prost, L., Cerf, M. and Jeuffroy, M.-H. 2007. Sharing the Design of a Cognitive Tool between Researchers and Potential Users. *Proceedings of the ECCE 2007 Conference*, London, UK. Pp. 67–74.
- [12] Ramstad, E. 2009. Developmental evaluation framework for innovation and learning networks: Integration of the structure, process and outcomes. *Journal of Workplace Learning*, 21(3), 181–197.
- [13] Syrjälä T., Kerosuo, H. and Toiviainen, H. 2008. Innovation beyond and learning across workplaces. Paper in *Insightful Encounters – Regional Development and Practice-Based Learning Conference on Regional Development and Innovation Processes March 5th–7th, 2008, Porvoo–Borgå, Finland*. Available: www.congreszon.fi/@Bin/1674521/3.Innovation.Syrjala.Kerosuo.Toiviainen.pdf (accessed 21 July 2009).
- [14] Toiviainen, H., Kerosuo, H. and Syrjälä, T. Development Radar – The co-configuration of a tool in a learning network. *Journal of Workplace Learning*, 2009. (In press).
- [15] Victor, B. and Boynton, A. C. 1998. *Invented here: Maximizing your organization’s internal growth and profitability*. Boston, MA: Harvard Business School Press.
- [16] Vygotsky, L. S. 1978. *Mind in society: The development of higher psychological processes*. Cambridge, Mass.: Harvard University Press.

The Qualitative Differences of the Effects of Technological Changes: Case Wood Procurement Process

Arja Ala-Laurinaho
Finnish Institute of
Occupational Health
Topeliuksenkatu 41 aA
FI-00250 Helsinki, Finland
arja.ala-laurinaho@ttl.fi

Marika Schaupp
Finnish Institute of
Occupational Health
Topeliuksenkatu 41 aA
FI-00250 Helsinki, Finland
marika.schaupp@ttl.fi

Arto Kariniemi
Metsäteho Oy
Snellmaninkatu 13
FI-00171 Helsinki, Finland
arto.kariniemi@metsateho.fi

ABSTRACT

We investigate the nature of the effects of technological changes on forest work activity, focusing on the effects on forestry professionals' tools and object of work. The analysis is based on the activity theoretical approach. The forest work has transformed into a technology mediated work, and the entire concept of the wood procurement process is in transition phase, accomplished by several independent companies operating in close co-operation and using shared on-line computer systems for coordination and planning of the process. Our case proves the need of more holistic and object-oriented design approaches in technological changes. There is a need to understand the activity of the entire network, and the changes in the object of work of the actors in that network in order to enhance new operating concepts and practices that the new technology would enable.

Keywords

technological change, change of work, wood procurement, forest work, activity theory

INTRODUCTION

At the beginning of the 20th century, there were the man, the horse and the axe. At the beginning of the 21st century, there are the man, the computer system and the harvester, as part of the wood procurement network. What is the nature and significance of the many technological changes that have happened during the last few decades in the forest work? Development based on mechanization and rationalization of work processes has happened in a couple of distinct technological cycles and decreased the cost per unit of timber (Figure 1). The most recent technological cycle includes the use of information technology, which has increased drastically during the last decade. What has mechanization brought to the industry besides increased efficiency? In this presentation, we investigate the nature of the effects of technological changes on forest work activity.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

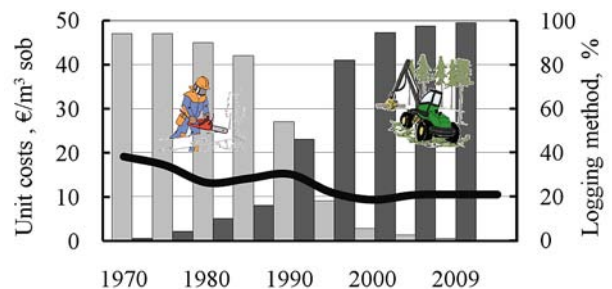


Figure 1. Real per unit cost development of logging and proportion of logging method in Finland.

The analysis is based on the activity theoretical approach, which suggests that collective work activities are motivated and defined by the object and the outcome of the activity [3]. As a case example we use procurement process and its social-technical environment. The data for the analysis is mainly collected from a developmental intervention in a wood supply district in Middle Finland, where we collectively investigated the changes in wood procurement together with the different actors in the process.

It is precisely the information technology that has enabled major organizational changes in wood procurement – among others the new division of work between the actors in the process. Technological innovations are often considered as a source of efficiency by being an improved tool for the activity. However, some technologies have the potential to start changing the whole logic of activity. In the presentation, we will focus on the effects of technological changes on forest professionals' tools and object of work through three cases, namely the work of the forest machine operator, the harvesting foreman and the forest entrepreneur. We will describe how, due to the new information technologies, the forest worker's traditional object of work – cutting and hauling the logs – has transformed into a technology mediated work, in which, besides operating the forest machine that incorporates highly advanced wood processing technologies, generating and processing of data for the collective information system has become a central part of the work. Also, we will show how the work of harvesting foreman has changed from face-to-face supervision of work and order giving out in the forest to computer-based production planning and management, which is mainly carried out in the office, and finally, we will depict how hierarchically supervised contractors have become professional entrepreneurs.

DATA AND METHODS

The Case and Intervention

The data was collected in a development project called "Developing the concept of wood procurement process – supporting the model of regional entrepreneurship and enhancing the well-being of forest machine operators" in 06/2008–03/2009. The project was organized as a Change Workshop intervention. A Change Workshop is a space for collective analysis of the shared work practices with the help of external interventionists, whose role is to support the learning process and not to provide any ready-made solutions. In this case the process included five 3-hour sessions. The first four of the sessions were organized at intervals of two or three weeks and the fifth one, which was an evaluation session, after three months experimenting period, during which the participants tested the ideas generated in the workshop in practice. The Change Workshop is an application of the Change Laboratory® method [2].

The intervention was based on the activity theoretical idea that in order to resolve an aggravating contradiction in the activity system (which manifests itself in problematic or unfeasible situations in everyday work) in an expansive way, the actors need to question the old activity, analyse the historical roots of the emerging contradiction, reinterpret the object of their work and build new concepts, tools, forms of collaboration and division of labor to support the new object. A central element in a Change Workshop is the concrete case material that mainly the participants but also the interventionists collect from the participants' daily work to be analyzed in the sessions. The participants were given assignments between the sessions to collect data especially about, on the one hand, the problematic situations in the work, and on the other, the innovations they have made to cope with or to overcome these problems. In the sessions we offered the participants (theoretical) models and representations (e.g. the activity system, see Figure 2) to help the participants to reorganize the concrete data and observations about the work, to see problems and dilemmas in a new light and to generate qualitatively new solutions to the problems. The participants also depicted the central elements of their work first in 1998 and then in 2008 using the activity system model. The idea was to make the changes, which often gradually sneak into the daily work, visible and more concrete. During the experimental period the participants discussed development proposals and planned the selected development tasks, and then tried the ideas in practice.

There were 10–15 participants in the workshop, a regional manager, a deputy regional manager and 6 field officers representing the forest department of a nationwide forest company, three harvesting and long-distance transportation entrepreneurs and three timber truck drivers/harvester operators. The data consisted of interviews and discussions in the workshop that were recorded and partially transcribed, visits to the regional forest department of the forest industry company and to a harvesting and long-distance transportation enterprise, as well as memos of the sessions and a report that were commented and checked for validity by the participants.

Object-Oriented Approach to Understand the Technological Changes

We used the activity theoretical approach and the model for activity systems (Figure 2) [1] in the analysis of the data. The core of the approach is that human activity is always object-oriented. The activity system forms a collective context and the shared object a durable motive for the more short-lived and

goal-directed actions. The subject's interaction with the object is mediated through signs and tools (including material artifacts and concepts and theories), rules, community, and a division of labour. These mediators carry cultural meanings and historical development within them and thus offer stability to the system.

The elements of an activity system are in constant interaction reshaping each other. Because every activity is connected to other activities, for example those of the customers, new elements and requirements can enter an activity system causing disturbances and deviations from the customary scripts of the activity. The innovations through which the new situations are managed start then to reshape the ways of working. Thus the activity system should be considered a dynamic entity that interacts with other activity systems. The wood procurement process was analysed as a network of activity systems connected to each other by the overarching object of the entire network.

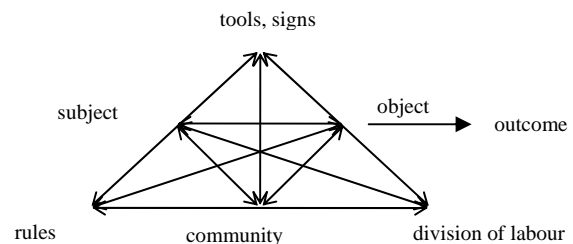


Figure 2. Model of the activity system [1].

TECHNOLOGICAL DEVELOPMENT TRENDS IN WOOD PROCUREMENT PROCESS

The technological development can be divided in three main paths. First, developing the haulage of the logs, which meant replacing horses with tractors and lorries, and later, with specifically forestry designed vehicles like forwarders and timber trucks. Second, developing the logging methods and tools from manual to motor-manual devices, and finally, to the single-grip harvesters with their high-tech tools and devices and ergonomic design. The third line of development concerns the information processing and flow of information in the wood procurement process, from oral instructions of the foreman about the desired log dimensions and thinning methods to the current detailed and on-line data flow between all the participants of the process – crossing the organizational borders of the wood processing mills and the harvesting and long-distance transportation enterprises, as well as the in-firm boundaries of managers, production planners, field officers, foremen and workers.

Periodically, these development paths are overlapping, having also effects on each others' progress. In Finland, the period of the ascendancy of frame saw and axe, as well as horses, reached its end in the beginning of 1960s (see also Figure 1) when hand tools were superseded by chain saw and horses with farm tractors. The second milestone was in the middle of 1970s when forwarders became more popular hauling technology than farm tractor based ones. The third milestone was the breakthrough of the single-grip harvester technology in the middle of 1980s. Only then became the harvesting by machines more effective than manual work by lumberjacks (e.g., in Finland the mechanization level of felling was only 16% in 1985, however reaching 46% at the end of the decade). Nowadays, practically all harvesting is carried out by single-grip harvesters and forwarders. The usage of information

technology can be seen as the forth milestone in the technological development, starting already in 70's with measuring techniques and management of crosscutting as first applications. Utilization of information technology was one of the key factors that enabled the success of the single-grip harvester, and today's shared information systems and on-line connections with up-to-date data concerning e.g. log information (dimensions, quality, quantity), transport possibilities, and digital maps of stands marked for cutting offer radically new possibilities to organize the wood procurement process. [6, 7]

EFFECTS OF TECHNOLOGICAL CHANGES ON WOOD PROCUREMENT PROCESS AND WORK

In this chapter, we will focus on the effects of technological changes on forest worker's work activity. We will show how harvester operators', harvesting foremen's and forest machine entrepreneurs' work and tasks have transformed into a technology mediated information work with data processing as an essential part of the wood procurement process.

Organizationally, the forest industry companies used to take care of the entire procurement process from forest to the factory. The forest departments of the companies bought the stands marked for cutting, planned the harvesting schedules, and organized both the haulage in the forest and the main log haul to the sawmills, pulp mills etc. industrial users. The forest workers were employed by the forest industry companies and the forest machinery was owned by contractors who worked under close hierarchical supervision of the foremen of the forest company.

During the last decade, there have been major changes in the organization of the wood procurement process, especially concerning the division of tasks and responsibilities of the actors in the process. The major driver of this development has been the implementation of IT systems and mobile technology, which have formed an integrative planning, control and information network covering the entire wood procurement process from forest to the mills. This change into a network-based co-operation can be observed in the changed tasks of harvester operator, harvesting foreman and entrepreneur, all closely connected to each other.

Until the end of the 90's, the main tasks of harvester operators and forwarder drivers were timber harvesting and hauling. They got the instructions for work from harvesting foreman of the forest industry company, usually face to face at the worksite. The work resulted to cut-to-size logs piled at roadside landing. Though during the decades the tools developed from axes to chain saws to first mechanized felling machines, the main tasks and targets remained basically the same. The new tools did make the work more efficient and safe, and improved ergonomics diminished especially the physical workload. The first applications of IT concerned measuring techniques of timber, and were used e.g. to support decision making on the most valuable combinations of log length and thickness, and thus improved the economic results of the harvesting. Still, those applications did not change the division of tasks in the forest work. [5]

The mobile and on-line information systems were intensively developed, and since the beginning of the 21st century, they allowed more and more work and interaction to be done via information systems. For example, all the instructions, information and maps on stands and worksites were transformed into the harvester cabins via telecommunications connections, and, when needed, the harvesting foremen gave clarifying instructions by

mobile phones. This radically released the resources of harvester foreman to be used in other tasks than travelling from stands to stands. Their work profile changed from face-to-face foreman on the fields into production planner with duties of short term planning and reporting, mainly accomplished using computer system within offices.

The new technology has significantly affected also the work of harvester operators. In addition to harvesting and hauling, crucial work tasks include now also worksite planning, quality assurance, and feeding and processing data as well as sending it further to the information systems of the wood procurement network. This means more independent work with enlarged responsibilities concerning especially quality of work and accuracy of information. The object and target of work changed from being merely timber piled on roadside into the raw material with specific destination mill and information on that raw material to be used in the further phases of procurement process, and even in the production planning of the destination mills.

At the same time, due to the new flexible way of organizing the work the information technology offers, the role of entrepreneurs in the wood procurement process has step by step changed from contractors who worked under close hierarchical supervision of the foremen of the forest company into professional entrepreneurs. They used to take care only of the execution of harvesting and hauling, and maintenance and transportation of the forestry machines. Nowadays they also do the fine scheduling of process, including both the planning and the execution of harvesting, hauling and long-distance transportation (depending of the scope of their business). A key prerequisite for this is the shared information system of the procurement network, as all the important data concerning stands marked for cutting, short term plans and schedules are available also for the entrepreneurs. As entrepreneurs, they are also responsible of quality and quantity requirements of their assignments. The procurement process is, thus, accomplished by several independent companies operating in close collaboration and using shared on-line computer systems for coordination and planning of the process.

DISCUSSION

Though the effects of technological changes on work practices are largely acknowledged by the actors in the procurement process, many of the changes have been so gradual that it is hard to recognize the entire range and intertwined consequences of effects. For example, the participants of the workshop described many easy-to-see changes of their work, like increased use of computers in all tasks. Also, problems related to changes were connected to specific actions and work tasks rather to entire operation mode. As an example, the field officers of the forest company felt that the new technology-related reporting tasks hindered them from executing their field work in the forest.

The activity theoretical analysis, however, showed a significant expansion in each stakeholder's responsibilities in the process and an increase in more autonomous planning and decision making regarding the tasks. The analysis gave a deeper meaning and practical content also for the dimensions that depicted the major lines of changes. It also made the participants to realise that even though the work that they do today is quite different from the one they did ten years ago, the interpretation of what they "should be doing" still rested on the old idea of each participant's work and professional identity. This realization gave them justification to let go of the old responsibilities and start to develop their work practices based on the new idea of work and the possibilities it offered.

We claim that the entire concept of the wood procurement process is in transition phase as depicted in the Figure 3. It is precisely the information technology that allowed the major organizational change – the new division of work in the procurement process. Many tasks and responsibilities were shifted “downwards” in the hierarchy, which had started to dissolve the old hierarchical organization of work that was based on direct and personal supervision of the foremen. The material activity had already changed into network-like technology supported collaboration. But, their orientations were lagging behind in ‘work in the forest’ orientation. This is shown in the transitional conception where the actual lack of time was a result of the misfit between materially existing form of network collaboration supported by the technology, and the old orientation. There was a general agreement that such a change from the work in the forest to a technology supported network-like collaboration is going on. However, some participants were concerned of the possible negative consequences saying that they anyway prefer to have social relations in their work.

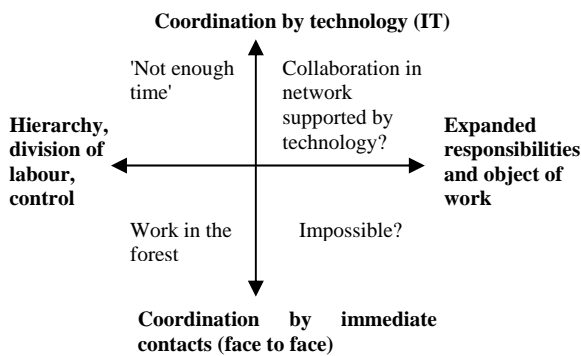


Figure 3. The change of operating concept of the wood procurement process.

The analysis of the case also brought up the critical question of designing, implementing and interpreting new technological applications and solutions. As the continuous, often as such gradual changes are implemented, at some point there is the risk of losing the benefits of the improved tools and systems. In order to get the benefits, the entire mode/concept of operation should be changed – the activity system, or the entire net of activity systems, should be analyzed and developed in relation to its historical developmental path. Treating the changes as merely of changes of tools will not be sufficient if there would be a need for change of the object of work and the activity system. For example, in our case, the chain saw and the early harvesters were predominantly changes of *tools*, though they had remarkable positive effects on ergonomics, managing of work load and efficiency of work. However, introducing the information technology changed radically the *object* of work: instead of just cutting of trees and logs, the new object of work included also information processing concerning both the stands of timber marked for cutting and the logs. This result is also in line with Pérez [10] and Freeman and Louçã [4] and many other researchers’ claim that the possibilities of IT may not be fully exploited if it is treated just as another improvement of tools, instead of radical new technology requiring also social innovations, new kind of production concepts, and new structures of organisations and institutions.

Norros (et.al) [8] has proposed in their writings on ecological design concept and later, on joint intelligent systems approach

[9] that the design should focus on the level of activity and systems instead of action and tools. They emphasize designing systems in usage, i.e. designing practices. This is especially important in case of information technology. Our case well proves this need of more holistic and object-oriented design approaches. It is not enough to design user-friendly interfaces on the bases of current work actions; instead, there is a need to understand the activity of the entire network of, e.g., wood procurement process, and the changes in the object of work of the actors in the network that the technological developments bring about in order to enhance new operating concepts and practices that the new technology would enable.

ACKNOWLEDGMENTS

The project was financed by Metsäteho Oy and Metsämiesten Säätiö. The field study was carried out in one of regional forest departments of Metsäliitto Osuuskunta. Participating of harvesting and long-distance transportation companies was essential.

REFERENCES

- [1] Engeström, Y. 1987. Learning by expanding. Gummerus Oy, Jyväskylä.
- [2] Engeström, Y. 2007. Putting Vygotsky to work: the Change Laboratory as an application of Double Stimulation. The Cambridge Companion to Vygotsky. H. Daniels, M. Cole and J. W. Wertsch, Cambridge University Press. Pp. 363–382.
- [3] Engeström, Y., Miettinen, R. and Punamäki, R.-L. (Eds). 1999. Perspectives on Activity Theory. New York: Cambridge University Press.
- [4] Freeman, C. and Louçã, F. 2001. As time goes by. From industrial revolution to information revolution. Oxford: Oxford University Press.
- [5] Kariniemi, A. 2006. Kuljettajakeskeinen hakkuukonetyön malli – työn suorituksen kognitiivinen tarkastelu. Helsingin yliopiston Metsävarojen käytön laitoksen julkaisuja 38. University of Helsinki. Summary: Operator-specific model for mechanical harvesting – cognitive approach to work performance.
- [6] Metsätalastollinen vuosikirja 2008. Finnish Statistical Yearbook of Forestry. Metsätutkimuslaitos. Finnish Forest Research Institute.
- [7] Konttinen, H and Drushka, K. 1997. Metsäkoneiden maailmanhistoria. Timberjack Group Oy. Otava, Helsinki.
- [8] Norros, L., Kuutti, K., Rämä, P. and Alakärppä, I. 2007. Ekologisen suunnittelukonseptin kehittäminen. In: Kaasinen, E., Norros, L (Eds.). 2007: Älykkäiden ympäristöjen suunnittelu – kohti ekologista systeemi-ajattelua. Teknologiateollisuus ry, Tammer-Paino Oy, Tampere.
- [9] Norros, L. and Salo, L. 2009. Design of joint systems: a theoretical challenge for cognitive systems engineering. Cogn Tech Work, 11, 43–56.
- [10] Pérez, C. 2002. Technological revolutions and financial capital. The dynamics and bubbles and golden ages. Cheltenham: Edward Elgar.

Can System Innovations be Facilitated by Societal Embedding?

Eveliina Saari

Technical Research Centre of Finland
P.O. Box 1000
02044 VTT, Finland
eveliina.saari@vtt.fi

Sirkku Kivisaari

Technical Research Centre of Finland
P.O. Box 1000
02044 VTT, Finland
sirkku.kivisaari@vtt.fi

ABSTRACT

As the concept of innovation has extended to cover social, organizational and system innovations, traditional innovation theories based on technology push or market-driven approaches are challenged. In this paper, we describe a research and development approach called societal embedding of innovations. It has been applied as an evaluation, research and development method in innovations which have been provoked by societal needs, such as health care, energy and environment. We present its main principles and ponder its future challenges.

Keywords

system innovation, societal embedding, learning, methods

INTRODUCTION

Concern for availability of high quality services at reasonable cost in the future has strengthened the need for sustainable change in health care system in all Western societies. In Finland, however, in spite of public policies and programmes encouraging system change and abundance of local experiments, innovations tend to reach only local scale. This paper explores how scaling up of innovations can be enhanced. Widening the scale of innovation from local experiments to a national level system innovation is a challenge which calls for initiating new kind of agency in the network of researcher-developers, service providers, users, and policy makers of various levels.

In this paper, we present a method called societal embedding of innovations. The method has been used as a research, development and evaluation approach in studies related to innovation that has been provoked by societal needs. We ponder to what extent researcher-developers can enhance the innovation process. They may play a part in reconfiguring or extending the network of actors who promote the innovation. They have also been said to activate processes of knowledge production and organizational learning [1]. Innovations are increasingly developed in complex multi-actor networks, and the role of R&D is becoming more dynamic and interactive. Knowledge is not just a part, which you buy from an R&D supplier and implement in the innovation process. Instead,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

innovations are more often co-created in the interactive learning processes in the multi-actor networks [2].

We open up the approach of societal embedding of innovations from the perspective of learning. We use the concepts and ideas of cultural-historical activity theory [3] to ponder the agency and object of a system innovation. In addition, the role of models and visual representations is perceived significant in facilitating learning in the innovation network. Finally, we explain how the role of a researcher-developer changes along the innovation path and in its different phases.

ROOTS OF SOCIETAL EMBEDDING OF INNOVATIONS

Societal embedding of innovations is a research and development approach, which aims at facilitating and initiating new sustainable innovations in a multi-actor network. The approach has been developed at VTT from the mid 1990s to promote the societal quality of potential innovations and facilitate the distribution of e.g. medical ICT solutions for wider use [4]. Typically these innovations needed new kind of collaboration between public and private sector in order to become sustainable. Since 1995 the concept of innovation has broadened to cover social and organizational aspects besides technological. At the same time industrial firms have gradually expanded their delivery from products to services. Innovation researchers realized that innovations, which were launched for the public sector, needed knowledge and understanding of political, legislative aspects as well as technological and social. The innovation network involved in the development process covered multiple actors from policy makers to potential users and service providers.

Multi-level perspective to study long-term technological and societal changes presented in transition management literature [5, 6] offered a broad framework for understanding the dynamics of system innovations. The perspective stresses that technological systems change as the interplay between landscape, regime and niche levels. Socio-technical landscape refers to relatively stable, slow-changing factors such as cultural and normative values, long-term economic developments and societal trends. Socio-technical regime refers to the semi-coherent set of rules carried by different actors such as users, policymakers, scientists, and public authorities. Niches represent the local level of initiatives and activity.

Transition management emphasizes both top-down and bottom-up processes [7]. On the one hand, it stresses the importance of defining common visions by the government. On the other hand, the approach emphasizes the need for diversity and experiments [6]. The scaling up of experiments has been pointed out [8] to be a critical phase in system innovation. We argue that widening the scale of innovation from local experiments to a national level system innovation is a challenge

which calls for initiating new kind of agency in the network of service providers, users, and policy makers of various levels.

When developing a sustainable innovation, multiple values and perspectives should be involved. The following figure is a simplified representation of the main actors and their main perspectives to innovation: (1) Providers: efficiency of service production process, (2) Users: usefulness and value to individual users (3) Purchasers: cost efficiency and correspondence with local needs, and (4) Societal actors: correspondence with substantial societal needs e.g. in terms of relevance and wide-ranging implications. Ensuring the societal quality calls for collaboration between different actors and stakeholders.

The societal value of the system innovation is co-constructed in a multi-actor network. Researchers' role relates to identifying and articulating the different perspectives of the actors who participate in development of the innovation who have something at stake, or who indirectly influence its development. By opening up the perspectives of the different actors we aim to produce mutual learning.

FACILITATING LEARNING BETWEEN ACTORS ALONG THE INNOVATION PATH

Societal embedding of innovation is a research and intervention method, which varies depending on the phase of the innovation process. In the following, we present a rough model of developmental phases of a system innovation and how researcher-developers could promote its development in each phase. The phases have been inspired by theory of expansive learning [3] and Dewey's model of reflective thought and action [9].

From Single Perspectives to Shared Understanding of the Problem

In order to create a new idea of an innovation, actors have different perspectives of the main problems in the current

activity or routine which should be solved. This pre-phase of creating an innovation benefits from workshops where the prevailing activity and its problems are made visible to all actors; users, providers, developers, and policy makers. A researcher-developer may show ethnographic analyses of the prevailing activity and provide statistical data about processes and their defects to trigger the problem-solving and questioning of the "old way of doing". Another important task for the researcher is to gather the innovation network together, and build and maintain interaction within the network.

A key question is to create a constructive dialogue between the different actors and give them a possibility to create a shared understanding of the elements of the solution. Different interests of the actors may become incompatible if they do not see the problems and its solution in a wider perspective than their own. Designing a locally attractive solution which could be scaled up is a challenge for the innovation network.

Designing a New Model and Its First Local Experiments

After gaining a shared understanding of prevailing activity and its problems in the multi-actor network, the researcher-developer may help the actors design a new model of the process, service or product to solve its problems. In this phase of innovation, the visualised plan is materialized into local experiments.

Learning from these local experiments between users, providers, developers and societal actors may be facilitated by a researcher-developer. Moving from the first experiments to the wider use of the solution is often critical. Common workshops of the innovation network may produce collective agency or help identify the dedicated actors to promote the innovation from the first experiments to its wider distribution.

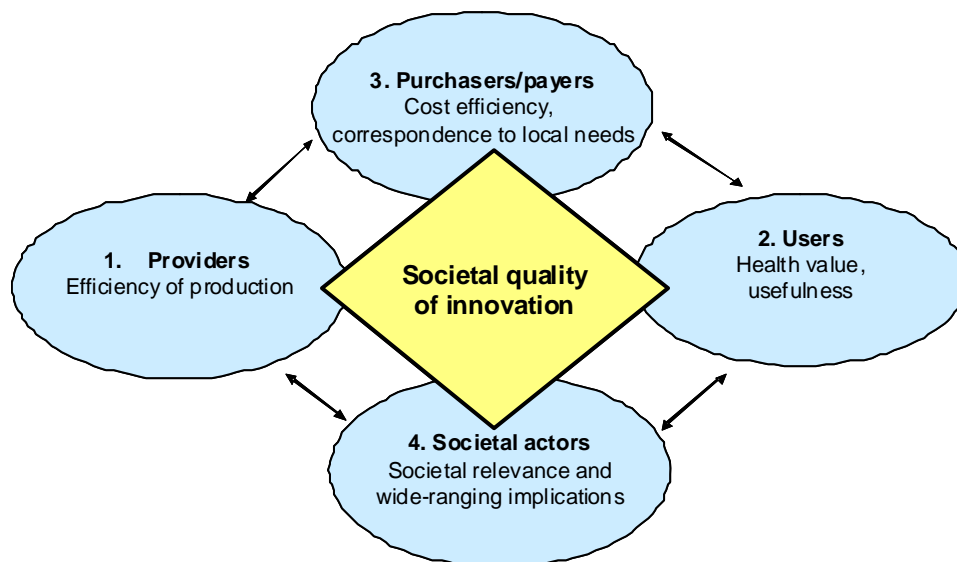


Figure 1. Key actors in societal embedding of innovation.

Expanding of the Experiments

Transferring a potential innovation into wider use is not a straightforward process. Often a new model of activity cannot be used as such in a new environment. The researcher-developer may help the actors identify which elements of the innovation embryo can be generalized and which parts need to be re-invented. New services and organization models consist of human activity, and should be adopted by a new local development network in order to be implemented.

The Figure 2 shows the developmental phases of an innovation as a learning cycle. In each phase, the multi-actor network of developers, purchasers, providers, users and societal actors are facilitated to have a dialogue on three aspects: What is the object we are developing? Who should be involved in developing it? How does this development object solve the prevailing problems and societal challenges?

Consolidation of the New Model and Evaluating the Impacts of Its Wider use

In this phase, the spread of the innovation can be evaluated and its societal impact can be analyzed with the help of local examples and larger scale evidence. Researcher-developer may facilitate developmental impact evaluation workshops [10] in which different perspectives of the multi-actor network are voiced. It is significant to evaluate how the innovation solves the current problems of the activity from the perspectives of the user and society, besides provider, purchaser and developer. This kind of pondering generates often new innovation ideas and the innovation cycle may start again.

CHALLENGES OF THE APPROACH IN INITIATING SYSTEM INNOVATIONS

Expansion of the Object in the Innovation Network and the Role of the Facilitator

According to cultural-historical activity theory, human activity is object-oriented [11]. The object is to be understood as a project under construction, moving from potential “raw form” to a meaningful shape and to an outcome. In science and technology studies, actor-network theory has analyzed innovation trajectories in which both the innovation itself and the network constructing, developing and distributing it, is transformed [12]. When we study and develop system innovations, e.g. new organizational forms between public and private health care, which are extended over single organization borders, the object of development is constructed by a complex and changing network of actors. The actor-network theory emphasizes the significant role of an innovation champion in promoting an innovation. Activity theory, in turn, considers learning, complementary knowledge, and actors’ expertise essential in the development of an innovation. We see that both political measures and grass-root level experiments are needed to develop a system innovation. Today, the role of researchers faces new challenges [13]. Researchers are becoming developers or facilitators, who construct new communities and dialogue across different actors and organizations during the innovation process. However, constructing a new kind of collective object and motive is not constructed overnight; it needs a systematic and planned intervention process.

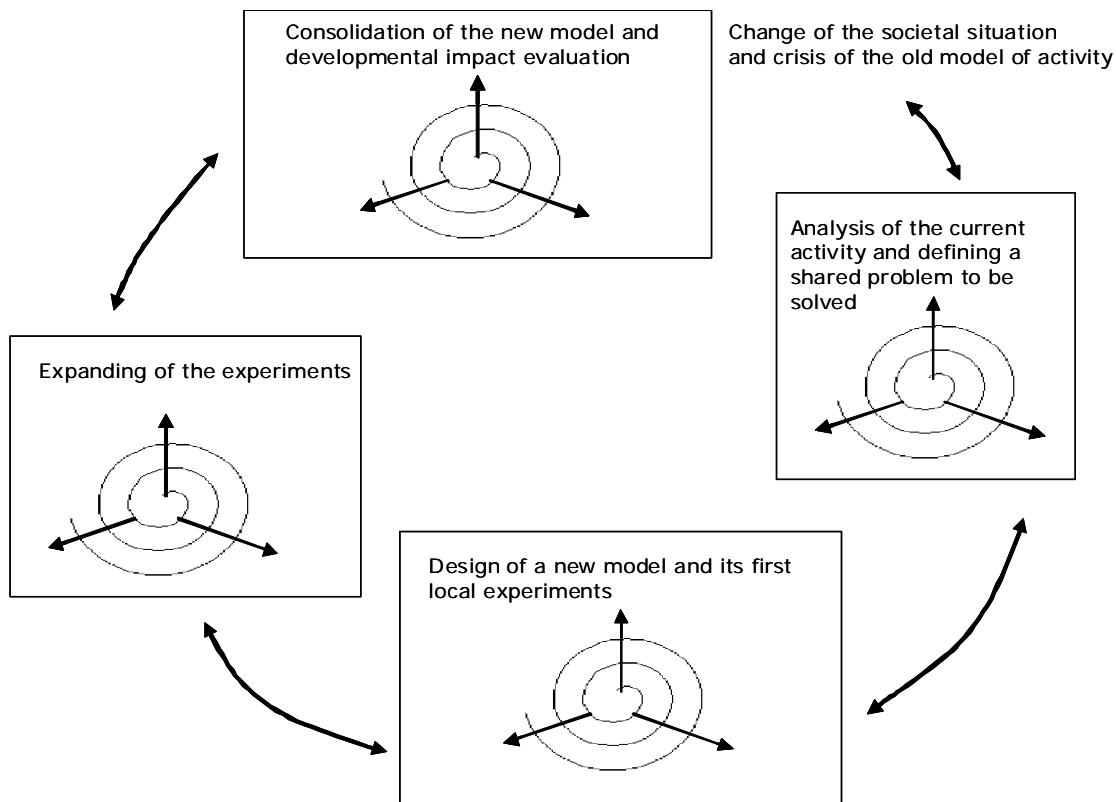


Figure 2. Societal embedding of innovation as a learning process.

Visual Representations of the Innovation as Tools for Reflecting and Developing

Visual and conceptual representations of the work processes have been stated to help the employees reflect on their work and make changes in it. In the multi-actor innovation network the researcher-developer has the expertise to construct new tools to materialize the new model of activity. The developer may make the motives of the different actors visible and make the development network realize the phase of the innovation. The participants need to have insight and make their own interpretations about the data, which has been collected e.g. from the prevailing problems and obstacles of the current model of activity. The developer provides the participants with both conceptual tools and the data in a learning occasion. In activity theory this is called a dual stimulation method [14].

The Evolution of the Agency in the Innovation Network

Innovation processes may take several years, even decades, from an idea to a feasible outcome. And the network and actors change along the path. Therefore, the researcher has an important task as a messenger and modeller of the innovation process, and thus serves the progress of the innovation.

In our projects, we have noticed that local level actors involved in the pilot phase of the innovation shift their positions, and become even entrepreneurs for spreading the innovation [15]. Consecutive workshops may generate “collective agency” for carrying the innovation embryo from one phase to another. The researcher-developer may also give a voice to those actors in the development process who do not usually have much power in prevailing working routines and models.

CONCLUSIONS

We have presented an overview of a method to enhance societal embedding of innovations, which emphasizes considering of societal values and sustainability of the innovation throughout the evolution of an innovation. We would like to see the researchers’ role as a more active interventionist in the future. However, the researcher-developer becomes not the owner of the innovation. Her role is more like a significant bridge-builder in the fragile and critical phases, when a development project ends and a new one is not established, yet. She is responsible for creating an optimal learning process to support the innovation path, and the network constructing it.

REFERENCES

[1] Gallouj, F. 2002. Interactional innovation. A neo-Schumpeterian model. In: J. Sundbo and L. Fuglsang (Eds.). *Innovation as strategic reflexivity*. London and New York: Routledge 2002.

[2] Sundbo, J. 2002. Innovation as a strategic process. In: J. Sundbo and L. Fuglsang (Eds.). *Innovation as strategic reflexivity*. London and New York: Routledge 2002.

[3] Engeström, Y. 2000. Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, 43, 7, 960–974.

[4] Kivisaari, S., Lovio, R. and Väyrynen, E. 2004. Managing experiments for transition. Examples of societal embedding in energy and health care sectors. In: B. Elzen, F. Geels and K. Green (Eds.). *System Innovation and the Transition to Sustainability: Theory, Evidence and Policy*. Cheltenham: Edward Elgar Publishing.

[5] Geels, F. W. 2002. Technological transitions as evolutionary reconfigurations processes. A multi-level perspective and a case study. *Research Policy*, 31(8/9), 1257–1274.

[6] Kemp, R., Rotmans, J. and Loorbach, D. 2007. Assessing the Dutch Energy Transition Policy: How Does it Deal with Dilemmas of Managing Transitions? *Journal of Environmental Policy and Planning*, 9, 315–334.

[7] Geels, F. W. 2004. Understanding system innovations: a critical literature review and a conceptual synthesis. In: B. Elzen, F. Geels and K. Green (Eds.). *System Innovation and the Transition to Sustainability: Theory, Evidence and Policy*. Cheltenham: Edward Elgar Publishing. Pp. 19–47.

[8] Weber, M. Hoogma, R., Lane, B. and Schot, J. 1999. *Experimenting with Sustainable Transport Innovations. A workbook for Strategic Niche Management*. University of Twente.

[9] Dewey, J. 1991/1938. *Logic. The theory of inquiry. The later works of John Dewey vol. 12* edited by Jo Ann Boydston. Carbondale & Edwardsville: Southern Illinois University Press.

[10] Saari, E., Kallio, K. and Hyytinen, K. 2008. Learning from the users of research: Defaces, insights and fragile endeavors as manifestations of expansive learning. Paper presented in *Organizational Learning and Capabilities (OLKC) Conference 28.4.–30.4.2008*. Copenhagen.

[11] Leontjev, A. N. 1978. *Activity, consciousness, and personality*. Englewoods Cliffs. NJ: Prentice Hall.

[12] Callon, M. 1992. The dynamics of techno-economic networks. In: R. Coombs, P. Saviotti and V. Walsh (Eds.). *Technological change and company strategies*. London: Academic press.

[13] Nowotny, H., Scott, P. and Gibbons, M. 2002. *Re-thinking science: Knowledge and the public in an age of uncertainty*. Malden, MA: Blackwell Publishers.

[14] Vygotsky, L. S. 1978. *Mind in society*. Cambridge, MA: Harvard University Press.

[15] Kivisaari, S. and Saari, E. 2009. Hybrid actors in developing system innovations – case of primary health care management. Paper presented in the *1st European Conference on Sustainability Transition: Dynamics & governance of transitions to sustainability*. Amsterdam 4th–6th June 2009.

Strategic Partnership as a Design Challenge – Applying Design Competence to Facilitate Innovation-Driven Relationships and Activities in Public Organization

Mervi Hasu
The Finnish Institute of
Occupational Health
Topeliuksenkatu 41 a A
00250 Helsinki, Finland
mervi.hasu@ttl.fi

Tuuli Mattelmäki
University of Art and Design
Helsinki, School of Design
Hämeentie 135 C 00560
Helsinki, Finland
tuuli.mattelmaki@taik.fi

Salu Ylirisku
University of Art and Design
Helsinki, School of Design
Hämeentie 135 C 00560
Helsinki, Finland
salu.ylirisku@taik.fi

ABSTRACT

In addition to the application of traditional expert competences, practitioners in service-intensive public organizations today must develop new skills for dealing with collaborative service concept development and various user-driven and customer-orientated participative work practices. This paper discusses an example case in which a design approach was applied to boost the innovation process in a knowledge-intensive public organization.

Keywords

public services, prototyping, design workshop, partnership, co-creation, collaborative learning, activity theory

ACM Classification Keywords

H.1.0 Information Systems, Models and Principles, General

INTRODUCTION

Public organizations today are faced with using design competence for public innovation. During the 2000s, the debate on innovations and innovation environments has expanded from technological, closed intra-organizational or controlled network environment towards social, service-orientated, user-driven, open environment innovation [e.g. 8]. The rise of service industries has also opened up discussion on the need for innovations in traditional public organizations, in which new hybrid forms of service production are increasingly taking place.

Strategic collaboration in R&D and the co-development of products and services with key customers are increasing. Advanced firms actively engage in strategic partnerships, i.e. strategic alliances [7], joint ventures [6] or regional networks [11] for various reasons, such as to acquire skills, to purchase or to obtain access to critical external resources, to gain benefits from another organization without owning it, to reduce risks, and to adapt to rapid technological or market changes.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Strategic partnerships offer potential to public sector organizations, in which partnering activity in general is a new phenomenon.

The public sector is seen as bureaucratic and reluctant to change, [4] which makes it a challenging environment for renewal and innovation. The image of public services is still very often hierarchical, slow, isolated, and customer unfriendly. Public services, organizations and their practitioners are being challenged to become innovative, i.e. open to everyday-life customer initiative and user experiences. This calls for a new type of expertise, expert identity, and attitude change among public sector practitioners. Isolated, abstract and theoretical knowledge is no longer enough. Instead, more network-orientated, collaborative, service-like and co-creative identities and competences are needed [5].

Thus a major transformation is taking place in many companies: instead of manufacturing and selling products to customers, the objective in service logic is “assisting customers in their own value-creation process” [16 pp 257]. Many innovative organizations today are advocating co-creation. Windsor [17] describes co-creation as a deep engagement with the internal team or engagement with customers and the culture in which they live. Successful co-operation and co-creation requires trust and engagement. Trust can be created through e.g. a process of negotiating common goals and values, realizing and organizing the network, communicating goals and identifying the roles of different players [9].

The building, nurturing and management of collaborative relationships are becoming an invaluable competence and a prerequisite for co-creation. But how does a traditional expert organization become a collaboration-intelligent community? For top management this does not yet seem to be a relevant question: strategy is the guide, middle management is the implementer. We propose that a designerly approach can support and drive both the strategic and everyday-life collaborative processes in organizations and networks. In this article, we aim at contributing to the development of new competences by discussing an example case in which a designerly approach was applied to boost innovation processes in a knowledge-intensive public organization.

CASE CONTEXT

We will briefly discuss a case in which design expertise was applied as a catalyst for exploring a phenomenon that was novel and unfamiliar to the organization. The Finnish Institute of Occupational Health (FIOH) is a public research institute which has recently undergone major organizational restructuring and renewed its strategy. It is a multidisciplinary organization that employs nearly 800 experts and has regional facilities

around the country. The target of FIOH's new strategy was to move towards the explicit interactive model of developing and implementing innovations through various modes of activity in close collaboration with partners and customers.

The new strategy created a major challenge for the development of expert competence and identity compared to traditional research expertise. It became a considerable learning challenge to both the management and personnel of FIOH. The management raised the question of how to organize, in practice, activities for innovation. How could they enhance and accelerate innovation within the new organizational structure? They came up with an idea of a kind of venture organization within FIOH. Two pilot venture units were established at the beginning of 2006, both of which already had considerable accumulated knowledge and permanent contacts with the main players in their sector. The units aim to make a major, novel contribution to the solution of particular needs or problems within the society, i.e. the solution is to be used by relevant societal partners and customers as part of their practices. The units can be interpreted as purposeful, time-pressured innovation pilots. That is, innovation can be deliberately enhanced and accelerated, at least partially, by managerial actions. The strategy plan included four broad phases: (1) planning (idea formulation), (2) start-up, (3) piloting and experimentation, and (4) customer-driven redesign and sustaining of the innovation.

The context of designer intervention, which we called the partnership mock-upping workshop, is associated with the Good Indoor Environment Quality venture unit. It is a multidisciplinary unit led by the Director (MD, professor). The group was comprised of two originally separate groups within FIOH and included 22 highly educated people (many of them PhDs), including several natural scientists and engineers. In order to accelerate innovation, a variety of expert competences were used for the benefit of the units. At the beginning of 2006, FIOH's Head of Research (first author of this paper) organized a small workshop in which external design experts (the other two authors) introduced methods and techniques for exploring and understanding user needs and user experience for product/service development. The Director was impressed, and later asked the same experts for help in planning how to approach one of the identified, most important (yet anticipated) partners of the unit. This organization, entitled here "the Properties", can be seen as one of FIOH's significant and strategic partners.

However, several questions arose. The potential common interest or practical target as well as the form of collaboration between the company and the venture unit were difficult to envision. How to get the firm interested in collaboration, how to present the unit's competences to the firm, how to open the negotiations and with what kind of ideas? How to engage the whole group in customer-orientated thinking and acting? It was decided together with the design experts that a workshop for exploring the matter was needed in order to provide a safe setting in which the group can experiment and learn the collaboration-building process together. The workshop took place during the start-up phase in autumn 2006. Its aim was to explore what the strategic partnership could consist of and what form it could take.

COLLABORATIVE PROTOTYPING

Kelley [10] characterizes prototyping as acting, exploring and perhaps even failing before finding the answers. Prototyping is applied for idea generation, communication and testing [14]. Mock-ups, i.e. low-fidelity prototypes enable exploring and sharing the form, scale and appearance of an idea. Mock-ups

are applied to consider particular features, to facilitate collaboration and to provide a hands-on feeling of the future product. The idea of building a strategic partnership, however, is a highly ambiguous whole when compared to traditional physical products. Kelley [10 p.36] advocates, "You can prototype just about anything – a new product or a service, or a special promotion. What counts is moving the ball forward, achieving some part of your goal." While prototypes concretize thoughts, and make them visible and debatable, they also foster playful exploration to get the feeling of things [13].

Early design phases are often characterised as 'fuzzy' and ambiguous. The challenging questions are: what are the concrete moves to be made, what do the activities mean to the team and to the organization, and what concrete results can these yield. Designing is about identifying alternatives which are discovered through exploring problems and solutions that are strongly intertwined. Similarly, the collaborative mock-upping of a complete process aims at discovering the elements of the process and helping to outline the actions to be taken. Thus, the objective of the workshop in this case was to give form to a process.

PARTNERSHIP WORKSHOP

The process mock-up workshop features three parts: contextualization, action, and reflection. Contextualization develops a shared understanding of why the workshop is organized, what the overall situation is, who and which organizations are involved, and what the aims are. During the action phase, workshop participants are encouraged to apply their knowledge, communicate, act, make quick decisions and produce a common understanding of the alternative solutions. The workshop activities are captured on video, which is used to facilitate reflection on the process and the decisions that were made. The experience and the video material support the team in planning the actual project.

The aim of the workshop organized at FIOH was to explore and develop strategic partnership. The workshop followed the idea of a user-focused collaborative prototyping of a process, which was piloted in concept design projects [1]. The planning of the workshop was built on the design experts' earlier experiences of user-centred product concept design projects, but was customized for this particular case in negotiations with the organizations' representatives.

The objective of the workshop was to make the first move from visionary words to a real life action plan. The organization and team members did not really have expertise in user- or customer-centred design mindsets or tools. The team realized, however, that it had to learn new strategies and practices in order to achieve a partnership with the key player. It had to be more than an expert institution; it had to be an attractive partner. The workshop's objective was thus to uncover what a partnership could be about, what the process of identifying and encountering the partner would be, how to maintain the partnership and furthermore, what shape the collaboration could take, what products, tools and methods would be applied. Team-building also needed support since the unit had only recently been established. The Head of Research and nine team members, including the Director, actively participated in the workshop.

The overall principles in the arrangements of the workshop were:

- 1) An authentic-like project organization is to be established, i.e. some participants were given specific roles such as Mr H, Project Manager and an evaluator (entitled financier)

- 2) Authentic-like deliverables must be created in every phase, i.e. abstract discussion has to be turned into actions, documents and solutions
- 3) Situations are to be explored by acting them out, and all roles should potentially be based on “real” characters, such as the development manager of the partner company
- 4) All activities are to be constrained by strict time limits in order to force intuitive action
- 5) Strategic decisions must be argued for in front of a critical ‘financier’, and a refined focus must be articulated during all reviews, i.e. exploration must be turned into solutions and the reviews must allow iterations.
- 6) Reflections are to be discussed with the help of the video documentary of the workshop and real plans are to be outlined based on these i.e. the reflective discussion enables the team to open up the experiences for analysis and iteration.

The structure of the workshop was roughly the following: 0) introduction 1) warm-up, 2) forming the project plan, 3) context study 4) review, 5) envisioning the future, 6) review of results, 7) reflection. The director explained the overall situation of the unit and the purpose of the workshop during the introduction. He also briefly described the approach that would be taken throughout the day and emphasized that they were all in the same situation, facing this novel challenge. Then the design experts explained the day’s agenda and the materials for the workshop (e.g. hats for role-playing). These pre-warm-up explanations aimed to create a motivating context and to positively affect the participants’ expectations of the workshop.

The video documentary of the workshop day was reviewed the next day. It created a vivid basis for a discussion on insights and potential ideas for the actual process. Immediate feedback revealed that the day had served its purpose well. The participants pointed out that through the process they had gained a clearer picture of the potential partnership, and considered this necessary for their progress. They also developed initial experiences regarding the possible tools and methods that might be employed during the next phases. Some of the methods that were tried out at the workshop, such as visits to the partner’s environment, interviews of the relevant actors, and observations of the work, could be implemented immediately.

THE NEXT STEP

In 2007, FIOH and the Properties launched a development project in order to assess how health and safety aspects could be more effectively integrated into real estate management. The Properties has even defined indoor environment as one of their most important targets for development. The development phase began in 2008 and was implemented by using participative workshops. The co-operation between FIOH and the Properties in this project will continue until at least 2010. Although the experimented relationship has now been realized in practice, the realistic future level of partnership, as well as the future of the unit, remains to be seen.

The workshop built the participants’ confidence as regards working with the anticipated partner organization. It was, however, only the first mock up of the partnership. Perhaps the most obvious evidence of the value of the workshop is the fact that the unit has expressed interest in organizing a new workshop to tackle the situation that they are currently facing. They now have a good start with the partner but are hesitant about how to move beyond the current level of collaboration. In the forthcoming new workshop, the participants would like

to create a new process mock-up to collectively evaluate the experiences of the current collaborative project and to invite the actual partner organization’s representatives to co-explore potential ways in which to proceed.

DISCUSSION

For designers, prototyping serves as a framework for the application of design competence in organizational change, which is still a rather unfamiliar arena for them. Designers are skilled in moving flexibly from one topic to another and applying knowledge, tools, theories and ideas from various fields of their work. However, it was realized that it is extremely useful to know and be able to communicate the reasons for utilizing the design approaches and processes in order to convince and motivate the stakeholders, and moreover, to translate the process and methods for the novel usages. In addition, the designers need to be sensitive in identifying and interpreting novel phenomena outside of their previous expertise. For example, the prototyping material consisted of human interactions, not plywood or foam, and one of the design components that had to be discussed was body language in role-playing.

We suggest that in addition to the application of traditional design competences, designers need to develop new skills for dealing with social practices, intangible processes, and complex systems.

For the participants, prototyping serves as a collective learning activity [15] for the building of new expert competence and identity in organizational change. In the workshop, the experts faced an unfamiliar situation. They had to leave their analytical, individually-bounded expert identity and “civilized” meeting room behaviour and throw themselves into a collective role-playing and exploration mode. They even had to reveal that they didn’t always understand what they were supposed to do. They needed to learn something that did not yet exist, in activity-theoretical terms; to move in the zone of proximal development [3] of the new expertise. Design process and the mode of ‘fuzzy’ exploration can feel chaotic for someone unfamiliar with design, as Kelley has also noted [10]. Based on the team’s feedback, facing the confusion and surviving with insightful results was rewarding. Going through this process fostered “out of the box thinking”, and being open to new opportunities and analogies.

The role-playing, and the process of producing deliverables at a fast pace were part of the ‘fuzzy’ exploration. Despite some confusion, the experts adapted to their roles easily and the whole workshop in fact proceeded as a play in which the participants improvised their lines by stating their status e.g. “From the perspective of the Properties’, I would like to point out ...” This attitude was already created at the beginning by the Head of Research acting as the financier, and the two “representatives” of the Properties. Success was also partly because the key persons played along and inspired others to join the game. Most importantly, the Director learned as part of the team, as a genuine member and not as somebody who knows best.

The workshop was part of a series of events meant to encourage innovative activity. The top management expected remarkable results from the venture unit. The participants were motivated to try out the design process. Since they were highly educated experts, they were also able to quickly observe the key elements, make outlines and translate some of the design-related assignments into a language that was closer to their own field of competence. Although we do not expect that such expertise is always needed, personal motivation and motivating

the team in different ways is necessary. The Director and the Head of Research played crucial roles, both of them highly engaged in the workshop activities and, moreover, in the unit's overall goal of achieving successful results.

Product design mock-ups are tangible. In this exercise the concreteness was achieved by the casting of roles, the seeking for meaningful situations, functionalities, and human to human interaction in everyday life. In addition, the team was guided to concretize, make sense and communicate through visualizations and acting out situations. During the exercises, the team learned to design the process together, "got the feeling" of the context, and through being engaged in different roles, exercised a human-centred approach that is extremely valuable in building trust and partnerships. Finally, the video documentary of the day serves as a reminder of the collaboratively created mock-up of the partnership process.

Based on our experiences, we suggest that the design approach helps represent complex collaborative processes and strategic partnership-building as concrete practice. It encourages management to perceive collaboration as a multi-level activity which is performed by real-life people and groups at all levels of organizations and networks [15]. It can be used intentionally to help participants nurture partnership as a temporal, stepwise activity in which mutual learning needs to take place. Learning may concern the co-creation of a target and mission of partnership; its means, tools, rules, and working patterns, as well as emerging future forms of collaboration.

REFERENCES

- [1] Battarbee, K. et al. 2005. Looking beyond the product: design research in industrial and academic collaboration. Proceedings of Joining Forces Conference. UIAH, Helsinki.
- [2] Burns, C., Cottam, H., Vanstone, C. and Winhall, J. 2006. RED PAPER 02: Transformation design. Design council. UK. [Online PDF] Available at www.designcouncil.info/mt/RED/transformationdesign/ [accessed 10 July 2009].
- [3] Engeström, Y. 1999. Innovative learning in work teams: Analyzing cycles of knowledge creation in practice. In: Y. Engeström, R. Miettinen and R.-L. Punamäki (Eds.). *Perspectives on Activity Theory*. Cambridge: Cambridge University Press. Pp. 377–404.
- [4] Fuglsang, L. 2008. Capturing the Benefits of Open innovation in Public Innovation: a Case Study. *Int.J. Services Technology and Management*, Vol. 9, Nos. 3/4.
- [5] Fuglsang, L. and Sundbo, J. 2005. The Organizational Innovation System: Three Modes. *Journal of Change Management*, Vol. 5, No. 3, 329–344.
- [6] Grandori, A. and Soda, G. 1995. Inter-firm Networks: Antecedents, Mechanisms and Forms. *Organization Studies*, 16(2), 183–214.
- [7] Håkansson, H. and Sharma, D. D. 1996. Strategic Alliances in a Network Perspective. In: D. Iacobucci (Ed.), *Networks in Marketing*. London: Sage Publications. Pp. 108–124.
- [8] von Hippel, E. 2006. *Democratizing Innovation*. Cambridge: The MIT Press.
- [9] Järvensivu, T. and Nykänen, K. 2008. Identifying basic elements of network management: Comparison between managing networks, Markets and Hierarchies. In: Anttila, M. and Rajala, A. (Eds.). *Fishing with business nets: keeping thoughts on the horizon* Professor Kristian Möller. HSE print. [Online publication] Available at hsepubl.lib.hse.fi/FI/publ/hse/b90 [accessed 10 July 2009]. Pp. 146–161
- [10] Kelley, T. 2001. Prototyping is the Shorthand of Design. *Design Management Journal*, 12(3), 35–42.
- [11] Perrow, C. 1992. Small-firm Networks. In: N. Nohria and R. G. Eccles (Eds.). *Networks and Organizations; Structure, Form and Action*. Boston, Massachusetts: Harvard Business School Press. Pp. 445–470.
- [12] Schlosser, A. E. 2003. Experiencing Products in the Virtual World: The Role of Goal and Imager in Influencing Attitudes versus Purchase Intentions. *Journal of Consumer Research*, 30(2), 184–198.
- [13] Schön, D. A. 1983. *The reflective practitioner*. Basic Books. New York, NY.
- [14] Säde, S. 2001. *Cardboard mock-ups and conversations*. Doctoral thesis. UIAH Helsinki. Finland.
- [15] Toiviainen, H. 2003. *Learning Across Levels. Challenges of collaboration in a small-firm network*. Helsinki: Helsinki University Press.
- [16] Vargo, S. L. and Lusch, R. F. 2008. From Goods to service(s): Divergences and convergences of logics. *Industrial Marketing Management*, 37, 254–259.
- [17] Winsor, J. 2006. *SPARK. Be more innovative through co-creation*. Dearborn Trade Publishing, US.

Representing Changes in Work for Design: An Activity-theoretical Perspective

Laura Seppänen

Finnish Institute of Occupational Health
Topeliuksenkatu 41 a A1
00250 Helsinki, Finland
laura.seppanen@ttl.fi

Arja Ala-Laurinaho

Finnish Institute of Occupational Health
Topeliuksenkatu 41 a A1
00250 Helsinki, Finland
arja.ala-laurinaho@ttl.fi

ABSTRACT

Representing long-term qualitative changes in work activities and practices is increasingly important for design. By drawing on activity theoretical understanding on representation and models, this paper shows how a model representing activity-level changes of work was designed and used in a developmental intervention. Features of developmental models to represent changes in work are discussed. The case example aims at contributing ideas for ecological system design in depicting functions of historicity in the models for developing human practices.

Keywords

models, representations, changes in work, developmental interventions, design, activity theory

INTRODUCTION

As work becomes more complex, and as different tasks and responsibilities are increasingly distributed and recombined, the representations of work and its changes become crucially important for design and managing work practices.

Changes in work both urge the need for design, but they also reveal the nature of work activities. Launis and Pihlaja [8] describe different types of changes in work, as experienced in five different work organizations. One type of change is a sneaking one - a bunch of gradual changes affects substantially the everyday work but these changes are not consciously and collectively examined, and thus work is still carried out with old practices. Another change type is the implementation of a new information system which affects the whole work organisation much more than anticipated. Interestingly, the authors find that almost all types of changes, as experienced by individuals, were connected to large-scale, fundamental activity-level changes going on at work.

To start with, we describe two classifications of representation from the activity-point of view. The aim of our paper is, first, to give an example of how a representation of changes in work was designed and used in a work setting of collaborative local food trade. Design is thus on two levels: first on designing a developmental model or tool, and then on designing work

patterns and practices through the use of the tool. Second, the aim is to discuss the developmental nature of this model and the possible functions of historicity (changes in time) in developmental design. At the end, we figure out features for models representing changes in work for developmental design purposes. We hope to contribute to the concept of ecological system design that investigates the multiple human-environment relations focusing especially on human practices [6].

CLASSIFICATIONS OF REPRESENTATIONS

The tools of an activity give a societal meaning and structure to individual perception and experience. Marx Wartofsky [13] suggests a three-level of hierarchy of artifacts used as tools in human activities. *Primary artifacts* are used on the operational level where the subject is mostly unaware of the means he or she is using. *Secondary artifacts* are internal or external representations which we activate when we think on the nature and use of tools – sort of reflexive embodiments of actions where primary tools are used. Blueprints and manuals of software are examples of secondary artifacts. Further, Yrjö Engeström [4] divides secondary artifacts in two types. *How-artifacts* are the ones that tell us how a certain object shall be handled with a corresponding artifact: they can be rules directly guiding the use and formation of primary artifacts. A procedure of an ergonomic analysis can be a how-artifact. *Why-artifacts*, in turn, are more general, giving us an idea of why an object behaves as it does, and thus justifies the selection of primary artifacts. They can be eg. explanatory models of cognitive or ergonomic phenomena.

Wartofsky [13] proposes still another category of *tertiary artifacts*. They give overarching perspective to activity formations, but still are more autonomous and more detached from the practical activity. Socio-political visions, scientific paradigms and religious creeds are examples of tertiary artifacts [4].

Representations of changes in work can be related to a classification of *work orientations*, which mean collective and individual models or patterns of thinking and which steer and regulate work. Engeström [3] distinguishes five types of orientation models:

1. *Spontaneous models* appear as experiential or figurative prototypes, eg. case examples of a phenomenon. It is often difficult for a person to consciously analyze spontaneous models.
2. *Anticipatory analyzers* are static models in which an object or a phenomenon has been divided into parts or categorized, listed or formed a hierarchy of it by some common and distinctive features.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

3. *Algorithms* are process descriptions or prescriptions for the performance of duties that proceed in phases. They do depict movement in time, but are unable to consider interactions and feedback.
4. *System representations* are models delineating feedbacks, interactions and changes taking place in the whole system. Often they turn to be complex and balanced descriptions of phenomena.
5. *Germ cell models* aim at portraying simple internal relations within a system as contradictory ones. The contradictory nature of these relations makes a need for development and change (development meaning a qualitative change in the work system).

The germ cell model is the most developmental type of orientation models. Showing the contradictions of a work system creates a tension that has importance from the point of learning and development. Besides activity theory, also other learning theories emphasize discrepancies or tensions in promoting learning [1, 9].

A representation basically means “a picture or image” of an object or a thing. Wartofsky [13] links representations, through perception, as part of human doing and making (communication and production). Following Knuuttila [7], we rather speak about models, in order to emphasize their use-character in two respects: how they are produced, and how they are used in developing work activities. Later, we will evaluate the model in the light of these classifications.

In the following, we first present our case and its intervention. Then, we describe, how the model was designed and used, and how the developmental effort continued after its use.

THE CASE OF LOCAL FOOD

Introduction to the case. The gap between small and craft-type of food enterprises and their large-scale and increasingly centralizing buyers is a problem in the fluency when trading the local food products. Therefore, an intervention study for promoting the collaboration and well-being between producers and buyers of local food was carried out in 2006-2007. Our case example of the study was a small wholesale company owned by three farmer families as partners, whose main idea was to deliver local food products to various buyers, mainly to big retailers, but also elsewhere. The company supplied food products of about ten other small food enterprises as well.

Before founding the company, the partners were individual farmers producing and trading their own local food products, having only occasional collaboration with each other. After the company was founded, the company marketed the products of the partners with their initial, enterprise-bound trademarks. Later in 2006, the company bought an existing brand, the trade of which was limited to two biggest Finnish retail chains. Together with this new brand, the company got many new suppliers and strengthened its trade with the big retail chains. The company had to invest in a new ordering software. Simultaneously with the trade with the new brand, the company was still trading the original trademarks of the partners. Moreover, each partner still had also individual trade, outside the company, of the products of their farm enterprises.

The intervention. The intervention was based on activity theoretical ideas of Developmental Work Research [5, 12] in which tensions or contradictions within activity systems, especially within their motivating objects give rise to learning challenges [11]. Roughly, the intervention consisted of four

phases. In the first phase, the researchers interviewed and followed the work of the case for preparing suitable data and methods to be used in the intervention. The second phase was the joint intervention meetings, the actors and researchers together, to discuss the history, state-of-the-art and future of the company, and to design a developmental experiment. The third phase was carrying out the experiment, and the fourth phase was the evaluation of the intervention and the experiment. The model we present below was created in the first phase and used in the second one. Because of the limitation in the length of this paper, we cannot explain here all the intervention methods and tools, which can be found elsewhere ([2], for this case; [5], in general).

THE DESIGN AND USE OF THE MODEL

How the model was designed. How the changes that had happened in the company should be discussed, and what type of model could be designed for promoting fruitful dialogue about the organizational design? It was first necessary to understand the general work of a wholesale activity. A reference from marketing literature was a list of important central operations: ordering, deliveries, sales promotion, product development, and networking. This list was a basis for structuring the work of the company as well as for designing the intervention methods and the model.

The interviews and the ethnographic work of the first phase helped understand the history and the main challenges experienced by the suppliers, buyers and the company partners.

Previously, the company had supplied food products mainly to regional buyers or to some other specialized small wholesalers or retailers, by using trademarks of each of the supplying producer. In 2006, the company had bought a new, national brand by which it obtained multiple new suppliers and also strengthened its trade with two main big retailers. While the company was carrying out simultaneously both old regional and new national forms of trade, the question of sales promotion became acute: should the producers promote their own trademarks as part of the company, or should efforts be put into joint promotion of the new national brand? Another challenge for company's internal collaboration was the ordering. Actually, the ordering is partly done by the company, but also individual partners do ordering, to their own customers, as well. These two challenges could be crystallized in a four-quadrant framework (Figure 1). The theoretical idea of the tensions in the object of activity guided the understanding of these challenges and the making of the model.



Figure 1. A model of the developmental transition going on in the company.

How the model was used. The model (Figure 1) was used in an intervention meeting together with the partners of the company. For its introduction, the history of the company was first discussed, as well as customers' and suppliers' expectations towards it, and the actual state-of-the-art of the company's activities. Also, the central operations that the company should take care of were discussed (ordering, delivery, sales promotion, product development and networking). For being able to have trade with the big retailers, the company had made relatively big investments in ordering technologies, which would allow and need much bigger sales volumes than the actual volumes. Then, the model (Figure 1) was shown to participants. Where is company now, and where it wants to go? These questions were posed together with the model. As a result, the actual (A) as well as the desired situation (B), as seen by the company partners, were placed in the framework (see Figure 1). The newly bought ordering software was crucial in having moved horizontally from left to right in the model. Sales promotion was desired to be the most important operation of the company.

How the intervention continued: The possibilities for enhanced sales promotion were discussed in the following meetings. Later on, an experiment of marketing a new product was carried out, which also produced guidelines of how initial sales promotion for big retailers should be made in the company.

THE DEVELOPMENTAL NATURE OF THIS MODEL

The axes of the model (Figure 1) suggest that the developmental tensions of the company were: 1. to distribute or to join the ordering functions, and 2. to have common sales promotion with the newly bought joint brand, or to put sales promotion efforts to individual products and their brands. In the classification by Engeström [4], the model (Figure 1) can be considered as an initial germ cell model, because it shows a hypothetical basic contradictory relation of the evolving wholesale trade activity. In the model, there is tension within one dimension as well as between different combinations of the two dimensions. The use of the model showed the current situation (A. common ordering, sales promotion distributed between products and individual partners). Also, the future wanted situation was found and discussed (B. Joint activity with the common brand, Figure 1). The model helped understand the gap between the actual pattern of work and the new envisioned form of work activity.

Within the hierarchy of artifacts [13], this model, based on its use in the intervention, is a secondary artifact. It was used both as a how-artifact [4], for designing future experiments, and as a why-artifact for overall reflection of the development of the work. In the latter sense, it could have been used as a reflective framework for depicting different orientations [3] the company and its partners can have in managing their trade. The actual and the future desired orientations are depicted as A and B in Figure 1. But, if we consider even an older historical phase, we can depict an orientation where farm-enterprises marketed the products and did the ordering themselves, independently of the company. Remarkably, the partners of the company still continue with the individual trade of their brands and products – this third oldest orientation clearly affecting the company is still present (C individual ordering and sales promotion, in Figure 1). The model does not only show these orientations, but give an idea or a hypothesis about their transformation in time.

The basic contradictory relation, as embedded in the model (Figure 1), was thus between individual-enterprise tasks, and joint ordering and sales promotion efforts of the company. Depicting and discussing all three orientations, including the oldest, still existing orientation would have been useful to understand better the basic contradictory relation and developmental challenges of the company. Questioning the old and new orientations the partners had in their every day work – whether to work individually to benefit their own enterprise, or to support the joint company – would have perhaps been difficult for the company partners, but it would have given broader perspectives towards the role and meaning of the company. In general terms, the historical patterns and orientations serve also as a mediator for understanding the linkages between tasks of individuals and the changes in collective work.

The model (Figure 1) representing changes in work can be taken as linear guide to a given, unidirectional goal, or a framework showing various alternatives. The fact that different orientations maintaining certain independence exist in the model reveals its open character: it is possible for the actors to go what ever direction, even to the old orientations. Still, the overall frame gives a hypothesis of the wanted direction. Activity-level developmental models of change may, of course, take also other forms of visual representation than the four-quadrant frame in our example.

DISCUSSION AND CONCLUSION

Based on this case, we want to draw some general features for creating and using developmental models about changes in work. They necessarily require that the models representing changes are constructed in collaboration between practitioners and experts (designers and researchers).

As the use of the model aims at enhancing a useful dialogue with actors, the models need to be close to practical contents of work and to practitioners' existing orientations towards it. In our case, both the empirical study (eg interviews and ethnography) and the literature reference (list of operations) were necessary to create a practically oriented model depicting major changes in work of the company.

To push forward the concrete developmental efforts, we need to make a link between these models and the tasks of individual participants. It is often in the individuals' tasks and experiments where creative agency for collective developing of work should be enhanced [12]. And, in order to enhance participants' agency in developing their work, the models need to be open to various alternatives. An influential way of connecting these models to tasks was the experiment carried out during the intervention. The preferable tasks or individual actions to be developed are not directly seen or derived from the models, but the model may offer a structure that helps in selecting relevant experiments. The model may also be of help in assessing the realization of the experiments.

An analysis of the changes in time is a basis for understanding the actual state-of-the-art [10] as well as for finding future developmental alternatives. We think that efficient developmental models representing changes need to include a tension to motivate questioning and reflection. In our case, changes (history) in work were used to create this tension. *The crucial developmental point was to show how the old and the new were both present in the actual activities.* Depicting tensions are useful for collective developmental design, but revealing the historical origins of the tensions of the work system may better help in managing changes and open insights for developmental organizational design. Further investigation about this is needed.

To conclude, the history, or depicting the path of changes, has many functions such as managing changes and designing for the future, creating a tension for developmental efforts and for connecting the work of individuals to the collective activities.

References

- [1] Argyris, C. and D. Schön. 199. Organizational learning II. Theory, Method and Practice. Menlo park: Addison-Wesley.
- [2] Arola, J. and Seppänen, L.. 2007. Kumppanusriihi: pk-tuottajien ja ostajien elintarvikeyhteistyön kehittäminen. Unpublished report.
- [3] Engeström, Y. 1985. Kehittävän työntutkimuksen peruskäsitteitä. *Aikuiskasvatus*, 4, 156–164.
- [4] Engeström, Y. 1990. When is a tool? Multiple meanings of artifacts in human activity. *Learning, working and imagining. Twelve studies in activity theory*. Orienta-Konstultit Oy. Pp. 171–195.
- [5] Engeström, Y. 2007. Putting Vygotsky to work: the Change Laboratory as an application of Double Stimulation. *The Cambridge Companion to Vygotsky*. H. Daniels, M. Cole and J. W. Wertsch, Cambridge University Press. Pp. 363–382.
- [6] Kaasinen, E. and L. Norros 2007. Älykkäiden ympäristöjen suunnittelu: kohti ekologista systeemiajattelua. *Teknologia-teollisuus*, julkaisuja nro 6/2007.
- [7] Knuuttila, T. 2005. Models as epistemic artefacts: Toward a non-representationalist account of scientific representation. Dept of Social and Moral Philosophy. University of Helsinki.
- [8] Launis, K., Pihlaja, J. et al. 2005. Tuotantotapamuutokset ja työhyvinvointi. Henkilöstö työpaikan muutoksen tekijänä. (Changes in ways of production and work-related well-being. Personnel as an agent of change, in Finnish) Hankeraportti Työsuojelurahastolle. Helsinki, Helsingin yliopisto, Toiminnan teorian ja kehittävän työntutkimuksen yksikkö.
- [9] Normann, R. 2001. Reframing business. When the map changes the landscape. John Wiley & Sons, Ltd.
- [10] Querol, M. A. and L. Seppänen. 2009. Learning as changes in activity systems: the emergence of on-farm biogas production for carbon credits. *Outlook on Agriculture* 38(2), 147–155.
- [11] Seppänen, L. 2004. Learning challenges in organic vegetable farming. An activity theoretical study of on-farm practices. Doctoral dissertation, University of Helsinki.
- [12] Virkkunen, J. 2006. Dilemmas in building shared transformative agency. *@ctivités*, 3(1), 44–66. <http://www.activites.org/v3n1/virkkunen-en.pdf>
- [13] Wartofsky, M. 1979. Models: Representation and the scientific understanding. D. Reidel Publishing Company.

Session 10: HSI Design for Collaborative Work

Being Virtually Everywhere: An Exploration of Teachers' Multitasking in a Hybrid Ecology of Collaboration

Rikard Harr, Victor Kaptelinin
Department of informatics
Umeå University
Sweden

rharr@informatik.umu.se, vklinin@informatik.umu.se

ABSTRACT

Collaboration mediated by digital technologies is typically considered an alternative to face-to-face collaboration. However, in real-life settings “virtual” and “physical” collaboration are often complementary, rather than mutually exclusive. This paper reports an empirical study of a hybrid physical/virtual ecology of collaboration at a senior high school in Sweden, a massively collaborative environment featuring different concurrently used groupware. The study focused on teachers' *collaborative multitasking*, that is, management of multiple collaborative activities. The findings indicate that the use of groupware in the setting presented a significant challenge for the teachers, who experienced *collaboration overload*. To keep themselves updated on current developments in their teams and projects, the teachers developed a variety of strategies for monitoring several collaboration spaces and switching between different technologies. The identified problems and strategies of collaborative physical/virtual multitasking are discussed in relation to existing research and design of supportive technology.

Keywords

multitasking, groupware, information overload, interaction overload, physical/virtual collaboration

ACM Classification Keywords

K.4.3 [Computers and Society] Organizational Impacts – Computer-supported collaborative work..

INTRODUCTION

A characteristic feature of the modern way of working, as well as life in general, is multitasking. Increasingly, people have no choice but constantly juggle several tasks, roles, and commitments. In human-computer interaction (HCI) and computer-supported collaborative work (CSCW) multitasking has been attracting significant attention and has become a key object of study [1,

2, 3]. With some notable exceptions [4, 5, 14], the main focus of multitasking research has been on how people manage their individual tasks (e.g. [2, 3]). This perspective is undoubtedly important, but not sufficient for understanding multitasking as a real-life phenomenon.

Work in modern organizations is increasingly performed by flexible configurations of collaborative units, such as projects, groups, or teams. This trend, combined with – and fueled by – the widespread use of information technologies, has transformed a wide range of work practices and has contributed to a proliferation of work contexts where individuals are commonly involved in several projects running in parallel [3, 6]. Multitasking in modern work settings is, essentially, collaborative multitasking, i.e. managing several threads of collaborative activities.

Collaborative multitasking is associated with a number of specific problems and challenges [14]. When an individual takes part in several collaborative projects, the amount of information and interaction one has to deal with may become massive. Richness of interaction and information sharing within an organization, which, on the one hand, is known to facilitate work processes [7, 8], can, on the other hand, cause ‘information overload’ [9] and ‘interaction overload’ [10].

To better understand challenges and solutions of the technological support for collaborative multitasking, the scope of research should be expanded from individual technologies to whole real-life contexts, in which multitasking is taking place. There is a growing trend in HCI and CSCW to analyze the use of technologies as contextualized in ecologies of people and artifacts [11, 12], including hybrid physical-virtual ecologies [13]. This paper aims to contribute to this research by reporting a study of collaboration among teachers at a senior high school in Northern Sweden. The setting featured a variety of groupware used in addition to traditional collaboration support and, therefore, can be described as a hybrid physical/virtual ecology of collaboration. The focus of the study was on the main challenges faced by the teachers when managing their multiple collaborative activities, the strategies employed by the teachers to achieve productive work, and how information technologies were used in the setting.

The rest of the paper is organized as follows. In the next section we give a brief overview of selected relevant work on multitasking and information overload. In the two sections that follow we describe, respectively, the method and the findings of our empirical study. Finally, in the last section we present the overall conclusions and suggestions for further research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

MULTITASKING AND INFORMATION OVERLOAD

Multitasking

The term “multitasking” generally refers to situations, where several tasks are performed in parallel or immediate seriatim [17] and significant attention has been assigned to multitasking in previous research [3, 14, 15]. Some studies indicate that multitasking is associated with positive outcomes, such as increased productivity [16], while other studies suggest that multitasking causes less positive effects, such as reduced performance and increased level of stress [15, 17].

Traditionally, studies of multitasking in organizations focused upon managers i.e. how managers arrange their working hours and spend their time [18, 19, 20]. These studies provide compelling evidence that managers are indeed continuously involved in numerous concurrent activities. Sproull [18], for example, shows that managers spend 80% of their time in different kinds of interactions and that their work is consequently best described as multitask processing. Hudson et al. [19] found that managers are constantly looking for uninterrupted periods of time but at the same time acknowledge interruptions and multitasking as a natural part of their work. A study conducted back in the 1970s [20] characterizes the work of CEOs in terms of “brevity, variety and fragmentation”.

More recently, multitasking research has also included other categories of workers. González and Mark [3], for instance, studied multitasking in various groups of information workers, including not only managers but also analysts, and software developers. Their study confirms that information work in general is very fragmented and that respondents: (a) normally spent three minutes of working on an event before moving on to work on another one, (b) spent an average of two minutes using an electronic tool, paper document or application before switching to another tool, and that (c) an average of eleven and a half minutes was spent on continuous work on a theme or a project before switching to another theme or project.

Information Overload

Projects, themes, or tasks that information workers face in their everyday work are usually associated with vast amount of information resources, such as reports, advice, or proposals. Participation in numerous parallel activities thus has the potential of creating a situation, where the individual is overloaded with information and interaction. A term used for describing this situation is “information overload” which, according to Eppler and Mengis [21] is often used to express the notion of receiving more information than one can handle. Other related terms have been used for describing similar experiences of not being able to cope with a situation, such as cognitive overload, communication overload etc. Inspired by the range of definitions that are used within existing research, Mulder et al. [22] suggest the following definition (also adopted in this paper): “*Information overload is the feeling of stress when the information load goes beyond the processing capacity*” [22, p. 245]. The negative effects of experiencing information overload are described as feelings of stress [6, 21, 23], anxiety [9, 21], and reduced or threatened productivity [9]. LaPlante [24] reported in the Computerworld magazine on a survey study of managers in the U.K, the U.S., Singapore and Hong Kong. The study found that 25% of the respondents experienced negative health effects (e.g., headaches and depression) due to the vast amounts of information they had to manage in their work. Even worse, 94% of the respondents

reported little hope for improvements in the future. There are even some reports (e.g., [25]) that information overload may have a persistent psychological effect by causing the development of the attention deficit trait (ADT). According to Mulder et al. [22], individuals who develop this trait find it difficult to structure their work, manage time, set priorities, and are constantly feeling a moderate level of guilt and panic.

EMPIRICAL STUDY

Research Site

The senior high school in which we conducted our study is located in Northern Sweden; in 2008 it had around 1100 pupils and employed approximately 140 teachers. Every teacher belonged to one workgroup that included all teachers involved in a certain educational program. In addition to belonging to a workgroup, each teacher was also a member of other constellations such as a workplace issues group, subject groups (which included all teachers teaching a particular subject), or less formal units, such as a marketing group or an anti-bullying group.

Method

Informed by ethnography [28] the study employed observations, interviews, and document analysis as data collection techniques. The study started out with a three-day observation of a teacher, followed by interviews and analysis of several documents. The interviews were conducted with seven teachers and two technical support persons (TSPs) (see Table 1).

Table 1. Background data of respondents (Y.O.P: years of practice).

Name	Role	Age	Gender	Y.O.P.	IT- experience
Robin	TSP	-	M	-	-
Kim	TSP	-	M	-	-
Fredrik	Teacher	35–45	M	0–10	Above average
Nils	Teacher	45–55	M	10–20	Above average
Martin	Teacher	45–55	M	10–20	Average
Sara	Teacher	55–65	F	10–20	Above average
Eva	Teacher	55–65	F	30–40	Average
Elisabet	Teacher	45–55	F	10–20	Above average
Jonas	Teacher	25–35	M	0–10	Above average

The TSPs were involved in the process of implementing a new learning management system (LMS), Fronter, at the school. They also provided us with several documents describing administration work and the use of IT at the school.

All interviews were conducted at the respondents’ workplaces and lasted between 45 to 90 minutes. The teachers were interviewed individually, while the TSPs were interviewed together. The teachers were asked about their everyday work in general with a specific focus on collaborative processes and the use of IT, while interviews with the TSPs focused on previous and current IT-support for collaboration at the school. All interviews were recorded and later transcribed by one of the authors.

In order to preserve the integrity of the respondents, their real names are substituted in Table 1 with fictional names and information that could reveal their identity have been blurred in

the figures. In addition, information about age and years of practice is presented in intervals to further protect respondents' identities. All teachers were given the opportunity to review the transcripts of their interviews before the analysis phase was initiated.

Technology Use at the School

The main systems used by the teachers were FirstClass (or FC) (Figure 1), an Intranet (Figure 2), and Fronter (Figure 3).

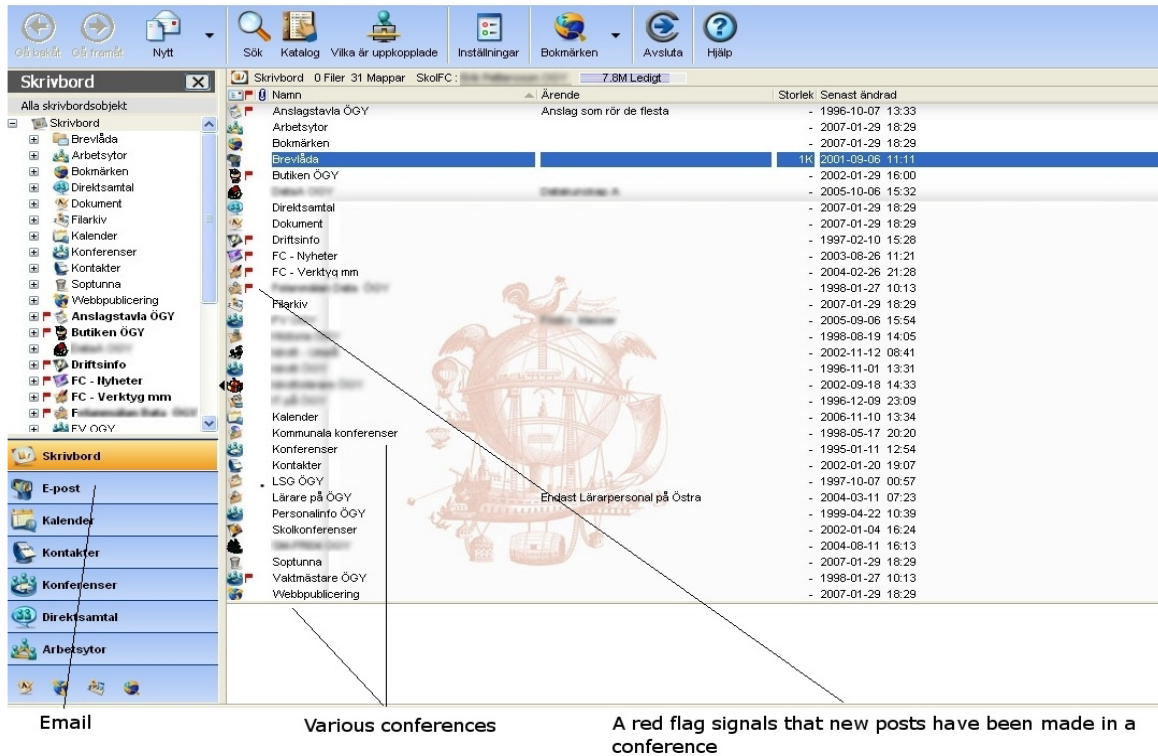


Figure 1. Screenshot of a specific teacher's FC interface.



Figure 2. A screenshot of the school Intranet.

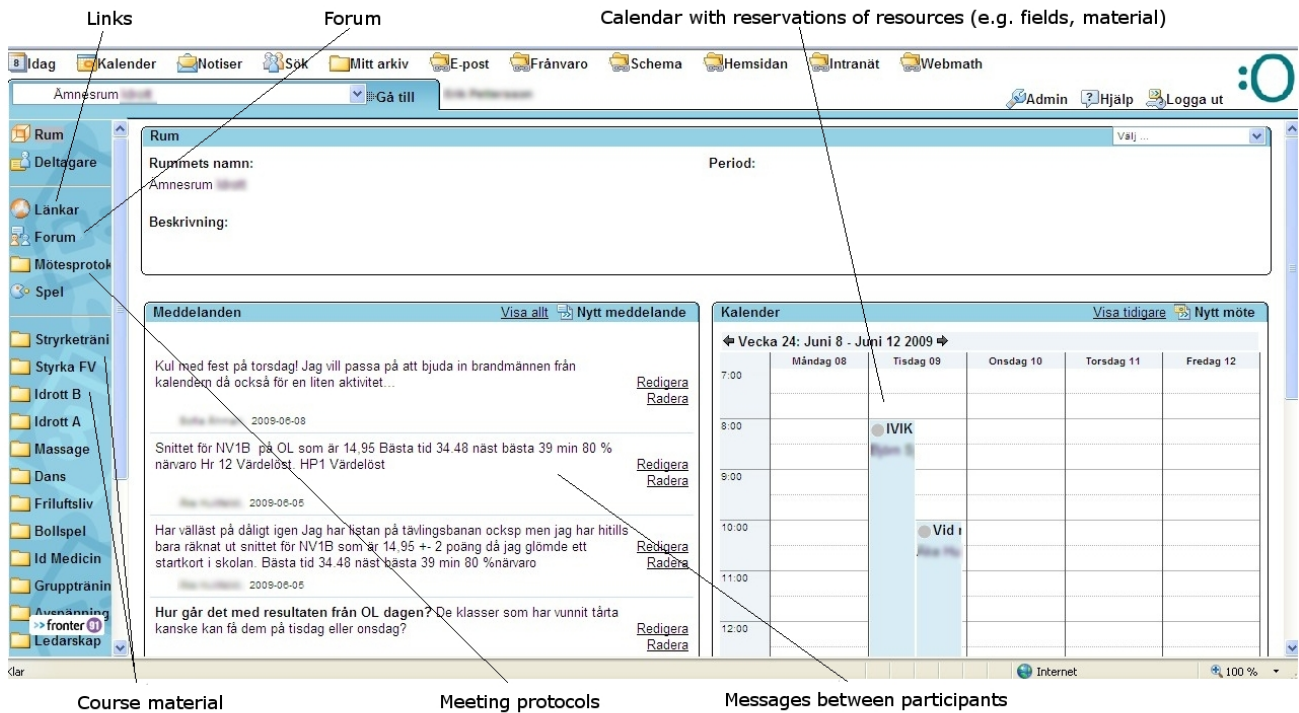


Figure 3. Screenshot of a Fronter room used by physical education teachers.

Phones were rarely used: due to teachers' mobility the usefulness of stationary phones was limited, while the use of cell phones was restrained by school policy. Before the implementation of Fronter, FC was the main system used for collaboration among teachers. The system consists of a suite of integrated applications supporting communication, collaboration, and knowledge sharing within an organization. It includes different components such as email and instant messaging, collaborative workspaces, calendars and file management. At the time the study was conducted the system was however mostly used as an email client. Before Fronter was introduced, each teacher group had one or several conferences in FC (see Figure 1), which were used for interaction, file sharing, and setting up agendas. Conferences were also used by other formations such as various interest-based groups and for top-down distribution of information. A feature in FC that was highly appreciated by many was how a red flag was displayed in connection to a conference if new posts had been made by one of its participants. This enabled teachers to, with little effort, get an overview of recent changes and added information.

Another system used at the school was an Intranet (see Figure 2) that was mainly employed for distributing general information (concerning the municipality, schools in the region, or specifically the school we studied). The Intranet was only used for information dissemination and was not employed to support teachers' collaborative activities.

In spring 2007 the Fronter system (see Figure 3) was introduced to provide a more advanced support for course administration and communication between various groups of users.

Fronter is a learning management system (LMS) designed based upon a room metaphor, which allows collaborating individuals to design and structure their shared workspaces. The system provides various tools for collaboration (e.g. shared calendars and tools for voting and resource booking) and

communication (e.g. messages, chat, and discussion), learning and teaching, administration, and publishing. Even though the system provides numerous tools for collaboration and communication, only a limited subset of these tools have been applied in the school. The figure above (see Figure 3) displays the fronter room of physical education teachers. The tools that this formation made use of for collaborative purposes was *links* in which its participants could add URLs they consider valuable, *forum* in which discussions could take place, *calendar* in which reservations of resources was displayed, *various folders* in which course material and meeting protocols were stored, and *messages* that served as a channel for informal communication and notifications.

RESULTS AND DISCUSSION

Interviews conducted within the study were transcribed and then individually analyzed by two researchers. Observational data and documents collected in the setting were also used in the analysis. A combined set of highlighted user quotes and researchers' comments was organized into key themes during collaborative interpretation sessions. A version of the affinity diagram technique [29] was used and four key themes were identified: (a) deployment of a new LMS system (i.e. Fronter), (b) information and collaboration overload, (c) coordination of physical and virtual collaboration in the setting, and (d) individual strategies for handling multiple collaborative activities.

Deployment of a New Technology

In 2007 the school deployed Fronter, an implementation based upon the result of an analysis of administrative processes and available computer support conducted in 2006. The analysis reports stated that the use of two information channels at the school, the intranet and the FC system, caused stress and fear among teachers to miss important information. The report

recommended ensuring an integrated information environment for all school related activities and stated:

The stress can be avoided if the information flow is properly restructured so that the user can find all information in one place. (“Basis for requirement specification, 2006”, translated from Swedish)

According to our respondents, the vision expressed in the statement above never came true. While Fronter was intended to be the only portal for information sharing within the school, several respondents (Martin, Fredrik, Eva, and Jonas) reported that FC continued to be an important tool that was used on a daily basis. Email handling in FC was for instance considered more suitable for one-to-one interaction than the room metaphor applied in Fronter. Using email for collaborative purposes is reported as important for teachers. One respondent (Martin) mentioned that email was his “main tool” and that he use to send between 30 and 50 emails every day. Even in situations where there were several recipients, email was sometimes preferred over using communication functionality in Fronter. Sara, for instance, mentioned creating a mailing list including email addresses of all teachers in her work group. Fronter as used in the school did not support email, instead teachers needed to share the same Fronter room to be able to send messages to each other.

In addition to the continued use of FC as the email tool of choice, school management reopened a conference in FC. The conference, named “staff info”, was before the deployment of Fronter actively used for all kinds of formal and informal discussions, but as a Fronter room replaced the conference conversations died away. According to Fredrik, the same thing happened to a conference assigned to discussing labor issues. Jonas commented that his workgroup avoided using Fronter for discussions; they preferred to have discussions during physical meetings and use Fronter for other aspects of collaborative work.

The use of the Intranet did not change to any large extent through the implementation of Fronter and continued to be a key tool for top-down information dissemination.

Information and Collaboration Overload

All teachers except Sara stated that there was too much information to handle. When describing their daily use of IT, they mentioned the need to browse not one but several different information/collaboration spaces (e.g., email, intranet, Fronter).

Managing numerous information and collaboration spaces was reported as being an especially demanding task due to a lack of knowledge about where to find which information. Elisabet mentioned the lack of consistency when it comes to where to look for certain information: “[...] *you can find information in different corners of the world, and it creates insecurity and stress.*” When following Elisabet at work we overheard her inform a puzzled colleague that information related to a certain issue had been moved over from a conference in FC to a room in Fronter. Sara mentioned that she had to inform a colleague, who apparently had been browsing, in vein, various Fronter rooms to make a reservation of a computer lab, while that type of reservation was supposed to be done by using a physical binder placed outside the teachers’ coffee room.

Before Fronter was introduced, virtual collaboration took place in a limited number of conferences in the FC system. When Fronter was introduced, much of virtual collaboration moved to Fronter rooms, where it became more fragmented. The number of Fronter rooms to which a teacher was connected soon reached the current level of 10 to 30 rooms. The rooms were

created for very different kinds of informal and formal units, ranging from a teacher and his or her students, to two teachers wanting to share course material, to a group of individuals with a common interest, such as union issues or the physical work environment. According to several respondents (Sara, Fredrik, Martin, Eva) the introduction of Fronter and the decreased use of FC resulted in a confusing information flow. Martin, who dubbed the ensuing situation “a communication crisis”, observed:

[...] it became so confusing that it was demoralizing. People started to be absent from meetings in a way not seen before. [...] That is what is most difficult, that you have so many engagements that you sometimes feel you have to be in two places at the same time, it is very important that communication works.

Eva reported a similar experience and mentioned that with Fronter as the main collaboration support tool it was difficult to develop some sort of overview, i.e. ensure that no relevant information was missed (in FC a red flag signaled that a conference contained new information and disappeared once the information had been read). In general, using several information channels available on different technological platforms presented a serious challenge for the teachers.

Physical/Virtual Collaboration

No respondent mentioned any instance of school collaboration on a group level that was solely conducted through groupware, and only one respondent (Eva) mentioned a group that was not using groupware at all. Most instances of collaboration were of a physical-virtual nature, where physical meetings were combined with the use of Fronter and FC. Some tasks were predominantly dealt with through physical collaboration, while others were typically completed through the use of groupware.

Tasks Managed through Physical Collaboration

Eva considered socializing an important aspect of meeting with colleagues in her workgroup. She felt somewhat isolated in her everyday work and really appreciated the informal conversations taking place during the meetings. When participating in one of the meetings we noticed that all teachers were willing to exchange information on current events with each other; they freely expressed their frustration, happiness, concerns, and ideas. According to Sara, this rarely happened in virtual collaboration. Jonas also thought that Fronter did not provide a proper support for discussions and while acknowledging that Fronter does help people make concrete, well-formulated contributions, he added:

[...], it is easier to have [face-to-face] meetings and talk, that’s how it is, Fronter is no discussion tool. It is no system through which you discuss, it is about sending information.

According to our respondents, a common approach adopted by many workgroups was to mainly use physical meetings for discussions and dividing labor. Sara described a project in her workgroup, which started with an initial discussion at a physical meeting, where Sara and a colleague received the task of looking into available options. The work on the task was reported to the group and a decision was made at a follow-up (physical) meeting. Jonas described a similar arrangement in his workgroup:

There have been six of us working actively and three who have contributed with their opinions. Then we [at a physical meeting] decided who should do which part, there are different parts in the exam, different areas, and then you take your part and finish it.

Tasks Managed through Virtual Collaboration

According to our respondents, some collaborative tasks were preferably dealt with through virtual collaboration. Such tasks included, for instance, announcing appointments and publishing schedules. Elisabet gave an example of a schedule of group meetings posted in the group Fronter room. Sara provided a similar example of a document, containing information about shared time between teachers on a program (something that was first discussed at a physical meeting), being subsequently made available in a Fronter room. Nils, who considered virtual collaboration spaces appropriate for sharing awareness information on current events, observes: “*you can leave short messages such as, during the hour that I had the group we did this and that*”. This is an example of explicit information sharing, but there are also examples of more implicit information sharing in the virtual collaborative spaces, for instance, through notifications that are automatically sent to all members about changes that have been made in a shared document.

Teachers also made use of Fronter for sharing course materials. Jonas described the work of developing a new course in his subject group and reported that they used Fronter for sharing material such as exams, key answers, lectures, presentations, etc. Nils gave a similar example of how he and his colleague invited each other to each others’ Fronter rooms for sharing lecturing materials.

Individual Strategies for Handling Multiple Collaborative Activities

Different strategies were developed by the teachers in order to manage their participation in numerous instances of physical-virtual collaboration.

Nils employed the strategy of never reading minutes from workgroup meetings, even from the meetings, which he did not attend (and he regularly missed meetings due to only working part time), this in order to avoid being overloaded by too much information. He explained that if there was something that concerned him, it was likely that someone of his colleagues would inform him about that.

Jonas adopted another strategy for dealing with information overload, the strategy of only paying attention to information that in a direct way concerned his students or colleagues, if this was not the case he threw it away. He said that if the information he filtered out was important, the information was likely to eventually reach him again in some other way.

Another common strategy was to only monitor certain information spaces. When teachers were asked to describe how they browsed the available systems and information spaces, almost everyone was able to give a thorough description of where they were looking for updates. Most teachers (all except Jonas) visited their Fronter rooms on a regular basis and several teachers mentioned that they sometimes look through parts of the Intranet. Eva however seemed to be avoiding the Intranet altogether and only went there to log on to Fronter, while Sara described her daily routine for information browsing as follows:

And then some time every week I might go further down into the folders and look if there are any new documents, on the Intranet. I look through the places that I find important.

Jonas reported a strategy of only heeding to information that was flagged by a notification. (No other teacher claimed to have adopted this strategy but Jonas talked about it as it were the most natural strategy.) Jonas explained:

In Fronter I check the notifications, if you have not made a notification then I will not see it. [...] You cannot add a new document in a room without flagging it, that is not writing a new message, and then it is your own fault if no one reads it. It doesn’t work, you cannot go into every room to see if something new have happened.

The most common strategy of how to use IT to handle information and interaction was to dedicate fixed time slots for these tasks. Several respondents (Jonas, Eva, Sara) gave quite specific descriptions of when they do this. Given the overall layout of a teacher’s workday, allocating specific times for activities in the virtual space is hardly surprising. When giving lectures it is difficult to collaborate with individuals other than those present in the classroom. Eva was an exception in how she managed virtual collaboration during lectures:

[...] I have time to start the lecture and then I log on and see if there are any messages, in the email. I use to go into Fronter as well and check if there is anything special, and then I rarely use it [i.e., the computer] during the day. [...] Then when the day is over I check again, my email, filling in attendance and so on.

Implications for Design

Exploring the design space of more advanced collaborative technologies was not an objective of our study. However, our findings appear to have some, if rather general, implications for design. In particular, they indicate that:

- There is a need for a more advanced cross-tool monitoring (cf. [30, 31]). To get updated on recent developments in a technology-based workspace, users typically have to actually login to the environment, which might cause significant overhead. Users should be provided with a possibility to monitor the most important developments in several virtual places without the need to actually “visit” them.
- Various levels of collaboration should be supported by collaborative technologies. Virtual collaboration as opposed to physical meetings does not necessarily mean that all participants pay their undivided attention to all tasks, discussions, documents, etc. For some participants less relevant information could be filtered out to make sure the information that really matters is salient enough.
- Since the participants are likely to frequently switch between various tasks, it is important to provide efficient task selection/ contextual clues to help the participants identify the tasks, which require their immediate attention, and restore task contexts as quickly as possible.
- The design of technologies analyzed in our study indicates little or no deliberate effort to provide support for integration of physical and virtual collaboration spaces. Since people do combine physical and virtual collaboration within the same project, providing tools and functionality for such an integration appears to be a promising direction for developing more advanced technological support for collaboration.

CONCLUSIONS

The main focus of our study was neither on evaluation of one particular technology, nor on comparison of computer-mediated and face-to-face collaboration. Instead, our aim was to gain a broad insight into key challenges associated with multitasking in a real-life massively collaborative physical-virtual setting, as well as the strategies used by the participants in the setting to deal with these challenges. A general issue that emerges from

our findings is a causal link the widespread use of technological support for virtual collaboration and the increasing fragmentation of modern work practices.

Bell and Dourish [32] observe that we live in “yesterday’s tomorrow”: much of what 10–20 years ago was considered a remote future has actually come true and became a reality of our everyday life. But, as Bell and Dourish argue, the reality has a tendency to differ from the original vision. Envisioned as a smart solution to making our lives easier, technology often causes fragmentation and messiness. Bell and Dourish specifically discuss ubiComp, but their general conclusions – as testified by the findings of our study – appear to be applicable to CSCW, as well.

Dynamically assembling new combinations of workers and making it possible for an individual worker to efficiently switch between different project teams, in which he or she is a member, presents a significant challenge if the work is accomplished through physical, face-to-face, meetings. The promise of CSCW was to provide virtual spaces for collaboration and thus support smooth and effortless switching between, and integration of, different collaboration contexts [33].

Undoubtedly, CSCW systems have made a profound positive impact on a wide range of work practices. However, our findings indicate that appropriation of groupware by organizations also causes a number of challenges.

First, in contrast to the physical space, which is a unique shared resource, there can be a number of virtual “realities” corresponding to different independent technologies, which might make collaboration fragmented. A straightforward approach to fighting fragmentation would be implementing one system, powerful enough to substitute all other competing technologies. That was exactly the rationale behind the implementation of the Fronter system at the school analyzed in our study. Our findings indicate that this straightforward approach could be unsuccessful. In our case, Fronter became just one more system employed in the setting.

Second, even within a single system it is possible to create multiple collaboration spaces, such as email lists, online groups, or virtual rooms. By supporting multiple collaboration spaces, technology may cause *collaboration overload*, rather than reduce collaboration-related time and effort. It is true that technology can make setting up a new collaboration space or joining an existing one relatively effortless. Low “entry costs” may, however, entice people to create or join too many collaboration loci. Some of our respondents, who became members of numerous Fronter rooms, eventually discovered that everyday handling of all these spaces could be a serious challenge.

Third, in the school we studied collaboration mediated by digital technologies complemented face-to-face meetings rather than constituted an alternative mode of collaboration. The setting provided a set of interrelated physical and virtual resources supporting collaborative activities, and can be described as a hybrid physical/virtual ecology of collaboration. Our respondents perceived physical meetings and online discussions as different aspects of the same collaborative activity, but as mentioned above, available technologies provided limited support for integrating physical and virtual collaboration.

In order to deal with the above challenges, the teachers we studied developed a number of strategies, with a varying degree of success. Knowledge, skills, and working solutions were predominantly acquired “the hard way”, through trial and error. The findings of our study suggest that people in the setting, as well as the organization as a whole, could benefit

from learning about others’ experience of dealing with similar problems, and from a more systematic and coordinated adoption of collaboration strategies and solutions.

It should be noted that our study was conducted in a specific organizational and cultural context, and therefore our findings cannot be directly generalized to other environments. More research is required to establish how collaborative multitasking is embedded in different types of physical, virtual, and organizational contexts, and what technological support is needed to help people deal with the specific challenges of collaborative multitasking.

REFERENCES

- [1] Cypher, A. 1986. The Structure of Users’ Activities. In: *User Centered System Design, New Perspectives on Human-Computer Interaction*, D. Norman, S. Draper (Eds.). Lawrence Erlbaum Associates Inc., Hillsdale, NJ. Pp. 243–264.
- [2] Czerwinski, M., Horvitz, E. and Wilhite, S. 2004. A Diary Study of Task Switching and Interruptions. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, April 24–29, 2004, CHI’04*. ACM Press, New York, NY, 175–182. DOI = <http://doi.acm.org/10.1145/985692.985715>
- [3] González, V. M. and Mark, G. 2004. “Constant, Constant, Multi-tasking Craziness”: Managing Multiple Working Spheres. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, April 24–29, 2004, CHI’04*. ACM Press, New York, NY. Pp. 113–120. DOI = <http://doi.acm.org/10.1145/985692.985707>.
- [4] Plaisant, C., Shneiderman, B., Baker, H. R., Duarte, N. B., Haririnia, A., Klimesmith, D. E., Lee, H., Velikovich, L. A., Wanga, A. O. and Westhoff, M. J. 2007. Personal Role Management: Overview and a Design Study of Email for University Students. In: *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments*, V. Kaptelinin and M. Czerwinski (Eds.). MIT Press, Cambridge, MA. Pp. 143–170.
- [5] Tolmie, P., Crabtree, A., Rodden, T. and Benford, S. 2008. “Are You Watching this Film or What?”: Interruption and the Juggling of Cohorts. In: *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work, San Diego, CA, USA, November 08–12, 2008, CSCW’08*. ACM Press, New York, NY. Pp. 257–266. DOI = <http://doi.acm.org/10.1145/1460563.1460605>
- [6] Leroy, S. and Sproull, L. 2004. When Team Work Means Working on Multiple Teams. Unpublished manuscript, NYU Stern School of Business, New York, NY.
- [7] Kraut, R., Egidio, C. and Galegher, J. 1990. Patterns of Contact and Communication in Scientific Research Collaboration. In: *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, R. Kraut, C. Egidio and J. Galegher (Eds.). Lawrence Erlbaum Associates Inc., Hillsdale, NJ. Pp. 149–171.
- [8] O’Connell, B. and Frohlich, D. 1995. Timespace in the Workplace: Dealing with Interruptions. *Conference Companion on Human Factors in Computing Systems, Denver, Colorado, United States, May 07–11, 1995, Chi’95*. ACM Press, New York, NY. Pp. 262–263. DOI = <http://doi.acm.org/10.1145/223355.223665>.

- [9] Farhoomand, A.F. and Drury, D.H. 2002. Managerial Information Overload. *Commun. ACM*, 45, 10, Oct. 2002, 127–131. DOI = <http://doi.acm.org/10.1145/570907.570909>.
- [10] Ljungberg, F. and Sörensen, C. 2000. Overload: From Transaction to Interaction. In: *Planet Internet*, K. Braa (Ed.). Studentlitteratur, Lund, Sweden. Pp. 113–136.
- [11] Nardi, B. and O'Day, V. 1999. *Information Ecologies: Using Technology with Heart*. MIT Press.
- [12] Krippendorf, K. 2006. *The Semantic Turn: A New Foundation for Design*. Taylor and Francis.
- [13] Crabtree, A. and Rodden, T. 2008. Hybrid Ecologies: Understanding Cooperative Interaction in Emerging Physical-Digital Environments. *Pers. Ubiqu. Comput.*, 12(7), 481–493. DOI = 10.1007/s00779-007-0142-7.
- [14] Gonzalez, V. M. and Mark, G. 2005. Managing Currents of Work: Multi-tasking Among Multiple Collaborations. In: *Proceedings of the Ninth Conference on European Conference on Computer Supported Cooperative Work*, Paris, France, September 18–22, 2005. ECSCW'05. Springer-Verlag, The Netherlands. Pp. 143–162.
- [15] Makato Su, N. and Mark, G. 2008. Communication Chains and Multitasking. In: *Proceeding of the twenty-sixth annual SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, April 05–10, 2008. CHI'08. ACM Press, New York, NY. Pp. 83–92. DOI = <http://doi.acm.org/10.1145/1357054.1357069>
- [16] Aral, S., Brynjolfsson, E. and Van Alstyne, M. 2007. Information, Technology, and Information Worker Productivity: Task Level Evidence. Working paper 13172. National Bureau of Economic Research, Cambridge.
- [17] Baron, N. S. 2008. *Always On*. Oxford University Press.
- [18] Sproull, L. S. 1984. The Nature of Managerial Attention. In: *Advances in Information Processing in Organizations*, L. S. Sproull and P. D. Larkey (Eds.). Jai Press, Greenwich, CT. Pp. 9–27.
- [19] Hudson, J., Christensen, J., Kellogg, W. A. and Erickson, T. 2002. "I'd Be Overwhelmed, But It's Just One More Thing to Do:" Availability and Interruption in Research Management. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Minneapolis, MN, USA, April 20–25, 2002. CHI'02. ACM Press, New York, NY. Pp. 97–104. DOI = <http://doi.acm.org/10.1145/503376.503394>.
- [20] Mintzberg, H. 1973. *The Nature of Managerial Work*. Englewood Cliffs.
- [21] Eppler, M. J. and Mengis, J. 2002. The Concept of Information Overload. A Review of literature from Organization Science, Marketing, Accounting, MIS, and related Disciplines. Retrieved June 24, 2009, from <http://www.netacademy.org>.
- [22] Mulder, I., de Poot, H., Verwij, C., Janssen, R. and Bijlsma, M. 2002. An Information Overload Study: Using Design Methods for Understanding. In: *Proceedings of the 18th Australia Conference on Computer-Human Interaction*, Sydney, Australia, November 20–24, 2006. OZCHI'06. ACM Press, New York, NY. Pp. 245–252. DOI = <http://doi.acm.org/10.1145/1228175.1228218>.
- [23] Franklin, D. 1997. Treat Depression with more than Drugs. *Health*, 11(3), 122.
- [24] LaPlante, A. 1997. Still drowning! *Computerworld*, 31(10), 69.
- [25] Hallowell, E. M. 2005. Overload Circuits: Why Smart People Underperform. *Harvard Business Review Article*. *Harvard Business Review*, March 3, 2009.
- [26] Harr, R. and Wiberg, M. 2008. Lost in Translation: Investigating the Ambiguity of Availability Cues in an Online Media Space. *Behav. Inform. Technol.*, 27(3), 243–262. DOI = 10.1080/01449290600874865.
- [27] Harr, R. and Kaptelinin, V. 2007. Unpacking the Social Dimension of external Interruptions. In: *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, Sanibel Island, FL, USA, November 04–07, 2007. GROUP'07. ACM Press, New York, NY. Pp. 399–408. DOI = <http://doi.acm.org/10.1145/1316624.1316686>.
- [28] Hammersley, M. and Atkinson, P. 1983. *Ethnography: Principles in Practice* (3rd edition). Routledge.
- [29] Beyer, H. and Holtzblatt, K. 1998. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann.
- [30] Boardman, R. and Sasse, A. M. 2004. "Stuff goes into the Computer and doesn't come out": A Cross-tool Study of Personal Information Management. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vienna, Austria, April 24–29, 2004. CHI'04. ACM Press, New York, NY. Pp. 583–590. DOI = <http://doi.acm.org/10.1145/985692.985766>.
- [31] Kaptelinin, V. and Boardman, R. 2007. Toward Integrated Work Environments: Application-Centric versus Workspace-Level Design. In: *Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments*, V. Kaptelinin and M. Czerwinski, (Eds.). MIT Press, Cambridge, MA. Pp. 295–331.
- [32] Bell, G. and Dourish, P. 2007. Yesterday's Tomorrows: Notes on Ubiquitous Computing's Dominant Vision. *Pers. Ubiquit. Comput.*, 11(2), 133–143. DOI = 10.1007/s00779-006-0071-x.
- [33] Bannon, L. 1993. CSCW: An Initial Exploration. *Scand. J. Inform. Syst.*, 5(3), 3–24.

Reconceptualising Clinical Handover: Information Sharing for Situation Awareness

Stephanie Wilson, Rebecca Randell, Julia Galliers, Peter Woodward

Centre for HCI Design
City University London
Northampton Square
London EC1V 0HB, UK

steph, rebecca.randell.1, jrg, peter.woodward.1@soi.city.ac.uk

ABSTRACT

Clinical handover, associated with the transfer of responsibility for patient care, is usually regarded as a single point of transition. Drawing on data from ethnographic studies of handover undertaken across a range of clinical settings, we suggest it may instead be useful to reconceptualise handover as a process that occurs over a period of time. We discuss the implications of this view and how it is compatible with construing the purpose of the information sharing that generally accompanies handover as being to promote good situation awareness in the distributed cognitive system of the clinical setting.

Keywords

handover, situation awareness, distributed cognition

ACM Classification Keywords

K.4.3 [Computers and Society]: Organizational Issues – Computer-Supported Collaborative Work.

INTRODUCTION

A common characteristic of many work systems, particularly many critical systems, is that of continuous operation. Work must continue twenty-four hours a day, every day of the year. This requirement for continuity necessitates a series of transitions between the human operators who are responsible for specific roles in the system, for example, the transitions of responsibility that occur between controllers in the domain of air traffic management.

For the last few years we have been investigating continuity of work in the healthcare domain. In this case, the concern is continuity in the provision of patient care across transitions in responsibility. Effective transitions are reported as contributing to a safe patient journey; or, from another perspective, poor transitions have been implicated in incidents of poor patient outcomes, patient harm and ineffective work [7], [17].

It is in hospital settings that care transitions are most evident. Care for patients in hospital is provided by complex, dynamic

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

and often unpredictable distributed cognitive systems that include people, information technologies, equipment and procedures. Care transitions are most evident here because hospital patients generally require frequent monitoring and regular treatment interventions. These continue across boundaries of time (the transitions of responsibility that occur as healthcare professionals change shift) and boundaries of space (the transitions of responsibility that occur as the patient progresses from one clinical setting to another, for example from Accident and Emergency (A&E) department to admitting ward). However, taking a holistic view, care transitions actually occur across a person's lifetime, in both hospital and community settings.

The transfer of responsibility for patient care at each of these points of discontinuity in time and space is commonly referred to as clinical handover (but see discussion in section 4). Clinical handover is generally regarded as a single point of synchronous transition, where responsibility for the system is simultaneously relinquished by one party and accepted by the other. An implication of this is that all information necessary for continuous safe care is passed and received at that point in time. In practice, this has been manifest in the staff who are handing over (primarily medical and nursing staff) preparing a handover document and/or giving a verbal summary to the receiving staff. The recent proliferation of work on "minimum data sets" for handover, i.e. the minimum information that should be communicated at every handover, has tended to reinforce this view of handover as "passing the baton". Our aim in this paper is to revisit this view and suggest that it is time to consider an alternative. Drawing on some of our recent studies, we propose a reconceptualisation of clinical handover as a process and suggest that the challenge of improving handover can then be construed in part as one of improving situation awareness.

BACKGROUND

For many years, clinical handover attracted relatively little attention from either healthcare professionals or researchers and, while transitions of responsibility obviously occurred, the practice of handover varied considerably. This has changed. Clinical handover and its contribution to patient safety have attracted substantial interest over the last few years, not least because of investigations such as [7] where poor handover was reported as a serious shortcoming. The importance of effective handover is also suggested by studies such as [9] which reported an increase in adverse events during cross-coverage (where a patient is temporarily assigned to a covering doctor who is not primarily responsible for their care) due to poor information transfer and [8] in which a survey at two teaching hospitals revealed that trainee doctors perceived problems with handover as the reason for 15% of mistakes. A further impetus to focus on handover has been changes in the organization of

healthcare work: both the increased specialization of medical work and the increase in shift work resulting from the reduction in working hours for junior doctors in the UK and elsewhere have led to more frequent transfers of patients. These changes mean more handovers and greater cross-coverage.

While continuity of care across transitions in responsibility is the primary goal of handover, this does not necessarily mean that the care is unchanged: a patient may be transferred to another setting precisely in order to enact change in their care. Handover has also been reported as achieving other important outcomes beyond the continuity of care for individual patients. For example, Wears et al [13] reported how shift changes in the Emergency Department can be a time for identifying problems in care provision and recovering from failure, while we previously discussed how handover provided an opportunity for developing treatment plans, checking that work had been completed, educating junior staff and promoting social cohesion [14].

Much of the current focus is on shift handovers, on the information communicated at handover and on improving handovers through standardization using mechanisms such as minimum data sets and standard operating protocols. For example, the Royal College of Physicians (RCP) in the UK undertook a review of published work on medical record keeping standards and a substantial consultation exercise, resulting in a suggested minimum set of data items to be included in medical documents for handover [11]. Other work, recognizing that what may be suitable for one setting is not necessarily appropriate for another, has tried to define the content and structure of handovers at a higher level. For example, the Australian Commission for Safety and Quality in Healthcare recommended the ISOBAR protocol (Identification of Identify, Observations, Background and History, Assessment and Actions, Responsibility and Risk management) in its recent guide [1]. The form and use of artefacts (e.g. see Figure 1) to support handover have also been examined through studies of current practice [10], [12]. However, a recurrent, implicit assumption in all this work seems to be that handover is a single, clearly defined episode in time.

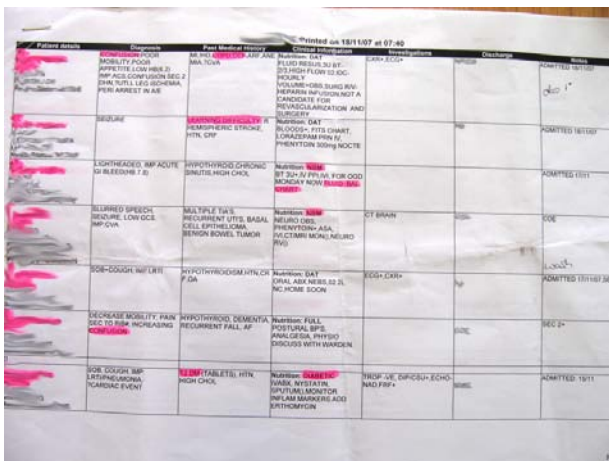


Figure 1. Example artefact to support handover: a “handover sheet”.

HANDOVER AS IT HAPPENS

We have undertaken substantial, ethnographic field studies of clinical handover in hospital settings over a two year period and these are still ongoing¹⁷. We have studied different kinds of handover, primarily involving nursing and medical staff: nursing shift handovers, medical shift handovers (including to night teams), temporary delegations of responsibility and transfers between settings.

In this paper, we draw on data collected in six of our field studies undertaken in varied hospital settings in the UK. The six field studies are summarized in Table 1. Two of them were undertaken in a District General Hospital (DGH): an Emergency Assessment Unit (EAU) where patients are transferred temporarily from A&E prior to either discharge or transfer to a specialist ward and a general medical ward where patients under the care of physicians stay on a longer term basis. The other four studies were undertaken in large, inner city teaching hospitals: a relatively small paediatric surgical ward which looks after children before and after elective and emergency surgery; a specialist ambulance transport service which transfers critically ill children from local hospitals to paediatric intensive care units; a high-dependency unit which looks after patients who require continuous electronic physiological monitoring (telemetry) and a Medical Assessment Unit (MAU) which is a short stay unit for patients arriving from A&E or EAU who are to be admitted to other wards. The studies took the form of non-participant observations recorded as field notes, audio recording of verbal communications and informal interviews with staff. We gathered examples of artefacts used to support handover and took photographs of the settings. In total, we undertook 660 hours of study in these six settings. Research Ethics Committee approval was obtained for this project and informed written consent was obtained from all staff and patients.

These studies have yielded a corpus of data distinguished by its breadth and depth. It is informing our work on understanding current practice in handover and the design of technology to support handover. We analyzed the data (as summarized below) to understand handover as it happens at present. However, it is not the purpose of this paper to report these results; rather to reflect on how undertaking the analysis caused us to face a number of questions about handover.

Following the data collection, a “cognitive landscape” was written for each setting. This was a narrative account of the setting, describing the physical environment, people, organisation and processes of work and, importantly, the cognitive artefacts that supported the work. Following this, a qualitative data analysis tool was used to index the data, identifying all handovers that were observed. The data for each setting was then analysed individually using a grounded theory approach so as to allow themes that were unique to each setting to emerge from the data. The field notes and audio transcripts for the handovers were firstly read and then coded. We paid particular attention to what was occurring and in what order, what was being accomplished and what strategies were used to achieve this. We identified the types of handover, duration, location, participants, artefacts, information communicated, structure, purposes and strategies. Differences across settings and types of handover became apparent.

¹⁷ Hence the field studies reported here are not necessarily the same as those in other reports of this work.

Table 1. Summary of six handover field studies.

	Hours of field study	Handover types	Unit description
Emergency Assessment Unit (EAU)	172	Medical shift handovers Nursing shift handovers Transfers	28-bed, short stay ward in District General Hospital (DGH). Patients transferred in from A&E or by direct referral from a General Practitioner (GP). Patients transferred out to appropriate wards or discharged.
General Medical Ward	104	Nursing shift handovers Transfers	20-bed general medical ward in District General Hospital (DGH). Mostly elderly patients.
Paediatric Surgical Ward	92	Medical shift handovers Nursing shift handovers Transfers	11-bed paediatric ward in large teaching hospital, catering for elective and emergency surgical patients. Patients under care of a range of surgical teams (e.g. orthopaedic, neurology).
High Dependency Unit	29	Medical shift handovers Nursing shift handovers	13-bed ward in large teaching hospital, for patients who require continuous telemetry monitoring and coronary care patients.
Ambulance Transport Service	111	Transfers	Service staffed by medical and nursing teams, transferring critically ill children from DGHs to paediatric intensive care units in other hospitals.
Medical Assessment Unit	152	Medical shift handovers Nursing shift handovers Transfers	28-bed, short stay ward in large teaching hospital. Patients admitted from A&E or EAU prior to transfer to another ward.

This process led us to question what we meant by “handover” through confronting questions such as “where are the handovers?”, “when does the handover commence and when does it end?” and “what information is shared?” and ultimately to reframe our definition of handover in order to arrive at a definition that encompassed and adequately characterized all the episodes that we might intuitively consider to be handover.

CLINICAL HANDOVER: A REFRAMING

The field studies highlighted the extent to which episodes referred to as “handover” by healthcare staff differed from one setting to another. The term “handover” was generally used to refer to a collaboration between two or more people, accompanying a transfer of responsibility, in which information was communicated in verbal and/or written form. People would talk about “taking” or “giving” the handover. Handovers were responsive to the context in which they occurred. They varied in terms of structure, information content, supporting artefacts, participants, location, duration etc. Local practice, while informed by guidelines and standard protocols, has generally evolved to satisfy local needs. Even within a given setting, the handovers varied depending on contingent circumstances. For example, when a patient was transferred into the EAU from A&E, the standard practice was for the A&E nurse to pass information directly to the EAU nurse who was assuming responsibility. However, if the EAU nurse was unavailable for any reason, the A&E nurse did not have time to wait and would instead pass the information to another EAU nurse who would later pass it on to the responsible nurse.

Yet this everyday sense of handover as a collaboration, reflected also in much of the literature, does not altogether match the more formal definitions. A widely accepted definition of handover offered by the British Medical Association in collaboration with the National Patient Safety Agency in the UK [4] equates it with the transfer of responsibility: “The transfer of professional responsibility and accountability for some or all aspects of care for a patient, or group of patients, to another person or professional group on a temporary or permanent basis”. The Australian Medical Association has adopted the same definition [2]. In contrast, the definition of

handover given in [5] emphasizes the exchange of information as being the defining characteristic of handover (“handoff”) while limiting the scope to information exchanges where there is also a change in control or responsibility: “the exchange between health professionals of information about a patient accompanying either a transfer of control over, or of responsibility for, the patient”.

A Process

As a starting point, and as mentioned in the introduction, we consider handover to occur when there is a transfer of responsibility for some aspect of patient care on a permanent or temporary basis. This means that we do not consider there to be a handover when one healthcare professional updates another without any change in responsibilities, but we do include temporary delegation of responsibility and resumption of responsibility.

We view clinical handover as a process that occurs when there is a transfer of responsibility for some aspect of patient care from one party to another (and note that many healthcare professionals may have responsibility for a patient at any point in time, each responsible for a specific aspect of care). This view distinguishes “handover” from “transfer”: it is not the transfer of responsibility itself, but the process within which responsibility is transferred.

The handover process occurs over a period of time. It is not a single point of transition. For example, while there was a formal handover meeting at medical shift change in several of our settings, the outgoing doctors may well start to update the oncoming staff on a less formal basis prior to the meeting and continue to do so afterwards. Likewise, oncoming staff might read handover documentation or medical notes prior to a handover meeting, as is evident in this excerpt from field notes for the general medical ward:

The outgoing nurse says that the patient is ‘for echo’ but the oncoming nurse disagrees. The outgoing nurse says that the patient is for ‘repeat echo’ but still the oncoming nurse disagrees. To resolve the issue, they get the patient’s medical record out of the trolley. In it, the Specialist Registrar has written a note saying that they

have agreed that a repeat echo is not needed. The oncoming nurse knows this from having looked through the medical notes before the handover.

(Field notes, General Medical Ward, nursing evening shift handover)

In some cases, the duration of the handover process may be very short (for example, in our studies, the handover process that occurred when a patient was transferred into the EAU from A&E was generally brief); in others it may be more prolonged (for example, in the ambulance transport service, the handover from a local hospital started with a phone call to the service and later continued when the ambulance team arrived at the hospital to collect the patient).

This reframing of handover as a process arose from the fact that, having set out to study handover, we were often confronted with the questions of when a handover had occurred and whether a particular information exchange was a handover at all. Rather than seeking an explicit identification of a moment at which the baton of responsibility was passed, which was almost inevitably not the same moment at which information was communicated, we concluded that it was more appropriate to look at the overall process within which both responsibility and information were transferred and define this as the handover.

Components of the Handover Process

Secondly, we conceive this handover process as consisting of three components:

- The passing of responsibility
- The acceptance of responsibility
- The sharing of relevant information.

Passing and Accepting Responsibility

We distinguish handover from other collaborations in the clinical setting by the fact that there must be a transfer of responsibility as part of the handover process. This transfer consists of the passing and acceptance of responsibility.

The transfer of responsibility in our settings was often implicit in the organization of the work. For example, at shift change one person would leave work and another would start, without any communication or other token exchange between them. When a patient was transferred by a porter from one ward to another, the patient's departure from one location and arrival in the other denoted the relinquishing and acceptance of responsibility. In these situations, the transfer of responsibility occurred irrespective of whether or not there was any additional, explicit indication. Sometimes the change in responsibility was less apparent (at least to us as observers). For example, when the ambulance service arrived at a DGH, they received information from the local medical staff about the patient and started to care for him/her. It was unclear whether responsibility had now passed to the ambulance service or whether this only happened when the patient was moved into the ambulance, their "space". It seemed that, in practice, there was a gradual passing of responsibility, with the ambulance clinical team starting to accept responsibility while the patient was still in the care of the DGH and its clinicians.

In other situations, particularly in more critical and rapidly changing situations, there was more visible flagging of the passing of responsibility, usually through verbal and/or written communication between the two parties. At nursing shift change on the general medical ward, an outgoing nurse would give a handover to the oncoming nurse and then leave the

ward, with responsibility automatically transferring to the oncoming nurse. On the paediatric surgical ward, a written document with details of all patients on the ward, the "doctors' list", was passed at medical shift change. This was accompanied by a verbal update that usually covered just those patients who might need to be seen or for whom there were outstanding tasks to be done.

[On call Senior House Officer] only tells [night Senior House Officer] about one [paediatric surgical] patient... bloods need to be chased... [On call Specialist Registrar] handed over one [paediatric surgical] patient - the patient with the distended tummy.

... [On call Senior House Officer] hands over the [paediatric surgical] patients. This takes about thirty seconds. He looks at the doctors' list for the paediatric surgical ward and says "There wasn't anything really. [Patient name]'s orthopod. Orthopaedic patient, liver patient, nothing for us to do" (as he points at the different names on the list). When she comes back, [night Specialist Registrar] asks what she has missed. [On call Senior House Officer] says about [ward name], there's nothing to do.'

(Field notes, Paediatric surgical ward, medical shift handover from on-call to night staff)

Perhaps surprisingly, the relinquishing and acceptance of responsibility is not always a clear, synchronous transition. There was sometimes an ambiguous intermediate state where responsibility was temporarily passed to a person or persons who would not ultimately be responsible for this aspect of patient care. For example, we observed nursing shift handover meetings where an outgoing nurse would 'handover' information for his/her patients to the oncoming team as a whole but responsibility for individual patients would only be assigned to staff at the end of the meeting.

In identifying the passing and acceptance of responsibility as two distinct components of the handover process, this reframing explicitly acknowledges the role of the person receiving handover. In current practice, the recipient is sometimes a passive participant in the process, particularly with regard to the acceptance of responsibility. We found little observable evidence of the acceptance of responsibility: as described above, the acceptance of responsibility at shift change and in inter-setting transfers was largely enshrined in the work practice. An exception was the ambulance transport service, where the consultant physician had to agree that the transfer could go ahead before the patient could be moved from the DGH into the full care of the service. This is an area that warrants further work to investigate mechanisms for more explicit acknowledgement of the acceptance of responsibility and their impact on safe patient care. In contrast, the recipient of a handover was more frequently an active participant in the sharing of information. Verbal handovers were not merely a one-way passing of information, they were dialogues where the recipients played an active role in ensuring they had acquired sufficient information to enable them to care for the patients. For example, in this excerpt from a shift handover we see the oncoming nurse not just accepting the information but asking questions of the outgoing nurse in order to connect disjoint pieces of information and form a bigger picture of the situation:

Outgoing nurse: he's had a CT scan, I thought we'd stopped his clexine, yep, and he is to go for a bronchial scope today.'

Oncoming nurse: 'That's why he's nil by mouth?'

Outgoing nurse: 'Yeah, nil by mouth for that because um [consultant] cancelled his clinic for the scopes yesterday'

so he's got all of yesterday's patients and whatever's built up today, so it could be anything up to 10 o'clock.

(EAU, nursing shift handover)

This was active participation in information sharing was particularly striking in the case of the ambulance transport service: the handover of information from the DGH would begin with the DGH doctor providing an overview of the case, but would gradually progress to the doctor and nurse from the ambulance service asking questions:

Ambulance doctor: What's the blood pressure?

DGH Anaesthetist: The last one, nineteen four over sixty three, the previous one I saw was one oh five systolic. Er his capillary refill is still sluggish.

Ambulance doctor: Alright.

DGH Anaesthetist: But better I think than it was.

Ambulance doctor: How much? Two? Or three?

DGH Anaesthetist: I couldn't quantify it...

(Ambulance transport service: handover from anaesthetist at DGH to ambulance service)

Information Sharing

The third component of the handover process is information sharing. This fits with practitioners' everyday sense of clinical handover and yet is not mentioned in the definition given in [4]. In some cases, the information sharing may be minimal even (in current practice) non-existent, though it is not at all clear that this is desirable in any situation.

As mentioned earlier, recent literature tends to promote a view of clinical handover as a distinct event and seeks to identify the set of information that should be shared at that point. This is particularly evident in the work on minimum data sets. Our field studies revealed that while the information sharing that occurs as part of the handover process does sometimes happen this way, it also frequently occurs in a variety of other ways and there are a number of (mainly practical) reasons for this. There was variation in when information was shared and what information was shared.

The information sharing may be removed in time from the transfer of responsibility. It sometimes occurs in advance, as in the case of patient transfers from the EAU where the sharing of information happened by phone and in advance of the patient being physically transported to an admitting ward. Alternatively, information may be shared some time after the passing and acceptance of responsibility. For example, on the general medical ward, no passing of information from the ward medical team to the on-call or night team was observed but those teams implicitly accepted responsibility for the patients when they came on shift. If a member of the on-call or night team was called to the ward, one of the nursing staff would provide information about the patient.

This example also highlights that in some situations there was no sharing of information. Transfers of responsibility to on-call teams were an obvious example but there were other cases as well. In the medical shift handovers in the EAU and paediatric surgical ward, information would be shared about only those patients that were likely to deteriorate or for whom there were tasks to be done. However, responsibility was assumed at the beginning of the shift for all patients, regardless of whether or the staff had received information about them. Doctors may be called to see a patient whom the outgoing doctor did not give them information about. Alternatively, doctors might not,

during their shift, come into contact with patients that they had been given information about.

Another variation on information sharing was evident in the ambulance transport service: in this case, there was explicit sharing of information, but it happened in stages. The service would receive an initial handover of information about the patient during a first telephone referral from a DGH. This was followed by a face-to-face handover when they arrived at the DGH and information may also be shared at other points as it became available.

As regards what information was shared, the details of this varied considerably depending on the setting which was related to factors such as how much was known about the patient, the nature of their complaint, what had been done so far and what needed to be done, the staff and their experience etc. As has been reported previously [3] [14], the information shared during the handover process is not just about specific patients but more generally about the status of the work system, e.g. the medical shift handover to the night team on the EAU would highlight the patients to be seen and the order in which they should be seen, while the nursing shift handovers that we observed on the paediatric surgical ward always began with a discussion of staffing issues before progressing to sharing information about individual patients and concluding with a summary of anticipated admissions. It was also common to share information about possible future changes to the status of individual patients or the system as a whole:

Having gone through the patients, [outgoing charge nurse] tells them about the expected admissions. She gets this information from the ward book. 'You've got two coming in but you've got one bed.'

(Paediatric surgical ward, nursing evening shift handover)

Finally, we deliberately use the term "sharing" to connote that the information flow in the handover process is not just one-way, a fact also reported by others such as [3]. We observed incidents where a person receiving a handover appeared to know as much about the patient as the person giving it:

Outgoing nurse: 'category not stated... slept quite late last night, about half past one, because she said she slept all day.'

Oncoming nurse: 'I think that leg looks horrific... she needs dermatology review.'

Outgoing nurse: 'They stopped IV fluid... they did a blood culture... she's going to a nursing home.'

Oncoming nurse: 'On Saturday.'

...

The outgoing nurse says the name of the next patient on the list, to which the oncoming nurse responds, 'I can't believe he's still here. He was in tears. How was he Monday night?... Have you seen his toes? They're black on both feet.'

(General medical ward, nursing shift handover)

[Oncoming Specialist Registrar] says that the patient has periods of vomiting/diarrhoea but [outgoing Senior House Officer] says he didn't know anything about this.

(Paediatric surgical ward, medical shift handover)

Information is distributed around the cognitive system of the clinical setting, in external cognitive artefacts and in the heads of the staff (and patients). In some cases, the information sharing that happens during the handover process is a "push" as those passing responsibility proactively bring some of this

information to the attention of those accepting responsibility for patient care: they highlight the information they judge the accepting party will require. In other cases, the information sharing is a “pull” as staff assuming responsibility seek out the information they require. Mostly it is a combination of the two. The information push or pull may happen before or after the passing of responsibility or both; it may happen verbally during a handover meeting or by reference to external cognitive artefacts or both. Either way, the sharing is impeded when the information captured in these artefacts is incomplete, out-of-date or inconsistent, and this happens frequently. This should not be surprising when one considers how effortful it is to maintain artefacts such as nursing and medical notes. It can be difficult to know where to locate information across the range of different artefacts and there are often gaps: some information is simply not captured in a tangible form, for example information about the setting as a whole or about social issues such as what the family have been told or that a patient is upset or aggressive.

In summary, we do not equate handover with just the transfer of responsibility or just the sharing of information; instead we define it as a process that comprises both elements. Reworking the phrasing in [4]: “Clinical handover is the process by which professional responsibility and accountability for some or all aspects of care for a patient, or group of patients, is relinquished by one person or professional group, accepted by another on a temporary or permanent basis, and in which relevant information is shared between the two parties”. This conception of handover allows us to limit our analysis to just those episodes where there is a relinquishing and acceptance of responsibility irrespective of how protracted they are and whether or not there is a sharing of information.

IMPLICATIONS

Reframing handover as a process offers a more holistic view of the role handover plays in the distributed cognitive system. The process view removes the emphasis from a single communication of information. Instead, we see the communication of information as an ongoing process of information sharing that promotes situation awareness.

We previously reported a study of medical shift handover [14] [16] in which we suggested that, in preparing for handover, doctors were creating their own mental representation of the “state-of-the-ward”. In the subsequent handover meeting, the doctor giving a verbal summary would use this representation to pass on the information he or she judged to be relevant – the information that would enable oncoming staff to create their own representation of the state-of-the-ward. We identified that one impediment to effective handover was the lack of a readily available, up-to-date physical representation of the state of the ward. Consequently, medical staff preparing for handover would have to glean the information from a variety of sources including colleagues, ward whiteboards, medical notes, previous handover documents etc. Extending this in line with the process view of handover offered here, we consider that clinicians have an ongoing awareness of the state of the system, including the patients for whom they are caring. If those assuming responsibility in the handover process have a good ongoing awareness of the state of the system, there is less onus on one point of information assimilation and transfer. Likewise, if those passing responsibility have ongoing awareness, preparations for transfer will be less effortful.

These ideas fit with the concept of situation awareness as used to describe the state of knowledge that workers have of the

dynamic environments in which they operate and which support their decision making [6]: “Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.” In the handover process, clinicians perceive visual and verbal elements in the environment, comprehend what they mean for individual patients in terms of their status and what they mean for the setting as a whole, and project how things might change. They use this situation awareness to organize and prioritize their work. It is a basis for making decisions about practical issues (such as bed management, staffing and transfers) as well as contributing to care and treatment plans for individual patients. This was particularly evident in the Medical Assessment Unit where a “board round” in front of a ward whiteboard was a forum for passing on information and making these kinds of decisions.

Considering the purpose of the handover process as being not just to transfer responsibility but also to promote good situation awareness, we take two implications for our work. Firstly, we suggest there is an opportunity for removing the reliance on a single point of information sharing during the handover process. In at least some clinical settings there are opportunities for the information sharing and construction of situation awareness to take place over a longer period of time. Second, we believe that the first part of the definition of situation awareness points to one of the primary obstacles for healthcare staff establishing good situation awareness: the difficulty they experience in perceiving relevant elements in the environment.

Technology Implications

Finally, we conclude that this reframing has implications for the design of work practice and supporting systems, including IT systems, to support the handover process. The work of the clinical setting needs to be captured on an on-going basis and made visible in a way that is accessible to those who will eventually assume responsibility. Healthcare staff need to be able to produce and consume this awareness in a non-effortful way.

This is not just an issue for the handover process. Our observations across the varied field settings consistently showed that, outside of specific interventions with patients, healthcare staff spend a vast amount of their time maintaining their own and others’ situation awareness: by asking questions, by answering questions and by “telling” each other things. The following snippet from a handover to the night team on the paediatric surgical ward gives a sense of the importance of verbal communication for the work:

... [On call Specialist Registrar] handed over one patient – the patient with the distended tummy - saying that the [ward] Senior House Officer came and asked him to see the patient but then he got called to A&E so he didn't go. He says that the paediatric consultant has been to see the patient but he doesn't know the outcome – he tells [night Specialist Registrar] to check the notes and see if the consultant wants anything done, although he says that the consultant would probably have called if he did.

(Paediatric surgical ward, medical shift handover to night staff)

Much of this maintenance of situation awareness is achieved through verbal communication, which is easy for those concerned to accomplish but which leaves no trace for others to benefit from. The information does not persist other than in the heads of those who heard it and this is therefore one factor that makes it difficult for others to perceive relevant elements of the environment. Other factors connected to with external cognitive artefacts have

already been mentioned (incompleteness, inconsistencies and inaccuracies).

Because handover has been seen as a passing of the baton, much of the recent work on providing support for handover has been concerned with either general guidelines (e.g. hold handover meetings in a dedicated space, at fixed times, with no bleeps), or with providing specific templates to capture the data sets [11] that should be passed on. This is reflected also in the IT systems to support handover, many of which are implemented to support local practice and have evolved from paper-based systems. They support the construction and sharing of a data set but have not been designed to promote ongoing situation awareness, although healthcare staff do sometimes use them in this way, referring to them and updating them outside of the handover process.

In previous work [15], we described a research intervention to introduce a large projected display into a handover meeting in order to improve information sharing. However, while this was a shared representation, its role was to support a relatively short-lived process: information sharing and decision making within the immediate setting of the meeting. In line with the reconceptualisation offered here, we are now looking at how technology can support handover as a process and how shared displays have a role in this. This is particularly relevant when the handover process is protracted, as in the case of the ambulance transport service, or when there is the opportunity for staff to build up situation awareness over a period of time, as on the general medical ward where the throughput of patients is slower than in the other settings that we studied. We are now developing these ideas in collaboration with the ambulance transport service. Our aim is to develop a technology intervention that supports information sharing between the distributed team: the clinical staff who travel in the ambulance to a DGH to stabilize and move the patient; the more senior clinical staff who generally remain at the base and, ultimately, the intensive care unit who will receive the patient. We have identified two distinct aspects of situation awareness that are important in this setting: knowledge of the current status of retrievals (e.g. where the ambulance team is at any point in time) and knowledge of the medical status of individual patients. Our system aims to capture this information on the fly, with minimal effort on the part of the ambulance team, because it is apparent that the significant effort required to create some external cognitive artefacts is a major factor impacting their utility. The information is then distributed to the staff at the base and the intensive care unit where shared displays make it readily available. Our expectation is that this will improve information sharing across the work as a whole, altering how information is shared in specific handover episodes and promoting better situation awareness. We have already collected baseline data and post-intervention evaluation studies will commence in the near future to investigate these issues.

SUMMARY

Drawing on substantial data from studies of handover across six varied healthcare settings, we have suggested a reframing of clinical handover as a process that occurs when there is a transfer of responsibility for some aspect of patient care. We have identified the relinquishing of responsibility, the accepting of responsibility and the sharing of information as distinct components of this process. All three should be considered in endeavours to improve handover. While we have not attempted to develop a detailed model of clinicians' situation awareness, and this is something that could be

explored in future work, it has been fruitful to view the purpose of the information sharing in the handover process from this perspective. Finally, the goal of improving information sharing in a non-effortful way across the handover process is driving our current work on investigating technology for the handover process.

ACKNOWLEDGMENTS

This work was funded by the UK Engineering and Physical Sciences Research Council (EPSRC) through the GHandI project (grant number EP/D078636/1). We would like to thank the staff in the clinical settings who have supported this work, as well as the patients who agreed to let us observe the handovers where they were discussed.

REFERENCES

- [1] ACSQHC. 2009. The OSSIE Guide to Clinical Handover Improvement. Sydney. Australian Commission on Safety and Quality in Health Care, ISBN 978-0-9806298-0-4.
- [2] Australian Medical Association. 2006. Safe Handover: Safe Patients. Guidance on clinical handover for clinicians and managers. <http://www.ama.com.au/web.nsf/doc/WEEN-6xFDKN>.
- [3] Behara, R. et al. 2006. A Conceptual Framework for Studying the Safety of Transitions in Emergency Care. *Advances in Patient Safety*, 2, 309–321.
- [4] British Medical Association, Junior Doctor's Committee. 2004. Safe handover: safe patients. http://www.saferhealthcare.org.uk/NR/rdonlyres/0F86311C-3B4B-4392-8FF6-4A5B7EE50D2D/0/Safe_Handover_Safe_Patients.pdf.
- [5] Cohen, M. D. and Hilligoss, P. B. 2009. Handoffs in Hospitals: A review of the literature on information exchange while transferring patient responsibility or control. Technical Report, University of Michigan, <http://hdl.handle.net/2027.42/61498>.
- [6] Endsley, M.R. 1995. Towards a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), 32–64.
- [7] The Healthcare Commission. 2009. Investigation into Mid Staffordshire NHS Foundation Trust. http://www.rcn.org.uk/__data/assets/pdf_file/0004/234976/Healthcare_Commission_report.pdf. Accessed 4th August 2009.
- [8] Jaggi, R., Kitch, B. T., Weinstein, D. F., Campbell, E. G., Hutter, M. and Weissman, J. S. 2005. Residents Report on Adverse Events and Their Causes. *Archives of Internal Medicine*, 165(22), 2607–2613.
- [9] Petersen, L. A., Brennan, T., O'Neil, A., Cook, E. and Lee, T. 1994. Does Housestaff Discontinuity of Care Increase the Risk for Preventable Adverse Events? *Annals of Internal Medicine*, 121(11), 866–872.
- [10] Randell, R., Woodward, P., Wilson, S. and Galliers, J. 2008. Public yet private: the status, durability and visibility of handover sheets. 21st IEEE International Symposium on Computer-Based Medical Systems, Jyväskylä, Finland.
- [11] Royal College of Physicians. 2008. A Clinician's Guide to Record Standards – Part 2: Standards for the structure and content of medical records and communications when patients are admitted to hospital. www.rcplondon.ac.uk/clinical-standards/hiu/medical-records

- [12] Tang, C. and Carpendale, S. 2007. An Observational Study on Information Flow during Nurses' Shift Change. CHI, San Jose, California, ACM.
- [13] Wears, R.L., Perry, S.J., Shapiro, M., Beach, C. Croskerry, P. and Behara, R. 2003. Shift Changes among Emergency Physicians: Best of Times, Worst of Times. Proc HFES 47th Annual Meeting. Pp. 1420–1423.
- [14] Wilson, S, Galliers, J. and Fone, J. 2005. Medical Handover: A Study and Implications for Information Technology. Healthcare Systems, Ergonomics and Patient Safety (HEPS 2005), Florence, Italy.
- [15] Wilson, S., Galliers, J. and Fone, J. 2006. Not All Sharing Is Equal: The Impact of a Large Display on Small Group Collaborative Work. Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, Banff, Alberta, Canada, ACM.
- [16] Wilson, S., Galliers, J. and Fone, J. 2007. Cognitive Artifacts in Support of Medical Shift Handover: An In Use, In Situ Evaluation. International Journal of Human-Computer Interaction, 22(1), 59–80.
- [17] Ye, K., Taylor, D. M., Knott, J., Dent, A. and MacBean, C. 2007. Handover in the emergency department: deficiencies and adverse effects. Emergency Medicine Australasia, 19(5), 433-41.

Design Guidelines for Negotiation Support Systems: An Expert Perspective Using Scenarios

Alina Pommeranz, Willem-Paul Brinkman, Pascal Wiggers,
Joost Broekens, Catholijn M. Jonker

Section Man-Machine Interaction, Delft University of Technology
Mekelweg 4, 2628CD, Delft,
The Netherlands

{a.pommeranz, w.p.brinkman, p.wiggers, j.broekens, c.m.jonker}@tudelft.nl

ABSTRACT

Negotiation support systems (NSS) can enhance humans' performance in negotiations. Much research in this area focuses on finding optimal bids. However, there is little research on human factors in technological negotiation support. We believe an in-depth analysis of the task involving experts and users is needed to build a new generation of NSS focusing on man-machine collaboration. We describe a scenario-based approach to gathering requirements for such a system. We wrote five scenarios containing part of the envisioned functionality in the most important use situations, e.g. face-to-face negotiation, on the phone, collaborative or mobile preparation. We used claims analysis to clarify our design decisions. To evaluate our claims we organized focus groups including six general and six job negotiation experts. The filmed scenarios were used together with two claims each to guide the discussion. Based on the data analysis we constructed 12 design guidelines for NSS.

Keywords

decision making and problem solving, decision aiding, human-system interaction, design approaches

ACM Classification Keywords

H.4.2 [Types of Systems] *Decision support.*

INTRODUCTION

Background

Existing Negotiation Support Systems (NSS) can enhance the human performance in negotiations and increase the number of win-win outcomes if the negotiation space is well-understood [9, 10]. This is because computers are good at coping with the computational complexity involved in calculating offers. However, there are a number of issues inherent in real life negotiations that are difficult to deal with using classical Artificial Intelligence (AI) approaches alone. These issues mostly relate to the interpretation of the social setting. Therefore, NSS are required that take into consideration the

strengths of both the machine and the human.

Current research on technological negotiation support is carried out in different areas, primarily in management science, electronic commerce and Artificial Intelligence [1, 14, 21]. It is hard to find studies in this area that include human factors [11], which is surprising since NSS are a type of interactive systems that offer rich possibilities for researching and designing human-computer interaction. However, different problems such as reaching optimal solutions and bids, formal descriptions of negotiations, the extraction of user preferences or problem representations remain the major foci of research efforts.

As formulated in [2, 19], NSS research concentrates on technological solutions, but the social problems that they intend to solve are secondary or even completely neglected. More in particular Swaab and colleagues [19] argue for a careful analysis of social and psychological processes in order to design good NSS. They claim that the success of an NSS is not only dependent on technological feasibility but also on the understanding of the activity that the system will support. These authors attempt to inform the design process of NSS.

However, they primarily look at two aspects that influence the outcomes of negotiations positively, namely common (cultural) identity and shared cognition. In this sense NSS can help by providing information to the opposing parties to establish a common understanding of the problem and possible solutions. Their studies show that the nature and representation of the information can influence negotiation outcomes. Another effort to emphasize the importance of social and also emotional issues in negotiation and their consideration for NSS has been made by Bui [2]. In his article the author points out problems that evolve from the fact that empirical research focuses only on the rational aspects of negotiation. For instance, the negotiation models that are implemented in NSS assuming strict economic rationalization ignore that people also take decisions based on social acceptability of different means to achieve a deal. Adding reasoning based on ethical and social norms to negotiation models will allow them to better represent the real life negotiation processes. Bui explores socio-emotional aspects such as conflict awareness, thoughts, emotions, intentions, trust and norms and their impact on negotiation. He creates a general list of aspects that NSS should help users with, such as identifying controversy, clarifying issues/criteria, equalizing parties or finding solutions and simulating impacts of potential decision. These can be seen as more generic guidelines for the functionality and design of NSS. Both cases [2, 19] refer to shared NSS used either collaboratively by all parties or as mediators. This is only one type of nss with special requirements.

An interesting research area where social aspects are actually considered is group decision support [e.g. 13]. However, also in those cases the focus is on collaboration and verbal

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

communication between the participants rather than aspects like thoughts, emotions, trust etc.

In this paper we focus on the challenges of developing a system that is used only by one party in the negotiation and takes into account the human factors occurring in negotiations. Our aim is to extract detailed guidelines for this negotiation setting that extend the generic guidelines presented in [1].

We explain our scenario-based approach for finding such design guidelines. The second section presents a thorough description of our qualitative method using scenarios, claims analysis, videos and expert focus groups. The data analysis is presented in the third section, followed by the results including the design guidelines in section four. Finally, we give a conclusion about our work in section five.

Research Goals

Our project goal is to build an NSS that supports one party in a dyad negotiation through all phases of the process (preparation, begin, analysis, bidding, closure). Since all negotiations differ and the domain of the negotiation has a major influence on the process, our project focusses on two example application domains: housing and job negotiations.

To achieve a good interaction between our system and the user it is essential that both explicitly share a generic task model. In order to implement such a model, we need to find out what task we will support and understand that task in depth. Furthermore, as pointed out in the background it is important to consider social, psychological and emotional aspects in the design of NSS. However, only a few researchers [2, 19] investigated such aspects and they focussed on broad social science concepts and their implications on negotiation. Therefore, the design guidelines that can be extracted from their research are generic. We believe additional in-depth analysis of negotiations is necessary to create specific guidelines that will inform the design process of the next generation of NSS. In addition to studying theoretical research from negotiation and social science literature, it is important to explore negotiation practice in the real world. We used a scenario-based design approach [6] involving general negotiation experts and job negotiation experts. Our primary research goal is to construct a number of design guidelines for NSS considering real life practice from an expert perspective. To reach that goal we have a sub-goal of understanding the users' task in context, as well as their behaviours and needs.

SCENARIO-BASED APPROACH

Overall we follow a user-centered design approach within this project. Our target users are people with different negotiation skill levels. As shown in the literature [20] most people are not very good negotiators although it is an almost daily task. Whereas users might not be able to explain themselves, negotiation experts and trainers have a good grasp of the common practice, mistakes and support that is needed. As pointed out by other researchers, a participatory design methodology making the user a co-designer from the beginning might not be sensible when the user knows little about the domain and is supposed to be taught about it by the software. We therefore adopted the informant design framework suggested by Scaife and his colleagues [15] that proposes to involve various participants at different stages in the design process. By this we can maximize their input and advance the development. The participants can be users as well as experts depending on the kind of data needed in certain design stages. At this early stage we will aim at gaining knowledge about

negotiation from domain experts. In order to get a structured overview of the situations and the ways our NSS should support users, we decided to organize a number of focus groups with experts. As will be explained in more detail below focus groups offer an interesting way to discuss first ideas due to their dynamic group element.

To be able to gather useful data we decided to guide the group discussions by using filmed scenarios of use situations of an NSS. Not only did we want to show our ideas about parts of the envisioned functionality but also get feedback on possible use contexts. Use contexts are determined by the different phases of a negotiation as well as the situation and conditions in which the negotiation takes place. A first brainstorming session with seven researchers of our project group helped to envision related functionality and situations in which different phases of the negotiation take place. This group of researchers is interdisciplinary consisting of HCI researchers with foci on psychology, emotions and user-system interaction as well as researchers from the fields of computer science and Artificial Intelligence. All researchers are familiar with classical negotiation literature. A first selection of feasible ideas was written separately on post-it notes which we clustered. Based on those clusters we generated, in close collaboration with a negotiation coach, five distinct use contexts that together cover all negotiation phases and types of system use (e.g. open, hidden etc.). The contexts are: face-to-face negotiation, remote negotiation (phone, internet etc.), preparing collaboratively for an upcoming negotiation and preparing for a negotiation with time constraints while being mobile (e.g. on the train). Each of these situations offers characteristics that influence the acceptability as well as the functionality of the NSS.

Scenarios

Scenarios are useful in the design process since they capture the consequences and trade-offs of designs [5]. The narrative nature of scenarios enables users, experts as well as designers to imagine the use situations and contexts of new or existing technology. For each of the five use contexts we wrote a scenario presented here in summary. Italic text is taken from the original texts of the scenarios. We chose to write two scenarios illustrating a job negotiation, two with real estate content and one about buying a car. We included one scenario set outside of our application domains in order to investigate how a completely different domain can influence the devices' role and functionality. All scenarios were checked by a professional negotiation coach to make sure that they were sufficiently realistic. Each scenario is briefly discussed below.

Mobile Preparation with Time Constraints. Preparation is one of the negotiation phases stressed in the literature, e.g. [8]. In this scenario we describe a preparation situation with special constraints. The job applicant Martin is already on his way to the interview. Therefore he has limited time to prepare himself. In addition, the mobile setting constitutes another constraint, namely limited resources. Both constraints require special regard when it comes to the functionality of the device. Just before getting on the train to the interview Martin has received a NSS on a pocket device. He uses the device's speed preparation function to prepare himself in the short time he has left. Among other functions the device allows him to receive knowledge about the job negotiation domain. *He wonders how much money he could ask for. He chooses 'expert opinion' on the interface and types in 'salary'. The PN suggests a website that has a forum where you can discuss current average salaries for IT consultants with an expert in the field. After reading through the forum Martin has a quite good idea what he can ask for*

with his kind of educational background and experience. With that knowledge he feels more secure and relieved.

Later in the scenario Martin makes use of the training module of the NSS which enables him to go through a simulated interview. He receives on-the-fly advice about his and the opponents' actions. The scenario ends with a more relaxed applicant, who knows what to expect in the upcoming negotiation.

Face-to-Face Negotiation, Secret Use. The situation described in this scenario is a negotiation between an employee, Bianca, and her boss. Bianca is using a pocket device with a NSS. She is hiding the fact that she has such support by telling her boss he is using her device to take notes.

Bianca has been working for a big telecommunication company in The Hague for 2 years now. Today her annual evaluation with her boss is due. Her boss is known as a quite friendly person, who hardly ever becomes aggressive or ill-tempered. However, he is very worried about his department's performance and likes to know exactly what his employees are doing. Bianca wants to take this meeting as an opportunity to re-negotiate some parts of her contract. Since her husband got a new job in another city, they decided to move further away. Therefore she wants to discuss with her boss about opportunities to handle the new situation. She knows that she worked hard and well in the last year and should get what she wants, but she does not consider herself a good negotiator. Therefore, she recently got the NSS and prepared herself for this negotiation with the device.

Throughout the negotiation described in the scenario Bianca receives help from the device. Several functions are described in this scenario including e.g. affect management, generating new options, and giving behavioral advice.

Bianca presses the button 'opponent concerned'. The NSS advises her to uncover the reasons for Mr. Smith's worries and show sympathy. Bianca asks: "May I ask you what your concern is?" Mr. Smith replies: "We always have a lot of spontaneous meetings to decide on how to proceed, which you will be missing if you were not here and since you are one of the main developers I think you should attend such meetings." "I really understand your worries, Mr. Smith. However, the welfare of my family is very important to me. But I am sure we can find a solution that considers both our concerns." The scenario ends with a deal in which both parties gain something and are satisfied with.

Collaborative Preparation. Negotiation involves a lot of emotions on both sides of the bargaining table, but also within a party, e.g. between a couple buying a house together. In this case the first step is to merge the demands and preferences of both partners before starting a negotiation with the opponent side. Our scenario describes a couple that is planning to buy a house together and uses the NSS during the preparation to sort out their preferences and to download domain knowledge about real estate. *They both sit close to each other on the sofa and look at the screen together. Mary starts the NSS and a virtual agent (VA) welcomes her. "What would you like to do?" he asks. Mary types in 'merge my partner's preferences with mine'. The 'collaborative preparation' module starts up. After a short introduction the NSS asks each of them to put in their preferences for a house separately. Since they also have the NSS software installed on their laptop they put in their preferences in parallel. From both preference profiles the NSS creates a matching profile and shows the clashes of their preferences. It advises the couple discussing the clashes and trying to find trade-offs between them that suit both. During this process of compromising the couple gets into a quarrel in which both*

insist on their own wishes without even communicating the underlying reasons in detail. In this case our device takes on a proactive role and interrupts the couple to give advice on how to handle the conflict. The NSS senses the noise and the angry voices in the room and assumes an argument. On the screen the VA appears and says "it became very loud in the room. Are you arguing?" Since the device does not get any attention a red LED starts blinking and a beep sounds. Both Mary and Piet stop talking and look at the NSS. Mary answers the NSS's question with yes. The NSS suggests calming down [...and...] prompts them to put in an emotional value on a scale from 'I don't care at all' to 'I would die for this' for each variable they have different preferences on."

After having sorted out all their preferences they start looking for houses. In the last scene of the scenario the couple visits a house and takes advantage of the NSS's feature of taking pictures and storing them together with other information about the house in a database.

Negotiation on the phone. A negotiation in which both parties are not situated in a face-to-face setting, but are distant from each other offers different design challenges for a NSS. First of all one party does not see the other party and therefore the use of a NSS can take place without each others' notice. Especially in real estate situations, e.g. when buying a house another aspect to consider is that the negotiation is split into a number of phone calls. This gives the user time in between the calls to use the system in each step of the negotiation. In our scenario a couple has decided to buy a house. Before the wife starts the negotiations with the real estate agent of the seller, the couple decides on a price. They use the NSS to download information about prices of similar objects in the same region to know what to expect. *Furthermore, the PN has downloaded housing domain knowledge, such as contracts and legal issues and the prices of similar houses in the neighborhood to take into account. Before Mary came to work this morning she had decided with Piet to set a first bid around 450.000 Euro.*

At work Mary calls the agent and starts negotiating. Before and during the phone calls she uses the NSS on her laptop to receive advice about different steps in the negotiation, e.g. *the NSS advises her to not start the negotiation with offering a price, but instead talk about other issues and options...*

The bidding goes on for a while and the NSS shows a visualization of the bids in the outcome space based on the preferences of Piet and Mary and the estimated preferences of the agent. After a while the NSS detects that the bidding is not reaching a win-win situation. After finding new variables to include in the negotiation to reach an agreement that suits both parties they finally close a deal.

Face-to-Face Negotiation, Open Use. We decided to include another scenario that has a face-to-face setting, but showing an open usage of the NSS. This scenario is about a couple buying a car. Our belief is that the car dealer's setting enables people to use the NSS more openly. When buying a car it is usually not necessary to stick to one specific car dealer. No long-term relationships need to be considered. Therefore, the couple in the scenario openly states that it will be using the device and explains what they can do with it. The focus of the scenario lies in the advice of time-outs at strategic points during the negotiation. During the process of looking at cars and refining their preferences for the new car, they enter information about the state of the negotiation into the NSS. They receive strategic advice on how to proceed and when to take the time to recapitulate.

He [the car dealer] shows them a range of more sporty looking family cars and the couple chooses their favorite. They enter that into the NSS. The NSS advises them to take a time-out and check whether they have considered all their preferences and whether all the information they need has been disclosed.

After they have found an interesting car the bidding starts in the car salesman's office. The NSS assists the couple by comparing prices with similar cars online. They disclose to the salesman that the market price is lower than his offer. The salesman drops his price. They negotiate about a few extras and finally leave with a new car and a deal they are satisfied with.

Storyboards and Videos

Due to their illustrative strength scenarios are a good means to communicate design ideas within the project team as well as to users or experts in the field. In order to exploit that strength even more we decided to visualize the scenarios. First we created a storyboard for each of the scenarios. For the collaborative preparation scenario see Figure 1. These storyboards then served as a basis for the shooting and editing of short (about two to three minutes) videos (for an example video see <http://mmi.tudelft.nl/video/scenario2/>). Using videos we were able to present the use contexts of our NSS very well. Much of the functionality of the NSS was kept open for interpretation to avoid limiting the discussion about the functionality. The videos were used in the design process as described in the focus group section. In the future they will also be used alongside a questionnaire on users' acceptability of an NSS in different use contexts.

Claims Analysis

Due to the scenarios' narrative nature many things are left implicit. Often causal facts and relations underlying the actions described are not revealed. Therefore it is useful to enumerate such causal relations separately. This can be done through claims analysis [5]. Each claim underlying a certain action or design feature in the scenario is listed together with its trade-offs. We used the claims slightly different, as proposed by Neerinx [12], namely in order to test our hypothesis about functionality and use contexts in the focus groups discussions with the experts. We wrote down four to six claims per

scenario based on our hypothesis. Due to space limitations we cannot list all the claims here, but only give examples. The first claim was written for the face-to-face scenario and the second for the negotiation on the phone scenario:

Advice claim: the NSS gives generic advice for different negotiation phases in a text-based form (e.g. ask for reason of concern, be sympathetic, and maintain the relationship).

+ Even though the user might know of such things due to a good preparation, the NSS's advice serves as a reminder during the negotiation process.

- The user might not be able to put the advice to practice or the way he tries to do so is not effective.

Graphical representation claim: the NSS shows the current status of the negotiation graphically including all variables etc.

+ The variables and their influences on the negotiation process are shown, so that the user can understand the process better.

+ The user can recapitulate and learn for future negotiations by looking at the current status and the influences of the variables.

- The number of variables and influences is high and the user finds it hard to learn from the graphical representation.

- The graphical representation is not understood by every type of user.

Focus Groups

Focus groups [16] have been widely used in marketing to exploit the dynamics of group discussions in order to receive attitudes towards ideas or products. Bruseberg and McDonagh-Philp [3] have shown that focus groups are also useful during the design process of new technologies. They help the participants to articulate their ideas and provide the researcher with inspiration for the design process. Lately, HCI researchers have adopted the method and refined the techniques used to stimulate the discussion. As for instance, Goodman and colleagues [7] found out, it is profitable to use visual help such as pictures and also scenarios in focus groups. Furthermore, tasks can start up a discussion. Based on these findings we used the previously described scenarios in form of videos in the focus groups.



Figure 1. Collaborative preparation scenario.

In total we had a number of 12 experts divided into three focus groups. We divided the experts into different focus groups according to their expertise. As explained by a number of researchers, e.g. [16], the homogeneity of the group plays an important role. The more similar the group members are the more likely they are to voice their opinions. Therefore, we formed one group with general negotiation experts, such as negotiation trainers, lawyers, a judge etc., and two with job negotiation experts, such as human resource employees and labour union representatives. In the beginning participants were introduced to each other and the project was described. Every participant received a questionnaire that contained two claims from the claims analysis (see previous section) per video. The claims, however, were reformulated into statements that allowed the experts to specify their level of agreement with. The two claims named in the previous section were presented as the following statements:

Statement: *General tips and strategic advice [e.g. try small talk, show sympathy for your opponents concerns] is more useful for the user than specific behavior- and decision-advice.*

Statement: *The NSS should focus on helping the user to understand the bidding process [e.g. graphical representation of the bidding including history of bidding] rather than proposing the next bid.*

After watching each video the participants individually specified their level of agreement on a 7-point Likert scale, and provided comments. We chose this method to give everyone a chance to think about their own attitudes and opinions in silence. As pointed out by e.g. Carey [4] less confident members may be encouraged to disclose more when having written down their views in advance. Once every member finished writing the moderator started a group discussion, by asking the participants in turn to react to the claims and discuss their ideas with the others. The moderator stimulated the discussion without enforcing any existing views from the project team. The discussion was taped for later analysis. In addition, two researchers in every group took notes. Taking notes is important since simple audio-recording cannot always capture what is happening between the members of the group.

DATA ANALYSIS

Our approach results in two types of data, i.e., data from the notes and data from the questionnaires. To analyze the questionnaire data (values on a Likert scale) we used a standard mean value calculation. Figure 2 presents the average level of agreement of the experts with the claims that were presented in the questionnaire. Considering the 95% confidence interval and the value four as the middle of the scale the results suggest that the majority of the experts leaned towards agreeing with the claims: 1) open use of the device when buying a car benefits the outcome; 2) device should help the user to understand the bidding rather than giving the next bid; 3) general tips are more useful than specific advice; 5) in preference elicitation ask for core concerns (instead of specific values); 6) short training and simulation enhances negotiation skills; and 7) short preparation

contribution positively to negotiation outcomes. The qualitative data explains the rationale behind these positions and provides additional ideas.

Focus groups provide large amounts of qualitative data, due to the dynamic nature of the group and the contextual setting. As discussed in detail in [4, 16] the data analysis of focus group data is delicate. Researchers have to be aware that focus groups are not meant to find consensus within the group and that empirical generalization from the data is not possible. However, according to Sim [16], the data from focus group can provide theoretical insights with sufficient level of universality to be projected to comparable contexts.

For the analysis of our data we used a method similar to interpretative phenomenological analysis [17], which is a bottom-up method often used in psychological qualitative research. The idea is to go through the data from one focus group to gather emerging themes from the text. Themes can be recurring ideas, thoughts or feelings from the participants. These themes are then clustered together and superordinate concepts might emerge. This process is repeated for the other focus groups and finally, the superordinate themes are compared and converged to final themes or theories, i.e. in our case transformed into design guidelines.

We analyzed the sessions separately on the basis of the notes by at least two researchers. The recordings from the sessions were only used in case the notes were not clear enough or incomplete. Every idea or attitude was written on a post-it note. Repeated ideas were not written down again, as we were not trying to get empirical generality and furthermore, in groups people tend to agree with or repeat thoughts and ideas.

To define the general themes that can be transformed into design guidelines four researchers independently clustered the post-it notes. We intentionally included one researcher unrelated to the project. Therefore, we could compare unbiased data with the data from the project researchers. Themes thus identified were then compared across all focus groups. This revealed that researchers used two arguments to categorize the themes. Either they considered the system's functionality or they looked at the phases in negotiations process. The system's functionality perspective led to four categories, namely negotiation tactics, usage of an NSS, information the NSS should provide, adaptivity of the NSS to the user, and the interaction with the interface. The negotiation perspective extracted categories for all negotiation phases, such as training, extracting preferences, context analysis, interaction with the opponent, and analysis of the bidding process. In particular, the participants emphasized that the device should motivate the user to prepare, as even a short preparation will be beneficial for the negotiation outcome. Furthermore, they expressed that the device should help people understand the bidding process instead of just proposing next bids. Note, that although the discussions were triggered by the statements and the filmed scenarios, they also gave insights that cannot be directly linked to the statements. New themes arose, e.g., the importance of context and the adaptivity of the system to the user. All themes fall into the categories resulting from the clustering.

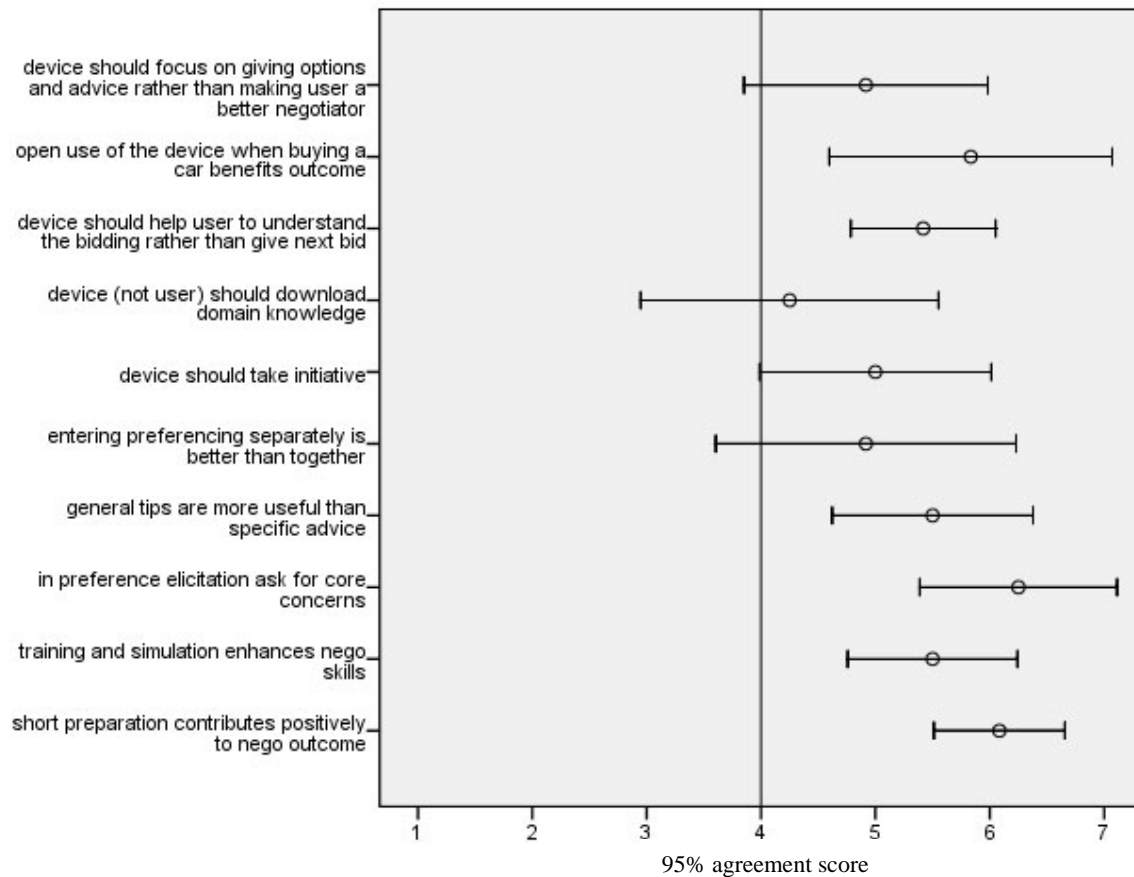


Figure 2. Mean values of agreement with claims (1 = strongly disagree, 7 = strongly agree) including 95% confidence interval.

FROM THEMES TO GUIDELINES

In the following we elaborate on the themes and construct eight design guidelines from the themes. Themes are presented in bold and guidelines in italics.

An NSS device adds higher value in the preparation and training phase than during a negotiation. Training needs to be interactive and the NSS needs to react intelligently.

All experts across the groups agreed on the fact that any preparation for a negotiation is useful. However, some experts mentioned that a technical device should add more value to the preparation than just reading a book on negotiation. They emphasized the importance of training and simulation and pointed out that the system needs to be able to respond to the user in an intelligent way. In detail, one idea that was mentioned was that the system needs to make people aware of what they can negotiate about. In addition, the system needs to ask questions to the user similar to the ones asked in job negotiations. In one group it was mentioned that multiple short sessions of preparation might be better than one long one.

1) An NSS should support interactive preparation sessions of different lengths.

2) The preparation module should have a simulation mode in which the user interacts with an intelligent negotiation agent.

In a face-to-face situation it is hard for the user to focus both on the device and the opponent.

Most experts were of the opinion that an NSS should not be used in face-to-face negotiations. Especially the job negotiation experts mentioned that the way the applicant or employee presents him/herself is important as well as focusing on the negotiation partner. While using a device the interaction with the opponent becomes awkward and might be embarrassing. Furthermore, the experts were concerned that understanding and processing the device's information and advice takes too much time and is too much cognitive load for the user in a face-to-face situation.

3) The cognitive load of the information representation provided by the nss during a face-to-face negotiation should be minimized.

The context including atmosphere, non-verbal communication and emotions plays a major role for the negotiation process.

In two focus groups it was emphasized that especially in job negotiations the non-verbal communication and the atmosphere in the room play an important role. Furthermore, emotions influence the decision-making process and the course of negotiation. This means that the system needs to be able to obtain this context information and take it into account when reasoning about next steps. People are generally better at interpreting emotions, non-verbal communication and atmosphere than computers. One way of enabling the system to understand the context is to build a context model within the system and let the user enter information about the context during the negotiation. To reduce the data that the user needs to feed into the system other techniques like emotion recognition or using (e.g. sound) sensors might be a solution.

4) *In the training module the user should be trained on being aware of the context.*

5) *Advice from an NSS should consider information about the context of the negotiation.*

The NSS is strong in the rational part of a negotiation, by offering new options and for storing and managing data. It should provide domain knowledge in terms of facts that the user can use to persuade.

Most experts agreed that the strength of a device would lie in handling the rational part of a negotiation. It can store and manage vast amounts of data, deal with the computational complexity during the bidding and offer new options to the user. Furthermore, domain knowledge should also include mainly facts, such as prices or salaries, which the user can use to persuade his/her opponent.

6) *An NSS should support the user by calculating bids and offering new options to negotiate on.*

7) *It should have a data storing and managing function that gives the user easy access to the information needed at a certain point in time.*

Both generic and specific advice is useful but needs to be applied carefully.

One of our claims was that generic advice is more useful than specific advice. The attitude towards this claim differed between the experts. Many of them saw a danger in specific advice because if the system cannot sense the context specific advice is often inappropriate. Generally both generic and specific advice could be useful but is dependent on the negotiation phase and the capabilities of system and user.

8) *An NSS should generally provide the user with more generic advice that the user can apply to the situation he/she is in.*

The NSS needs to adapt to the user's behavior and his knowledge or experience.

At several points in the discussion it was mentioned that the system advice or reactions need to be adapted to the experience of the user and his/her behavior. Regarding advice given by the system it was mentioned that novice users who are not good negotiators should get more specific advice whereas more advanced users are able to apply more generic advice. During the bidding the system should adapt its behavior to that of the user and recalculate the next bids in case the user changed his/her strategy.

9) *An NSS should be able to adapt to the user's skill level and experience and more in specific to the user's bidding behavior.*

10) *System advice should be based on the capabilities of the user to apply them in practice.*

Interruptions are seen controversial. Time-outs, however, are good.

The majority of the experts thought that active interruptions by the system through vibrating and beeping during a tense situation are not useful. The users would either ignore the system or become more upset. However, most experts agreed that time-outs are very useful for reflection of the negotiation process. As the user is not always aware of when to take a time-out the system should suggest it.

11) *An NSS should suggest time-outs at appropriate stages in the negotiation process.*

Preferences of collaborating partner's should be put in separately.

Across the focus groups there was a consensus that in the process of generating a preference profile for collaborating partner's, e.g. couples, they should put in their preferences separately. That avoids that one partner is more dominant than another. In our scenario we proposed that the system then merges the preferences and shows the clashes to the users. The experts did not agree on doing it this way. They pointed out that showing those clashes triggers arguments between the partners instead of a discussion about underlying values. It is more important that the partners talk about such values and come to a conclusion. The system could also directly suggest solutions. It was also proposed that a user indicates the importance of every preference.

12) *Partners should put in their preferences separately and assign an (emotional) value to each preference.*

CONCLUSION

Overall these guidelines boil down to the following insight: the preparation phase of a negotiation and the actual negotiation with an opponent require different interaction styles. In the preparation phase NSS should provide a negotiation training that is rich, content-full and contextual. Preferably it should make use of an adaptive scenario including socially intelligent opponents to provide a real setting. During the negotiation with an opponent, on the contrary, the system should provide concrete, personalized advice regarding offers and generic advice regarding the negotiation process with easy interpretable hints. The interaction style in this case should be as little interrupting as possible.

The major implication of these guidelines is that NSS need to have intelligence and reasoning capabilities in order to process the information entered by the users and give personalized output. Furthermore, the system needs to possess an accurate user model that is updated during the interaction to be able to adapt to the user. Furthermore, the interaction styles need to be carefully selected for each phase of the negotiation.

With regard to our approach we learned that the addition of video material as stimuli in focus groups facilitates idea generation and discussion within the group. Participants were able to directly reflect upon the potential usage of the NSS. The discussion was vivid and constructive. During the focus groups we got a detailed account of real life negotiations from the viewpoint of negotiation experts such as negotiation trainers, judges, labor union representatives and human resource employees. This enabled us to understand the task negotiators are facing and the mistakes people make. We learned what kind of support an NSS should give to its users and in which form.

A major drawback of making concrete stimulus material is that several experts also commented on and discussed particular implementations, while these were only included in the videos as examples and not as intended functionality. This happened even though experts were explicitly instructed not to pay attention to these details. We conclude that careful weighing is necessary regarding the amount of detail put into concrete stimulus material in order for a focus group to react upon the right level of abstraction.

In the future we will test users' acceptability of NSS in the different use contexts and conduct field studies in order to get a grasp of negotiation practice from a users' point of view.

ACKNOWLEDGMENTS

We would like to thank the project team as well as the negotiation experts for valuable input. This research is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs. It is part of the Pocket Negotiator project with grant number VICI-project 08075.

REFERENCES

- [1] Bellucci, E., Lodder, A. R. and Zeleznikow, J. 2004 Integrating artificial intelligence, argumentation and game theory to develop an online dispute resolution environment. *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*. Pp. 749–754.
- [2] Bui, T. 1994 Evaluating Negotiation Support Systems: A Conceptualization. In: *27th Annual Hawaii International Conference on System Sciences*. IEEE Press, Hawaii. Pp. 316–324.
- [3] Bruseberg, A. and McDonagh-Philp, D. 2002 Focus groups to support the industrial/product designer: a review based on current literature and designers' feedback. *Applied Ergonomics*, 33(1), 27–38.
- [4] Carey M. A. 1995 Comment: concerns in the analysis of focus group data. *Qualitative Health Research* 5, 487–495.
- [5] Carroll, J. M. 2000 *Making Use: Scenario-based Design of Human-Computer Interactions*. MIT Press, Cambridge.
- [6] Carroll, J. M. 2000 Five reasons for scenario-based design. *Interacting with Computers*, 13, 43–60.
- [7] Goodman, J., Dickinson, A. and Syme, A. 2004 Gathering Requirements for Mobile Devices using Focus Groups with Older People. To appear in *Designing a More Inclusive World*, *Proceedings of the 2nd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)*, Springer.
- [8] *Harvard Business Essentials: Negotiation*. Harvard Business School Publishing Corporation, Boston 2003.
- [9] Hindriks, K. V. and Jonker C. M. 2008 Creating Human-Machine Synergy in Negotiation Support Systems: Towards the Pocket Negotiator. In: *HuCom*. Delft. Pp. 47–55
- [10] Kersten, G. E. and Lo, G. 2003 Aspire: an integrated negotiation support system and software agents for e-business negotiation. *International Journal of Internet and Enterprise Management*, 1(3), 293–315.
- [11] Lee K. C., Kang I. and Kim, J. S. 2007 Exploring the user interface of negotiation support systems from the user acceptance perspective. *Computers in Human Behavior*, 23(1), 220–239.
- [12] Neerincx, M. A. 2003 Cognitive task load design: model, methods and examples. In: E. Hollnagel (Ed.). *Handbook of Cognitive Task Design*. Chapter 13. Mahwah, NJ: Lawrence Erlbaum Associates. Pp. 283–305.
- [13] Nunamaker, J. F., Briggs, R. O. and Mittleman, D. D. 1996. Lessons from a Decade of Group Support Systems. *Proceedings of the 29th Hawaii International Conference on System Sciences Volume 3: Collaboration Systems and Technology*, IEEE Computer Society.
- [14] Rangaswamy, A. and Shell, G. R. 1997. Using Computers to Realize Joint Gains in Negotiations: Toward an Electronic Bargaining Table. *Management Science*, 43(8), 1147–1163.
- [15] Scaife, M., Rogers, Y., Aldrich, F. and Davies, M. 1997. Designing for or designing with? Informant design for interactive learning environments. In: *Proceedings of CHI'97: Human Factors in Computing Systems*, ACM, New York. Pp. 343–350.
- [16] Sim, J. 2001 Collecting and analysing qualitative data: issues raised by the focus group. *Journal of Advanced Nursing*, 28(2), 345–352.
- [17] Smith, J. A. and Osborn, M. 2003. Interpretative phenomenological analysis. In: J. A. Smith (Ed.), *Qualitative Psychology: A Practical Guide to Methods*. London: Sage.
- [18] Stewart, D. and Shamdasani, P. 1990. *Focus Groups: Theory and Practice*. Sage.
- [19] Swaab, R., Postmes, T. and Neijens P. 2004. Negotiation Support Systems: Communication and Information as Antecedents of Negotiation Settlement. *International Negotiation* 9, 59–78.
- [20] Thompson, L. L. 2005. *The Heart and Mind of the Negotiator*. Pearson Prentice Hall, New Jersey.
- [21] Vetschera, R., Kersten, G. and Koeszegi, S. 2006. User Assessment of Internet-Based Negotiation Support Systems: An exploratory Study. *Journal of Organizational Computing and Electronic Commerce*, 16(2), 123–148.

Session 11: Cognitive Studies in HSI Design

Effects of Repetitive Practice on Interruption Costs: An Empirical Review and Theoretical Implications

Antti Oulasvirta

Helsinki Institute for Information Technology HIIT
Helsinki University of Technology TKK
P.O. Box 9800, FI-02015 TKK, FINLAND
antti.oulasvirta@hiit.fi

K. Anders Ericsson

Department of Psychology
Florida State University
Tallahassee, FL 32306-1051, USA
ericsson@psy.fsu.edu

ABSTRACT

It has been argued that a worker's ability to overcome interruptions depends on his or her level of expertise in the interrupted main task. The effects of repetitive practice (repeated experience) with new task have not been systematically analyzed. This paper reviews practice effects as reported from interruption experiments. The theory of long-term working memory (Ericsson and Kintsch, 1995) predicts that repetitive practice (experience) does not by itself lead to those changes in the structure of a novel memory-demanding task that are necessary for resisting interference. The review supports this prediction but also uncovers conditions in which repetitive practice produces a benefit, albeit limited, for interruption tolerance.

Keywords

interruptions, human-computer interaction, repetitive practice, long-term memory, working memory

ACM Classification Keywords

H.1.2 [User/Machine Systems] Human factors.

INTRODUCTION

An *interruption* is an “externally generated, randomly occurring, discrete event that breaks continuity of cognitive focus on a primary task” (Corragio, 1990, p. 19). It typically requires immediate attention and insists on action. These characteristics set interruptions apart from *distractions*, which are briefer and do not require full attention. Interruptions are a pervasive characteristic of human multitasking and are becoming a challenge in the information age.

Researchers have traditionally focused on *negative* effects of interruptions – ranging from increases in resumption time to decreases in accuracy and reduction in working memory (e.g., Altmann and Trafton, 2007; Cades et al., 2006; Czerwinski et al., 1991; Gillie and Broadbent, 1989; Hess and Detweiler, 1994; Hodgetts and Jones, 2006; McFarlane, 2002; Oulasvirta and Saariluoma, 2004; Trafton et al., 2003). We know much less about the factors that can reduce the effects of interruptions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

This paper examines how practice can improve a worker's ability to overcome interruptions. The most common experience at work and in leisure is *repetitive practice*. By repetitive practice we mean repetition of a task with no major intentional shifts in processing strategies, performance goals, or task materials. If a few hours of repetitive practice would prove sufficient for achieving reasonable *tolerance* to interruptions, there will be important implications for application in professional education. This result would imply that investing in costly improvements in organizational practices, training, and interfaces and task environments may not be necessary. If, on the other hand, repetitive practice does not produce sufficient improvements, such types of interventions must be explored.

The theory of *long-term working memory* (LTWM) (Ericsson and Kintsch, 1995) predicts that repetitive practice can improve interruption tolerance only if the properties of the task representation in *long-term memory* (LTM) change. Under random and unpredictable task conditions, repetitive practice would likely lead to only temporary and limited improvements. Under predictable task conditions, it should be possible for motivated participants to improve their anticipation, and the effects of practice would be more pronounced. However, in that case, the effects of repetitive practice should be eliminated when the task materials and other task characteristics are altered. Acquisition of memory skills that can exhibit transfer across sets of materials has been found to require extended periods of deliberate practice (Ericsson, 1985, 2006a; Ericsson and Kintsch, 1995).

To address these hypotheses, we review evidence from empirical studies on the effects of interruptions as a function of repetitive experience.

THEORETICAL APPROACH

According to the theory of LTWM, practice effects are mediated by more efficient use of LTM. Experts are known to use extensive semantic knowledge to achieve superior levels of performance. For example, it has been posited that chess grandmasters rely on 50,000 familiar chess patterns (“chunks”) in LTM to mediate their performance of chess-related tasks (Simon and Chase, 1973). In an alternative account, Ericsson and Kintsch (1995) have proposed that expert performers develop skills allowing them to encode domain-specific information meaningfully in LTM during representative activities. The experts' encoding processes allow the same information to be retrieved easily whenever it becomes relevant during subsequent processing of the task. Thus, experts not only acquire content knowledge and chunks; they also develop skills that enable them to efficiently apply their knowledge and manage their WM (Ericsson, 2006a; Ericsson and Staszewski, 1989). In working-memory-demanding tasks, such as planning and reasoning related to potential chess moves, skills for

encoding information with associations with sets of semantic and structural retrieval cues (retrieval structures) are developed. Selective activation of appropriate retrieval cues enables retrieval of the relevant information from LTM. These encoding processes and the retrieval cues are tightly integrated aspects of a given skilled performance and must be acquired in close connection with the development of other aspects of that performance. An important consequence is that the ability to protect information during interruptions depends on effective and retrievable encodings in LTM. Conversely, when people can resume tasks after memory-demanding interruptions without significant effects to the accuracy of continued performance would suggest storage in LTWM of information necessary for the completion of the interrupted main task. Figure 1 illustrates these ideas.

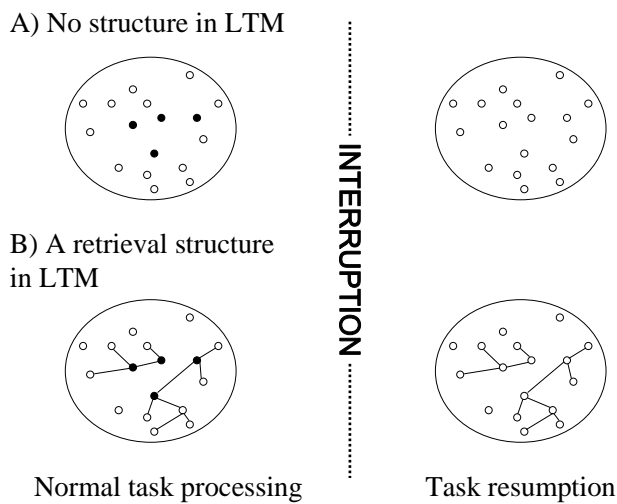


Figure 1. Memory contents before an interruption (left) and after (right). A) When inter-item associations in long-term memory (LTM) are weak or nonexistent at the end of the interruption, tasks cannot be resumed, because of interference. B) When associations with retrieval structures are established in LTM, it is possible to resume the task by reactivating the items previously available in working memory.

The key mechanism of LTWM is called a *retrieval structure*—an abstract, hierarchical knowledge representation built into LTM that mediates encoding, updating, and retrieval of intermediate products of cognitive processes carried out during skilled task performance. Certain encoding skills with acquired retrieval structures permit reliable storage of complex representations of task-related information for extended periods of time. Importantly for resumption of a task after an interruption, these representations can be accessed with the accuracy, reliability, and speed typical of short-term WM.

Encoding of relevant information with associations to a retrieval structure takes time, and thus by pacing participants at processing rate that exceeds participants' maximum speed would lead to interruption costs. Oulasvirta and Saariluoma (2006) tested this idea in the task of text comprehension, using expository texts. A 30-second multiplication verification task was utilized as the interruption task. In a set of three experiments, plenty of time was given for reading a piece of text (six seconds per sentence, or around 120 words/min), and no costs of interruption were found. In the fourth, critical experiment, the pace was raised to or possibly beyond the subjects' maximum

(2.5 seconds per sentence, around 290 words/min) – some researchers suggest that 250 words/min is the highest rate of reading with full comprehension (Carver, 1990). When the reading speed was increased, subsequent recognition memory was reduced significantly in the interruption condition in comparison to the uninterrupted condition, although the absolute difference in accuracy between conditions was small. Furthermore, the reduction in recognition memory was limited to the sentences studied *after* the interruption, indicating that participants had difficulties caused not by an interruption leading to *forgetting* the first part of sentences but by problems in *reinstating* the relevant information from the first sentence when reading the second. *Integrating* the second part of sentences with information about the first part was difficult either due to the insufficient time to reactivate retrieval cues and access the information, or due to inadequate encoding of the first sentences, or some combination.

How does practice produce interruption tolerance? People's primary goal in learning everyday activities such as driving a car, typing, or playing golf is to reach an acceptable level of performance. In the first phase of learning of everyday skills (Fitts and Posner, 1967; Anderson, 1982), beginners try to understand the activity and focus their attention on completing their attempts. During this *initial phase*, errors are perceptually salient and have immediate consequences, such as missing the tennis ball with the racquet and thus causing the end of the rally or being shot down in a video game. These frequent failures reduce the inherent fun of the activity, and often it is a parent or teacher who helps the beginner to succeed. With more experience (*the middle phase of learning*), gross mistakes become increasingly rare, sequences of performance steps are elicited more smoothly, and learners can elicit their performance without the need to actively deploy their attention to control their performance. After a limited period of training and experience – frequently under 50 hours for most recreational activities – learners attain an acceptable standard of performance, which can be executed with much reduced concentration. At this point, most individuals do not perceive a need for further improvements, which typically leads to a plateau where the same level of performance is maintained for months, years, and decades, requiring mere engagement in regular domain-related activities. In direct contrast, individuals aspiring to attain expert performance never allow their performance to be fully automated but continue to seek out, with the help of their teachers, new training activities wherein they need to engage in problem-solving to alter their cognitive representations and thus allow continued improvement of the mechanisms mediating performance. Some aspiring performers will eventually stop pushing their limits after longer periods of training. The performance level at the end of training will be eventually automated, and development will be prematurely arrested at that level.

Many skills activated in the normal course of a day have been acquired through experience and training, so one would expect some interruption tolerance in everyday activities. Skills acquired in educational systems, such as reading and algebra, and all sorts of other everyday skills, like driving and reading newspapers, have a very long experiential history for most adults. In these domains, however, most people hardly perform at expert and exceptional levels; they perform at a proficient level meeting their current needs. Many individuals seem satisfied with reaching a merely acceptable level of performance, as amateur tennis players and golfers generally are, and they attempt to reach such a level while minimizing the amount of effortful skill acquisition. Once an acceptable level has been reached, they need only maintain a stable performance, and often they

do so with minimal effort for years and even decades. As a consequence, researchers often have found that work experience in absolute terms has only a weak correlation with job performance beyond the first two years (McDaniel, Schmidt and Hunter, 1988). More generally, the amount of professional experience is not systematically related to superior performance (Ericsson and Lehmann, 1996). Recent reviews of the performance of doctors and nurses show that extended experience (beyond the first couple of years) after graduation from medical and nursing school is not associated with continued improvements but, instead, continued further experience is associated with decrements in objective performance (see Ericsson, 2004).

The activities of typical amateurs and many professionals contrast with those of future expert performers, who actively seek out challenging tasks for achieving an ever higher level of performance. This type of *deliberate practice* will depend on the training goals involved and on individuals' pre-existing skills in monitoring and controlling their performance (Ericsson, 2006b). For example, a chess player presented with the task of finding the best move for a specific chess position will engage in planning to find it. If unable to find the best move, the player will perform analysis to figure out why he or she did not find the best move, to avoid making similar mistakes in the future. Similarly, an athlete attempting to increase strength in a movement will repeat the movement with appropriate weights to reach exhaustion of associated muscles. The central assumption is that an individual's performance in a training task will vary as a function of focus of attention, type of strategy, and many other situational factors. If one wants to reach one's best performance consistently or even exceed one's highest current level, one has to identify the optimal conditions. Consequently, the performer needs to be fully rested at the start of the deliberate-practice activity if he or she is to maximize the gains in performance. The performer should also be fully prepared before initiating the practice activity, be given immediate feedback on the outcome, and then be allowed to repeat the task or perform a similar one with gradual modifications. Performing the practice task under these optimal conditions is much more effective for attaining higher performance than is performing a similar task when it is encountered, sometimes unexpectedly, within the natural context of performance. For example, imagine an amateur tennis player who misses a volley at the net. Play will continue until some time later when a similar situation emerges unexpectedly, with a similar problem for the player. Contrast this type of on-the-job learning with a session with a tennis coach. The coach would set up situations where the player would stand at the net and be ready to execute a volley. Upon mastery of the easy volleys, the coach can increase the difficulty of the shots and eventually embed volley shots into the rallies. It is essential to gradually embed the task in its natural context with regular time constraints and less predictable occurrence. It is easy to see that a few hours of this type of training would improve a player's volley skills more than would tens or hundreds of hours of regular tennis play against other amateurs.

Conditions of repetitive practice cannot be assumed to share these attributes. Importantly, training goals may be lower, the quality of feedback poorer, and possibilities (and motivations) for strategy improvement missing. More stable interruption tolerance would require qualitative changes in learning. Indeed, although experts are rarely assessed in interruption studies, their memory performance indicates virtually no cost of interruption. For example, Ericsson and Polson (1988) devised a laboratory analogue to ordering dinners in a restaurant to investigate the memory abilities of an expert waiter. They used a card with a

picture to represent each customer at a table, with each order being a random combination of meat entrees, level of meat doneness, salad dressings, and starches. Presentation was self-paced. Interestingly, they conducted surprise tests to examine how many of the orders earlier in a day at the restaurant were remembered. Expert waiter JC recalled information from 15 out of 18 tables, of which he correctly recalled 80%, incorrectly recalled 10%, and omitted 10%. That JC was able to recall virtually all customers for a given four-hour work period later, at the time of unexpected recall, is relevant but not necessarily evidence of what would happen during an interruption with the dinner ordering task. However, Ericsson and Polson argued that JC's ability to chat with customers during the period of them giving their dinner orders is evidence for robustness against interruptions. Ericsson and Staszewski (1989) have reported similar post-session performance for two students trained in mental calculation over 3–4 years.

More direct evidence for the hypothesis that only deliberate practice can lead to robustness against interruptions is found in a study in which bartenders of three levels of expertise were tested in a memory-demanding task similar to the order-taking of Ericsson and Polson. Beach (1993) compared novices, those at an intermediate level, and experts (teachers) in the task of drink-mixing. The bartenders were presented with a list of four drinks that they had to memorize before starting the mixing. The researchers interrupted mixing with a task of counting backwards. They found that interruption disrupted recent graduates but not experts, when measured in terms of drink errors. The author's explanation was that novices utilized *rehearsal* as their maintenance strategy. Beach (1993) also introduced a condition in the experiment wherein he replaced the normal glasses with opaque black glasses that were all the same shape. The unexpected finding was that those with moderate expertise suffered a 17-fold increase in post-interruption drink errors due to this change while no such effect was found for novices. Those at the intermediate level appeared to rely on, not on rehearsal, but available cues in retrieving the drinks upon task resumption, and removal of these cues led to a significant decrease in the accuracy of recall performance. By contrast, the recall of high-level experts was not degraded by the interruption or the changing of glasses, indicating that they were able to reinstate their retrieval cues without relying on those environmental cues.

The question remains, then, of the extent to which repetitive practice, which describes the early stages of a novice's learning, is sufficient for safeguarding information from interruptions and if there are differences among types of tasks. Our theoretical focus directs us toward domains wherein interruptions are relatively lengthy and demanding, such that they require reactivation of WM contents during resumption of the tasks. For example, driving or hammering does not necessarily meet these criteria, because interruptions are brief and these activities can be resumed with reliance on perceptual cues in the external environment. We therefore acknowledge that in a complex cognitive task, the effects of practice on a person's interruption tolerance are not solely due to improvements of memory (e.g., see Taatgen et al., 2008, on effects of practice on multitasking). Maintaining memory access is only one aspect of a highly integrated system that mediates superior task performance. Therefore, where necessary, we will extend our discussion.

REVIEW OF THE EVIDENCE

Only a minority of the papers on interruptions report on practice effects *within an experiment*. These experiments are

suitable for studying the effects of repetition, because experimenters have taken measures to ensure that the types of tasks are the same from one trial to another.

Two caveats are in order: First, in everyday activities, people are typically able to exhibit more control than in these experiments. Second, these experiments are typically only about one hour long and permit study of only the initial effects of repetitive practice.

We review evidence from five domains in which practice effects have been reported:

1. Military situation assessment
2. The Tower of London puzzle
3. Sentence production
4. The Gillie–Broadbent game
5. Mental calculation.

The practice in the studies reported upon counts as repetitive because the same task is repeated in the same conditions, with prior interruption being the only difference.

Military Situation Assessment

The military assessment task has involves monitoring several information windows presented by a PC and executing multiple actions to resolve a military situation. For example, an enemy attack is repelled by deploying a number of tanks in an appropriate manner. The interruption utilized in these studies is a two-choice classification task involving a radar screen. The main measurement used for interruption cost is the *resumption lag*: the time from completion of the interruption task to execution of the first subsequent command in the primary task. To assess the effect of interruption, this measurement can be compared to the mean *inter-action interval*, which denotes the time required for issuing a command under *uninterrupted* conditions.

In the first study reported upon with this task by Trafton et al. (2003), a practice effect was observed, but only for a group receiving no prior warning of the interruption (duration: 30 s). This group's resumption lag decreased from eight to four seconds across three 20-minute sessions. The other group, which had the prior warning (duration: 8 s) and used the time for rehearsing the main display before moving on to the interruption, showed no reliable improvement in resumption lag, which was consistently around 4 s across the three sessions. This difference in practice effects between two groups held even when improvements in baseline speed were taken into account. The authors report that 62% of the subjects' utterances during the interruption lag in the warning group concerned the goal state after the interruption (proactive preparation) and 38% referred to the state at the time of the interruption. In contrast, the no-warning group rarely rehearsed any states. Thus, a mere 20 minutes of practice almost halved the resumption lag and brought the no-warning group's resumption lag to the level achievable with 8 s prior warning.

In a more recent study of Altmann and Trafton (2007), three 20-minute blocks of this task were completed, each interrupted by the same radar task (duration: 42 s), after which the main screen was immediately restored. The cost of interruption was around 2.5 s. A decrease of 130 ms in interruption cost was found from the first to the third session, measured as the difference between resumption lags and inter-action intervals. This difference between sessions was not significant in a paired *t*-test, $t(374) = 1.59$, $p = .11$. (This test relied on the data that Altmann and Trafton have published on their website).

In the third experiment with this paradigm, Cades et al. (2006) examined the effects of practicing the main task by itself and practice with interruptions. This question is important for practitioners: If learning to resume a task is a consequence of experience with that task, training does not need to include interruptions. If, however, learning to resume a task is a consequence of practicing resumption itself, then the training need to mimic the interruptions with the task in everyday life. To address this question, three groups, with differing *interruption schedules*, were examined. The first group was exposed to interruptions only in the last of three sessions (OOX), the second group in the last two sessions (OXX), and the third group in all three (XXX). Twelve 30-second interruptions with the radar task occurred in each 20-min session. Across the three sessions, the mean inter-action interval decreased about 300 ms from the first to the last 20-min session, while full experience with interruptions (in the XXX group) led to a decrease of about one second in the resumption lag (from 5.2 s to 4.2 s). Also the group experiencing two sessions with interruptions (OXX) became quicker in resumption by the end of the final session. Importantly, the first group, which had practiced the main task, but not the interruption task, for the first two sessions (OOX), exhibited an interruption cost that did not significantly differ from the other groups' *first encounter* with an interruption.

Summing up these three experiments, the resumption lag decreased significantly already during the first 20 minutes of practice with interruptions (Trafton et al., 2003), but no reliable improvement was observed in the follow-up experiment (Altmann and Trafton, 2007). The implication of the study of Cades et al. (2006) is that the benefits achieved in Trafton's study were most likely not due to improvements in encoding the main task robustly to LTM. Had this been the case, practice of type OOX should have also decreased interruption costs in the third session.

What, then, explains this effect of practice? Our hypothesis at this point is that practice may enable *quicker response* to the re-presented perceptual situation. This hypothesis could be tested by including a condition in which the main task state is changed to something else after the interruption, in contrast to the interrupted state. If the resumption lags were equally short, one could infer that interruption costs are caused not by reactivation of memory but by the generation of a new response to the presented situation.

The Tower of London Puzzle

Hodgetts and Jones (2006) have studied a five-disk version of the Tower of London (ToL) task. In the task, participants are instructed to move rings from peg to peg – only one ring at a time – to achieve a final goal state with as few moves as possible. Participants were interrupted by a brief (~6 s) mood rating task that appeared at the time of the third move (out of six moves required to solve the ToL problem). After the interruption, the participants resumed the task and needed to move the next ring as soon as possible.

In the first experiment done by Hodgetts and Jones, the resumption cost was not significantly decreased after experience with the interruptions, an observation that could be attributed to the infrequent nature of the interruptions: Only six out of 25 trials included interruptions. When the interruption schedule was changed to 8/25 in a follow-up experiment (Exp. 3), a practice effect was observed. One can compare these two interruption frequencies to the setup in the experiment of Cades et al. (2006). In the 8/25 condition, the subject has more practice with task resumption.

Interestingly, in the latter experiment, the authors also manipulated the perceptual availability of the main task during the interruption. The observed practice-induced improvement in resumption times was uniform across these manipulations. Most likely, the subjects were fully concentrating on the brief mood rating task and did not pay attention to the main task that remained in the background.

Sentence Production

Swets (2006) interrupted sentence production with a brief (3–4 s) interruption that was either similar in nature (a verbal task) or dissimilar (an arithmetic task). The main task's stimuli involved arrays of three images depicting movement of colored objects above, below, or next to each other. The target sentences were two-clause items such as "The spoon moves above the box, and the bowl moves below the box." The timing of the interruption was manipulated: interruptions could occur during the production of the first clause, between clauses, or during the second clause. The images reappeared immediately upon task resumption. The results show that interruptions are most disruptive at earlier points in sentence production and are less disruptive when at clause boundaries, as measured by repetitions and incomplete sentences. The authors note, however, that repeating parts of a sentence does not necessarily reflect an inability to keep in mind the sentence but conversational conventions like repeating what was being said after being interrupted.

As an effect of practice, resumption lag dropped from 1218 ms in the first half of the experiment to 980 ms in the second. However, no effects of practice were found for variables that describe the quality of the sentences produced (such as repetition, skipping, and words in error). Because this experiment also involved repeated interruptions under conditions where the main task is simple and perceptually available upon task resumption, this practice effect may reflect similar speed-up in response speed as we hypothesized for Cades et al. (2006).

The Gillie-Broadbent Computer Game

The Gillie–Broadbent computer game is a task that places more weight on the role of WM for maintaining intermediate results in the main task during the interruption interval. In order to handle an interruption successfully, the participants must recall action items that need to be completed in the game after resumption. If this is not possible, they need to re-study the list of items to be memorized.

In the original experiment, Gillie and Broadbent's (1989) participants were first shown and asked to memorize a list of items that they would purchase by visiting shops in a virtual city (a map displayed by a PC). They were to pick up each item in the order prescribed by the list in a designated store. The map of the town consisted of 19 locations, and the list consisted of 10 items to be picked up. The task was self-paced. In a set of four experiments, Gillie and Broadbent were surprised to observe that interruptions as long as 2.75 minutes had no significant effect on completion times for the main task when the interrupting task consisted of an arithmetic task involving digits, but they found a significant effect when the interrupting tasks involved manipulations of letters and words, thus interfering with the main task content. When users have more time and degrees of freedom in encoding materials, it is more difficult to create disruptive interruptions. However, effects of practice were not examined.

In their critical follow-up experiment, Edwards and Gronlund (1998) studied the effects of practice on interruptions for the

Gillie–Broadbent task. Half of the 128 participants experienced an interruption with contents similar to those of the main task (map locations), and the other half was given a dissimilar interruption (a math task). To ensure that the interruption was not anticipated, only a single interruption was administered—during the sixth and final trial. This unique characteristic distinguishes Edwards and Gronlund's study from the previously examined studies, where interruptions were administered multiple times during the experiments.

The critical manipulation was related to the order of presentation of the shopping list: in the *fixed-order group*, the order in which items were picked up was always the same, across all six trials. In the *randomized-order group*, the order of items was randomly rearranged from trial to trial. Immediately after the interruption during the sixth trial, participants in both groups were queried regarding the *completed* subtask (e.g., picking up eggs) and the *unfinished* task (i.e., on what was being fetched when the interruption occurred).

The fixed-order group was significantly faster in retrieving all items, but there was no difference in the number of commands issued. The fixed-order group also correctly recalled significantly more items (50–53%) that still needed to be fetched than the randomized-order group did (22–34%). In other words, their memory of the interrupted content was better. Memory for the unfinished item in the randomized-order group was at chance level (22%) after a similar interruption, whereas the fixed-order group was clearly above this level (50% correct). The fixed group also spent less time studying the to-be-fetched items at the beginning of a trial and used a reminding function available in the game less often than the randomized-order group did, indicating that they were more efficient in encoding the task contents. For the last, interrupted trial, the randomized-order group (59%) was at the level of performance of the fixed-order group (63%) when the interruption was dissimilar to the main task (a math task) but not when it was similar (25%) (a map task), indicating greater vulnerability to retroactive interference in that condition. Moreover, memory for the unfinished item was consistently poorer in the randomized-order group (22–34%) than in the fixed-order group (50–53%).

In sum, the study of Edwards and Gronlund (1998) is the first to show that an effect of experience in the main task may be attributable to more efficient storage in LTM. The study indicates that the duration of a single experiment is enough to develop a representation that significantly improves tolerance to interruptions. The fixed-order group accomplished the task faster and had better memory for unfinished and completed items, and their memory was better after interruptions. When encoding differed from trial to trial (i.e., for the randomized-order group), the representation was more vulnerable to retroactive interference caused by similar interruptions. However, the benefit in the fixed-order group most likely depends on the predictable task and should be dramatically reduced during transfer to new task materials as well as varied orders of items to be collected.

Mental Calculation

The main challenge of mental calculation is that of remembering the operands of the problem and the intermediate results produced in the course of calculation. If memory for these intermediate items is lost during an interruption, the participant would have to start solving the problem from the beginning, which leads to high interruption costs.

Hess and Detweiler (1994) were the first to examine interruptions in this domain. The task set required mental calculation of a formula from pairs of numbers presented one pair at a time. In

the first experiment, participants experienced interruptions in all three sessions (XXX), while the second experiment was otherwise similar except that it introduced interruptions only in the last session (OOX). The experiments followed a 2 x 2 design. The subjects were interrupted after either the first number pair (low load) or the second pair (high-load condition) with one of two types of interruptions: a math memory task (remembering the result of a simple formula such as $X+Y$) and a memory task involving recall of two highly concrete nouns.

In the group that experienced consistent practice with interruptions (XXX), a practice effect was observed: an interruption cost of 0.30 in the absolute accuracy score in Session 2 for a demanding interruption was reduced in Session 3, showing a cost of a mere 0.05. In the last session, differences in accuracy among the different interruption conditions were no longer significant. In other words, the participants gained ability to handle an interruption occurring in the midst of calculation. However, in the group in which only the third and final session had interruptions (OOX), reliable negative effects of interruption type were found in the third session, paralleling the study of Cades et al. (2006). It seems that repetitive practice with the main task on its own is not sufficient for acquiring interruption tolerance in this task.

In a second paper utilizing the same task (Detweiler, Hess and Phelps, unpublished), the task was completed in either a varied-sequence or a consistent-sequence condition. For the varied-sequence group, each piece of information was presented in a different order across trials, whereas in the consistent-sequence condition the task information was presented in the same order. Again, a 25-second interruption occurred in step 2 or 3 (out of four steps in total). Participants in the varied group exhibited an increase in interruption cost (latency) of about 3 s to roughly 5 s between step 2 and 3, respectively. By contrast, subjects in the consistent-sequence group suffered a loss of only 3 s in both load conditions. In other words, reliable effects of practice were observed only in resumption lags, but not in accuracy as in the study of Hess and Detweiler (1994).

With these results taken together with those reviewed earlier, two effects of repetitive practice in memory-intensive tasks can be distinguished. First, practice in consistent task presentation conditions can lead to improvements in interruption tolerance in short period of time, whereas it is not clear if the same time is enough for interrupted performance to increase under inconsistent conditions. We suspect that since the effect emerges with a particular order of presentation but not when this order varies unpredictably from trial to trial, the effect is most likely not transferable to conditions with varied presentation. Second, practice with the interruption itself seems to be beneficial as well. The studies indicate that in conditions where interruptions occur more often, people improve their tolerance to them. This may be the result of one or more effects. For example, in tasks where reactivation can be done quickly on the basis of perceptually available information, such as ToL, it may be that this process of *generating* actions speeds up and there is no need to engage in recall of previously generated information. In more memory-intensive tasks, practicing the act of recall may establish a strategy for re-accessing retrieval cues upon resumption of the task (strategy learning), whereas when interruptions have not been encountered, people may struggle in their first attempt at resuming the task. Yet another hypothesis is that by practicing the main task in interrupted conditions, people learn to *anticipate* the memory demands posed by task resumption. This anticipation can lead to an attempt to encode retrieval cues or impose a consistent structure to encodings. We see some

evidence for the latter idea in a study that traced interruption costs in mental arithmetic over three days of practice.

In the study by Woelki and colleagues (2008), solving of arithmetic problem was studied with equations of varying levels of difficulty. Here, the elements were not presented to the participants one at a time as they were in Hess and Detweiler's paradigm; the intermediate steps had to be both generated and maintained in WM, placing more demands on WM during an interruption. An equation was presented on the display, and the result was to be typed via a keyboard. Occasionally, the task was interrupted by a 60 s color patch *n*-back task, where *n* was selected individually for each participant on the basis of a pre-test. After the interruption, the participant was asked to recall the equation that was being solved. In other words, this experiment measured effects of practice not directly but indirectly, from accuracy of recall.

In the study, six freshman-level students' development was traced for a total of nine hours over three consecutive days. Within each day, subjects' memory after an interruption improved toward the last session: recall of an equation after an interruption was 15% correct at the start of a day and had improved to 26% by the end of the day. But the improvement did not carry over to the next day; at the beginning of the following day, recall after an interruption was again poorer. Reaction times to memory probes exhibited an analogous pattern. In the final trial of each day, the subjects thought aloud during problem-solving, and they were interviewed on their solution strategies. The participants reported learning to solve the equations in a particular order of intermediate steps. Although there is no data from the beginning trials of a day to compare against, it is likely that the participants learned to impose their own structure for the solution (see also Detweiler et al., unpublished). The emergence of a consistent encoding strategy may explain why interruption costs decreased.

IMPLICATIONS

To conclude the paper, we sketch an explanation of how LTM mediates interrupted performance at different stages of repetitive practice.

Performing a novel task without prior experience. First, when initially encountering a novel task, people have limited resources to encode the information processed in the main task meaningfully. Their processes are varied, and pieces of information are likely to be lost unless actively rehearsed. In this initial stage, the participants' task representations are highly vulnerable to almost any kind of interruption that diverts attention from the main task. The level of activation of information still remaining upon resumption of the main task may be the primary mechanism mediating interruption tolerance. This would suggest that, at this stage, increased duration of the interruption would affect task resumption negatively. Also at this stage, costs are greater when interruptions occur at points of high memory load for the main task. Since the person's interruption tolerance is low, he or she may wish to resort to an *active maintenance strategy*. That is, during interruptions, one may try to rehearse contents relevant to the main task or might seek external markers and reminders (e.g., Trafton et al., 2003).

Repeated experience with the main task. Our review suggests that repetitive performance of the main task *alone* is not *always* sufficient to improve interruption tolerance. In the Tower of London (Hodgetts and Jones, 2006), the military assessment task (Cades et al., 2006), and the inhabitability index task (Hess and Detweiler, 1994), the observed gains in resumption times occurred in conditions where interruption of the main

task was encountered frequently but not when interruptions were rare.

Repetition under predictable task conditions. The results show that even a low number of repetitions in conditions where the main task contents are predictable and presented in the same order can significantly improve one's ability to complete the task after interruptions. Presenting equations (Detweiler et al., unpublished) or planning shopping for items (Edwards and Gronlund, 1998) in the same order facilitates the formation of a structure in LTM. In such conditions, interruptees are already less vulnerable to interruptions after only six or so trials. The data suggest that under these conditions people are less influenced by the similarity of the materials in the interruption and main task, or by whether the interruption occurred during high or low memory load. Most likely, this benefit of practice is very specific: the effect should disappear if the order of presentation or the nature of the task contents is suddenly altered.

Practice with the interruption task. The review indicates that repetitive practice of a main task *with frequent interruptions* improves speed in task resumption (Cades et al., 2006; Hess and Detweiler, 1994; Hodgetts and Jones, 2006; Swets, 2006). In some studies, the magnitude of this effect on resumption lag can be sizable. For example, in Trafton et al.'s (2003) study, resumption lags were halved from 8 s to 4 s with practice—but costs of 1–4 s were still reported with this paradigm, even in the last sessions of experiments. However, current evidence restricts to resumption lag; thus far has shown that practice *with interruptions* would be critical for improving also *accuracy* in the main task or memory for it after interruptions.

We are not sure what mediates this effect. At the moment, we have two hypotheses. First, it may be that trainees learn faster methods for reorienting to the display after an interruption. In a task that does not require carryover of intermediate results of processing, subjects who experience more interruptions also experience more situations where they have to regenerate an action from scratch. Second, repetitive practice may lead to learning to *anticipate the demands of task resumption*.

For example, Byrne and Bovair (1997) report an effect of practice on post-completion errors in a “transporter task.” The task requires completion of three sub-goals interactively, after which feedback is given on accomplishment of the goal. The participants were required to remember to do one final thing after this feedback, and they were trained and instructed in this requirement. Failing to do so manifested what the authors called a *post-completion error*. Interestingly, over the course of a session, the frequency of post-completion errors decreased from 65% to about 45%, which indicates that the participants increased their association of this subtask with the main task representation even when there were no perceptual cues to remind them.

Structure in the main task. The study of Woelki et al. (2008) tentatively suggests that with the extension of repetitive practice in a memory-intensive task subjects may learn to impose an order to processing in the main task and thereby improve their tolerance to interruption. Interruptions are still disruptive, and should be more so if task conditions change to familiar to unfamiliar. Woelki et al.'s (2008) data hint that the benefits gained at this stage may be short-lived and do not carry over beyond a day. It might be that more sustainable benefits may only be achieved when practice is more systematic.

Furthermore, evidence from Beach (1993) suggests that improvements in encoding of task-relevant information in LTM may preempt the need to engage in an active maintenance strategy such as rehearsal. In Beach's study, novices were

significantly hampered by interruptions, whereas more experienced bartenders were not. However, intermediate-level bartenders relied on environmental cues for task resumption, whereas higher-level experts were not dependent on either rehearsal or environmental support. The reduced use of rehearsal makes sense in view of the argument that active rehearsal is effortful and not very effective in tasks wherein the task-related information is complex and meaningfully organized (Oulasvirta and Saariluoma, 2006).

Data from studies of interruption costs when advancing even further to the middle phase of learning indicate that interruptions will still be somewhat costly. At this stage, costs of interruptions on memory are small in absolute terms—for example, 4–16% in memory accuracy in studies of reading comprehension reported by Oulasvirta and Saariluoma (2004, 2006). These costs will be more pronounced under conditions in which interrupting tasks are similar to the main task.

CONCLUSIONS

Our conclusion is that repetitive practice can bring about statistically significant improvements to the main task, especially when it takes place in conditions where 1) the main task offers a consistent structure for materials, 2) conditions involving interruption of the main task are practiced, and 3) controlled encoding strategies can be developed. However, the virtually perfect post-interruption or post-session recall seen in some experts (e.g., Charness, 1976; Ericsson and Polson, 1988; Ericsson and Staszewski, 1989) will not be achieved with repetitive practice alone. Further improvements may require deliberate efforts to develop effective encoding and retrieval of task-related information.

It is difficult to draw many conclusions for the training of workers from the current studies of interruptions in a laboratory setting. Many of the tasks are practiced for only one or two hours, and sometimes the accuracy in completing them is not perfect under uninterrupted conditions. By contrast, virtually perfect task completion would seem necessary for most professional work activities. When work-related tasks are essentially perfectly completed and the interruption costs only involve an increase in time to resume the main task, what would amount to a practically important increase in completion times? In both ToL and the military assessment task, the effects of interruptions were restricted to a few seconds in resumption time. In ToL (Hodgetts and Jones, 2006), no effects of interruptions on accuracy were found. Although errors in the military assessment task would be the most meaningful dependent variable, the task does not lend itself well to analysis of accuracy, because there are not clear right and wrong next steps or outcomes (Cades, personal communication, 2009). Altmann and Trafton (2007) report that error metrics were too difficult to define for that task, and overall game scores too variable for registering any effects of interruptions. Such interruption costs may correspond to a small fraction of work productivity if occurrences of interruptions are rare. In the future, it is important to focus on those types of interruptions in work environments that are costly in terms of productivity and errors and omissions causing accidents and injuries.

ACKNOWLEDGMENTS

Antti Oulasvirta's work was funded by the Emil Aaltonen Foundation.

REFERENCES

- [1] Altmann, E. M. and Trafton J. G. 2007. Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review*, 14(6), 1079–1084.
- [2] Anderson, J. R. 1982. Acquisition of cognitive skill. *Psychological Review*, 89(4), 369–406.
- [3] Beach, K. 1993. Becoming a bartender: The role of external memory cues in a work-directed educational activity. *Applied Cognitive Psychology*, 7(3).
- [4] Byrne, M. D. and Bovair, S. 1997. A working memory model of a common procedural error. *Cognitive Science*, 21(1), 31–61.
- [5] Cades, D. M., Trafton, J. G. and Boehm-Davis, D. A. 2006. Mitigating disruptions: Can resuming an interrupted task be trained. Proceedings of the 50th Annual Human Factors and Ergonomics Society Conference, San Francisco, CA.
- [6] Carver, R. P. 1990. *Reading Rate: A Review of Research and Theory*. San Diego: Academic Press.
- [7] Charness, N. 1976. Memory for chess positions: resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 641–653.
- [8] Coraggio L. 1990. Deleterious Effects of Intermittent Interruptions on the Task Performance of Knowledge Workers (A Laboratory Investigation). Doctoral dissertation, University of Arizona, College of Business and Public Administration
- [9] Czerwinski M., Chrisman S. and Schumacher B. 1991. The effects of warnings and display similarities on interruptions in multitasking environments, *ACM SIGCHI Bulletin*, 23, 38–39.
- [10] Detweiler, M. C., Hess, S. M. and Phelps, M. P. (unpublished). Interruptions and working memory. Unpublished technical report. The Pennsylvania State University, University Park, PA.
- [11] Ericsson, K. A. and Kintsch, W. 1995. Long-term working memory. *Psychological Review*, 102, 211–245.
- [12] Ericsson, K. A. and Lehmann, A. C. 1996. Expert and exceptional performance: Evidence for maximal adaptation to task constraints. *Annual Review of Psychology*, 47, 273–305.
- [13] Ericsson, K. A. and Polson, P. 1988. Experimental analysis of the mechanisms of memory skill. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 305–316.
- [14] Ericsson, K. A. and Staszewski, J. J. 1989. Skilled memory and expertise: Mechanisms of exceptional performance. In: D. Klahr and K. Kotovsky (Eds.). *Complex information processing: The impact of Herbert A. Simon*, Lawrence Erlbaum, Hillsdale, NJ. Pp. 235–267.
- [15] Ericsson, K. A. 1985. Memory skill. *Canadian Journal of Psychology*, 39, 188–231.
- [16] Ericsson, K. A. 2004. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, 79, S70–S81.
- [17] Ericsson, K. A. 2006. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative task. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance*. Cambridge, UK: Cambridge University Press. Pp. 223–242.
- [18] Ericsson, K. A. 2006b. The influence of experience and deliberate practice on the development of superior expert performance. In: K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance*. Cambridge, UK: Cambridge University Press. Pp. 685–706.
- [19] Fitts, P. M. and Posner, M. I. 1967. *Human performance*. Brooks/Cole Publishing Co.
- [20] Gillie, T. and Broadbent, D. 1989. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, 50, 243–250.
- [21] Hess, S. M. and Detweiler, M. C. 1994. Training to reduce the disruptive effects of interruptions. Proceedings of the 38th Human Factors and Ergonomics Society Annual Meeting. Pp. 1173–1177.
- [22] Hodgetts, H. M. and Jones D. M. 2006. Interruption of the tower of London task: Support for a goal-activation approach. *Journal of Experimental Psychology-General*, 135, 103–114.
- [23] McDaniel, M. A., Schmidt, F. L. and Hunter, J. E. 1988. Job experience correlates of job performance. *Journal of Applied Psychology*, 73, 327–330.
- [24] McFarlane D. C. 2002. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction, *Human-Computer Interaction*, 17, 63–139.
- [25] Oulasvirta, A. and Saariluoma, P. 2004. Long-term working memory and interrupting messages in human-computer interaction. *Behavior & Information Technology*, 23, 53–64.
- [26] Oulasvirta, A. and Saariluoma, P. 2006. Surviving task interruptions: Investigating the implications of long-term working memory. *International Journal of Human-Computer Studies*, 64, 941–961.
- [27] Simon, H. A. and Chase, W. G. 1973. Skill in chess. *American Scientist*, 61, 394–403.
- [28] Swets, B. 2006. Where was I? A psycholinguistic investigation of conversational interruptions. A doctoral thesis submitted for acceptance at the Michigan State University.
- [29] Taatgen, N. A., Huss, D., Dickison, D. and Anderson, J. R. 2008. The acquisition of robust and flexible cognitive skills. *Journal of Experimental Psychology: General*, 137, 548–565.
- [30] Trafton, J. G., Altmann, E. M., Brock, D. P. and Mintz, F. E. 2003. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58, 583–603.
- [31] Woelki, D., Oulasvirta, A., Kiefer, J. and Lischke, R. 2008. Practice effects on interruption tolerance in algebraic problem-solving. Proceedings of CogSci2008.

Occlusion Technique – Valid Metrics for Testing Visual Distraction?

Tuomo Kujala
Agora Human Technology Center
P.O. Box 35
University of Jyväskylä
+358 14 260 4660
tuomo.kujala@jyu.fi

ABSTRACT

The occlusion technique provides popular metrics for assessing tasks' visual demands and interruptability. In order to test the external validity of the occlusion technique for revealing distraction effects of visual in-vehicle information systems, the results of two driving simulation experiments were compared to those obtained with the occlusion technique. The effects of two different text types on time-sharing and driving performance metrics in simulated driving were compared to the results of the occlusion tests on the same visual secondary tasks. Besides the long task completion times, the occlusion data underestimated the distraction effects of the visual tasks found in simulated driving. Considering sensitivity for the potential distraction effects of display properties, task completion times occluded or unoccluded could not discriminate reliably between the two different text types. Time-sharing metrics could do this in the driving simulation already with a small number of participants. The R-ratio provided contradictory data on task interruptability compared to time-sharing in simulated driving with these types of tasks. The results suggest that the external validity of the occlusion metrics can be questioned. There are visual tasks that do not suit being tested with the occlusion technique as defined in the ISO standard. These tasks involve at least information-filled displays that provide drivers with more freedom in interaction styles than simpler manual and visual controls. The underestimations of visual distraction indicate that additional evaluations with time-sharing metrics, involving actual or simulated driving, can be highly recommended when assessing in-vehicle display prototypes with the occlusion technique.

Keywords

driver distraction, in-vehicle information systems, displays, visual distraction, interruptability, resumability, time-sharing

ACM Classification Keywords

H.1.2 [Models and Principles] User/Machine Systems – *human factors, human information processing*; H.5.2 [Information Interfaces and Presentation (e.g., HCI)] User Interfaces – *ergonomics, evaluation/methodology, screen design, theory and methods*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

INTRODUCTION

Information-filled displays of modern in-vehicle information systems (IVISs) and mobile devices can present severe safety risks in traffic due to driver distraction [20]. Because of the fast development of mobile technology and services, legislation is lagging far behind. These issues underline the need to find valid ways to test and develop the visual-manual controls and displays of these systems in order to find safer means of interaction while driving.

Lee, Regan and Young [13] define driver distraction as a breakdown of control in the driving task on multiple levels following Michon's [14] three-level cognitive control model. Above the operational level of a vehicle's lateral and longitudinal control, on a tactical level distraction is defined as the failure of task timing and on a strategic level it refers to inappropriate priority calibration between driving and secondary tasks [13]. Secondary task properties such as task interruptability, task resumability, task predictability, and task ignorability, relate closely to distraction on these higher levels of control [13]. In addition, drivers' lack of awareness of task demands and of their own capabilities is a factor that can contribute to distraction at these levels of control. The existing metrics of driver distraction are typically focused on measuring the effects of distraction on the level of operational control, such as deviations in lateral control [11], [13]. For this reason, there is an obvious demand for metrics that can provide information on drivers' capabilities to combine secondary tasks with the driving task on a level that does not necessarily interact directly with the measures of driving task performance. The question is, thus: how can distraction and the related secondary task attributes, such as task interruptability, be measured on the levels of tactical and strategic control?

The occlusion technique [3], [9], [22] is widely used and provides candidate metrics for assessing any secondary task's visual demand, and in particular, task interruptability. Task interruptability is not a well-defined concept. Yet, it should mean that the secondary task can be easily disengaged and resumed after interruptions, that are common while driving and necessary for enabling updates of relevant information in the driving scene [13]. Following the original technique of Senders, Kristofferson, Levison, Dietrich, and Ward [24], targeted in determining the visual demands of driving, the standardized occlusion procedure uses goggles or a screen to occlude participants' visual view to the visual IVIS under evaluation in periods of 1.5 seconds [9]. Task completion times are measured as Total Shutter Open Time (TSOT). The metrics include a measure for task interruptability, the Resumability (R)-ratio metric, which is calculated as TSOT/total task time unoccluded. R-ratios over 1 are assumed to indicate a cost of interrupting a secondary task while driving, and that the task or device may not be well-suited for use while driving because the task cannot be easily interrupted or resumed after the interruption [3].

Despite the merits of the occlusion technique of being fast, simple and cost-effective, the technique has been criticized due to its lack of dual-task condition and its insufficient metrics (e.g., [15], [16]). In particular, the occlusion technique may be limited because it does not involve simulation of the cognitive demand associated with visual sampling of the driving scene during occlusion periods (e.g., [15]). Further, the validity of the R-ratio for revealing differences in interruptability of different tasks has been questioned [6], [27]. Even though the reliability of the occlusion technique, i.e. the repeatability and consistency of the method, has been shown in a recent study to be at a satisfactory level [4], the external validity of the metrics can still be questioned [12], [15]. It seems, for example, that there are visual tasks for which the technique suits poorly [3], [6].

Time-sharing metrics are another candidate for assessing task interruptability and especially, driver distraction at the tactical level of control [11]. Time-sharing is defined here as allocation of visual attention in time between tasks. Time-sharing-metrics utilized in the context of driving provide information on the glance duration distributions towards a visual secondary task in *self-paced* task settings, and thus, on the total efficiency of the allocation of visual attention to an in-vehicle display [8], [11], [29], [30]. Very short glances at a display, as well as rare but significantly long glances in relation to the driving situation, can indicate inefficient search behaviors and interaction strategies. Thus, time-sharing metrics could provide information regarding distraction on the tactical level of dual-task control, and on task interruptability in dynamic and ecologically valid experimental conditions.

In this paper, time-sharing data obtained in two experiments with simulated driving are compared to data obtained with the occlusion technique. The visual secondary task of finding textual information on an in-vehicle display is the same in both cases. The text type on the display varied between participants, and the effects of the text type were evaluated with both metrics. The objective was to assess the sensitivity and validity of the occlusion metrics when dealing with in-vehicle tasks involving information-filled displays, when comparing the data to results obtained in task settings enabling higher external validity of the data due to the included driving task and the following dual-task condition [15]. Other objectives were to assess the scope of tasks of which visual distraction effects are suited to being evaluated with the occlusion technique, and to compare the sensitivity of the occlusion metrics to that of the time-sharing metrics in discriminating between display properties' effects on task interruptability.

We hypothesized according to the reviewed literature that the occlusion technique would underestimate the observed distraction effects of the reading tasks compared to driving simulation data. We expected significant effects of text types on time-sharing efficiency in simulated driving but that the occlusion data could not discriminate reliably in a similar manner between the distraction effects of the text types. These findings would mean that the selected visual secondary tasks represent tasks that are not suitable for evaluation with the occlusion technique.

TIME-SHARING EFFICIENCY IN SIMULATED DRIVING

In this section, we make a combined analysis of the results of two experiments on visual search tasks' distraction effects conducted in a driving simulation. Experiment 2 can be considered as a repetition of the first one, testing the reliability of the findings of Experiment 1 with small improvements in the experimental design explicated in detail below.

Research Method

Participants

A total of 34 volunteers (Experiment 1: 16, Experiment 2: 18) was recruited via public university e-mail lists. They included 18 women and 16 men between the ages of 20 and 34 (Exp. 1: $M = 24.3$, $SD = 3.9$; Exp. 2: $M = 27.0$, $SD = 3.0$). Aged drivers were excluded from the samples in order to control the effects of aging on time-sharing abilities [30]. All the participants had a valid driving license and lifetime driving experience of 2 to 500 thousand kilometers (Exp. 1: $M = 45$, $SD = 40$; Exp. 2: $M = 144$, $SD = 153$), and normal or corrected vision. Twenty-four of them were classified as experienced drivers ($> = 30,000$ km) and ten as novice drivers ($< 30,000$ km). All the participants in Experiment 2 were considered to be experienced drivers with 25,000 to 500,000 km life-time experience and driving on a weekly basis. This was to more carefully mitigate the known effects of low levels of driving experience on time-sharing efficiency [29]. The participants were randomly assigned to groups with different text types, although the groups were balanced over driving experience and gender. The experiments were conducted in Finnish with fluent Finnish-speakers.

Environment and Tools

The fixed-base driving simulation environment is located in the Agora User Psychology Laboratory at the University of Jyväskylä. The driving simulation is based on high-quality open-source car simulation software (www.racer.nl). All trials were driven in a simulated Ford Focus on a road-like environment simulating the Polish countryside. The road included straight parts and bends with varying curvature. The width of the road and other possible factors affecting the position of the car were fixed. A simulated racetrack was used for practice. The projected driving scene included a speedometer and tachometer just above the steering wheel. The 17" secondary task display was located 20 centimeters below the driving view and over 45° from the normal sight axis, on the right side of the participant (see Figure 1). Other research equipment included a helmet-mounted iView X HED eye-tracking system with a 50Hz sampling rate, A/V capturing devices, and a computer for controlling the secondary tasks.



Figure 1. The position of the secondary task display in relation to the driving scene and participant.

Design and Procedure

The procedure consisted of a single driving task and a driving task with a series of visual secondary tasks. In the secondary tasks, the participants orally answered questions located in the upper part of the display by searching the right sentence or part of it from the text chapter below the question. The experimental design was a within-subject design over dual-task condition (driving task and driving task with dual-tasking) and a

between-subjects design over text type (two different text types in the dual-task condition). A between-subjects design was selected in order to see how the different text types support the development of interaction during first-time contact with this type of dual-task situation through several task repetitions.

The text chapters consisted of a spaced or compressed text (see Figure 2). The spaced text had an empty line between the sentences while the compressed text was comprised of a single paragraph. The discriminability of the text was selected as an independent variable, because it was assumed that the spaced text would support more interruptible and resumable reading strategies which should be visible in the time-sharing data. E.g., the sentence the participant checked last was assumed to be faster to identify after an interruption with the spaced text. The text and the related question changed after each correct answer. The participants were instructed to try again if the task did not change after their answer (i.e. in the case of the wrong answer). The task design was intended to imitate situations in which the driver is reading a newsfeed or an e-mail message with an in-car Internet system while driving. One example of the questions was: "How often do disturbances caused by passengers occur during domestic flights?". The text chapter in this task was news about disturbance on a flight, and at the end of the fifth sentence, an officer tells that similar cases happen a few times a year. There was no time pressure in completing the secondary tasks and the tasks were thus self-paced, but the trial lasted as long as the participant took to complete the secondary tasks. In Experiment 1 there were a total of five secondary tasks, while in Experiment 2 there were total of four secondary tasks designed to be similar to those in the first experiment.

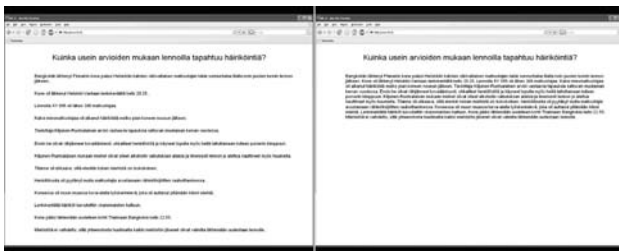


Figure 2. Secondary task display with the two alternative text types, spaced text (left) and compressed text.

At the beginning of the experiments, after the collection of general background information, a practice at driving for about 5 minutes on the racetrack was provided to the participants. After the practice, the participants completed the trials with and without secondary tasks on a same road. In Experiment 1 the order of these trials was counterbalanced to mitigate learning effects on the driving performance data. The order of the trials was not counterbalanced in Experiment 2, because the effects of dual-task condition in driving performance between the trials was not our main interest. Moreover, we wanted to make the dual-task condition as similar as possible for every participant in order to get further support for the outcome of Experiment 1. Driving speed was kept within 40–60 km/h by task instructions. To make the participants prioritize the driving task, they were informed that the 10 most accurate participants in the driving task would be rewarded with movie tickets. Accuracy was defined as total time spent outside own lane and below 40 km/h or above 60 km/h. Experiment 2 featured other upcoming traffic on the road in the form of four cars at preset points on the road. The traffic was included for further validation of visual behavior due to the possibility of unexpected events. Before the dual-task trial each participant was given one secondary practice task without driving. The dual-task trial was

designed to last around 10 minutes depending on the participant's performance. Finally, the participants were interviewed about their time-sharing strategies on the dual-task trial and the found strategies were classified according to common elements. The main questions were: "How did you perform the search tasks; did you have or did you develop certain ways of searching for the answers during the trial?"; and "Would you consider reading these types of news or messages while driving?".

Variables and Analysis

Independent variables for analysis included the dual-task condition and the text type. Among the dependent variables of driving performance were the total frequency (i.e. number) and total duration of lane excursions, which were defined as occurring when the visible part of the car's bonnet was over the lane markings for measuring safety-related deviations in lateral position. Dependent variables of time-sharing efficiency included standard deviations and maximum lengths of glance durations, as well as the frequency of glances longer than 2 seconds, in total and while driving in curves. The metrics of time-sharing efficiency were intended to measure participants' abilities to combine the visual secondary tasks with the driving task in a safe way on the level of tactical control. Glance durations of more than 2 seconds can be considered unsafe in many circumstances ([10], see also [8] and [29]). Frequency of over-2-second glances while driving in curves measured the participant's abilities to assess the visual demands of the driving situation and the extent to which they were able to adapt their task switching according to this information. In Experiment 1, the scoring of the over-2-second glances in curves was done manually, while in Experiment 2 this was done automatically by a script comparing the steering wheel data to the synchronized eye-tracking data.

The controlled variables included driving experience and gender, which were balanced between the groups in both experiments. Display properties other than the one varied, such as font (Arial, 12 pt), line spacing (single) and text starting locations, were fixed between the different display designs. In addition, the location of the answer varied between tasks within text between the first and last lines. Secondary task starting points varied with the participant's performance.

Video material of 25 frames per second from the eye-tracking system and the driving scene was scored frame-by-frame for lane excursions and glance durations with advanced video scoring software for behavioral research. A glance at the secondary task display was scored to begin at the frame the participant's gaze was off the road scene and to end at the frame with the gaze back in the road scene, following the SAE J2396 definition [21]. Other collected data included audio recordings from the interviews.

Repeated measures ANOVA and two-tailed *t* tests were used in order to find statistical significance and interaction effects in the results. Effect sizes were estimated using standardized differences in means and standard errors in the analysis of the combined results of the two experiments. Z-test was used in order to find statistical significance in the combined results. Alpha level of .05 was used in statistical testing.

Results and discussion

In Experiment 1, the dual-task condition had a significant increasing effect on the mean frequency of lane excursions, from 3.38 to 12.63, $F(1, 14) = 19.38$, $p = 0.001$, and on the mean total duration of lane excursions, from 4.50 to 20.66 seconds, $F(1, 14) = 8.62$, $p = 0.011$. However, the text type did

not have a significant effect on the frequency or duration of lane excursions. No interaction effects were found.

The text type had a significant effect on the time-sharing measures (see Table 1). The Compressed-group had significantly greater maximums, $t(14) = 2.66, p = 0.030$, and standard deviations, $t(14) = .36, p = 0.010$ of glance durations at the text, when compared to the Spaced-group. They also made significantly more over-2-second glances at the display in total, $t(14) = 2.93, p = 0.015$, and while driving in curves, $t(14) = 2.97, p = 0.017$, than the Spaced-group. There were no significant effects of text type on total glance times (spaced: $M = 209.92, SD = 28.24$; compressed: $M = 249.82, SD = 35.99$) or total frequency of glances (spaced: $M = 230.50, SD = 30.08$; compressed: $M = 174.50, SD = 22.72$) at the display.

Table 1. Mean values of time-sharing measures (standard errors) in Experiment 1.

	Spaced text (n = 8)	Compressed text (n = 8)
Standard deviation of glance durations (at the display) (s)**	0.59 (0.05)	1.33 (0.21)
Max duration (s)*	3.47 (0.44)	8.89 (1.99)
Frequency of >2s glances*	13.38 (3.88)	38.75 (7.74)
Frequency of >2s glances in curves*	2.00 (0.68)	8.88 (2.22)

Note. *: significant difference at .05 level, **: significant difference at .01 level.

Also in Experiment 2 the dual-task condition had a significant increasing effect on the mean frequency of lane excursions committed, from 2.61 to 6.11, $F(1, 16) = 10.69, p = 0.005$. Significant differences were not found in the frequency or durations of lane excursions across the different text types.

The time-sharing data revealed that the Compressed-group had again significantly longer maximum glance durations at the display, $t(16) = 2.39, p = 0.041$ (see Table 2). In addition, the frequency of over-2-second glances in total, $t(16) = 2.30, p = 0.039$, as well as while driving in curves, $t(16) = 2.25, p = 0.048$, were significantly larger in the Compressed-group. The distributions of glance durations in Figure 3 illustrate the effects of the text types. There were no significant effects of text type on total frequency of glances (spaced: $M = 116.44, SD = 9.38$; compressed: $M = 139.67, SD = 9.70$) at the display, but the total glance times were this time significantly lower for the spaced ($M = 157.08, SD = 8.61$) than for the compressed text type ($M = 255.08, SD = 32.44$), $t(16) = 2.92, p = .017$.

Table 2. Mean values of time-sharing measures (standard errors) in the experiment 2.

	Spaced text (n = 9)	Compressed text (n = 9)
Standard deviation of glance durations (s)	0.49 (0.06)	0.97 (0.23)
Max duration (s)*	2.97 (0.34)	6.15 (1.29)
Frequency of >2s glances*	13.33 (5.15)	36.56 (8.68)
Frequency of >2s glances in curves*	2.33 (1.64)	12.67 (4.30)

Note. *: significant difference at .05 level.

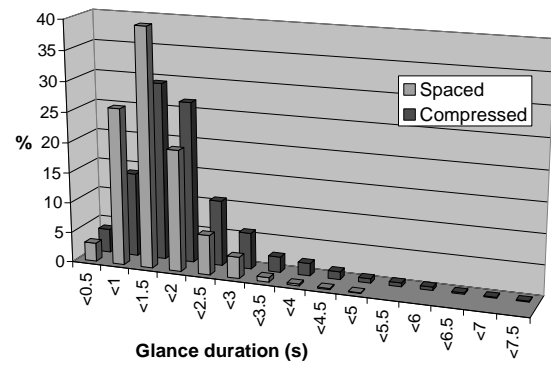


Figure 3. Glance duration distributions by text type in Experiment 2 (N = 18).

The combined results of the two experiments indicate that the spaced text type had a significantly positive effect on time-sharing efficiency with all time-sharing metrics compared to the compressed text type. Analysis of the combined effect sizes in time-sharing metrics is presented in Table 3. Smaller effect sizes of Experiment 2 may possibly be explained with the added oncoming traffic to the environment, and the resultant more careful visual behavior of the participants in the Compressed-group compared to the Experiment 1. The effects of the text type on time-sharing are still highly similar and parallel in the both experiments.

Table 3. Combined effect sizes in Experiments 1 and 2

Compr. – Spaced	Effect Size π (95% CI) Exp 1	Effect Size (95% CI) Exp 2	Combined Effect Size (95% CI)	Z	p
Standard deviation of glance durations	1.68 (.54, 2.82)	.95 (-.02, 1.93)	1.26 (.52, 2.00)	3.33	.001
Max glance durations	1.33 (.25, 2.41)	1.13 (.13, 2.12)	1.22 (.49, 1.95)	3.26	.001
Over-2-second glances	1.47 (.36, 2.57)	1.08 (.09, 2.07)	1.25 (.52, 1.99)	3.33	.001
Over-2-second glances in curves	1.48 (.38, 2.59)	1.06 (.07, 2.04)	1.25 (.51, 1.98)	3.32	.001

Note. π : Standardized differences in means.

However, it should be noted that the time-sharing behavior was not at a risk-free level in either group, because there were considerable amounts of over-2-second glances in tasks with both text types. The frequencies of over-2-second glances in both groups were far above the frequencies observed in a study with a similar design assessing Point-of-Interest search tasks with mobile navigation software [11].

Analysis of the effect sizes of the dual-task condition in the driving performance metrics revealed that the dual-task condition had a significantly negative impact in both groups in frequency and total duration of lane excursions. The combined

effect sizes were 0.92 (CI = .52, 1.32; $Z = 4.47$; $p = .000$) for the frequency and 0.55 (CI = .19, .91; $Z = 2.97$; $p = .003$) for the total duration of lane excursions. The smaller effect sizes in Experiment 2 (.41; CI = -.07, .89) for the duration of lane excursions compared to Experiment 1 (.74; CI = .19, 1.29) can possibly be explained by the controlled learning effects in Experiment 1.

There were no significant effects of text type in either measures of driving performance, contrary to the metrics of time-sharing. This finding indicates that lapses in visual behavior on a tactical level are not necessarily in direct relation with driving performance on the operational level of control (see [13]). This does not mean that lapses on the tactical level would be risk-free if they do not lead to decreases in driving performance. A possible explanation for the non-significant differences in the average lane maintenance measures could be that mean glance durations stayed on a relatively safe and typical level [28] with both text types in both experiments; the overall average was 1.43 seconds ($SD = .55$; spaced: $M = 1.20$, $SD = 0.41$; compressed: $M = 1.66$, $SD = 0.58$). This is inline with Wierwille's [28] visual sampling model, which can be taken as an indication for the similar level of the simulated driving tasks' visual demand compared to real driving. The mean total task time for the dual-task trials was 9.34 minutes ($SD = 3.23$).

The sensitivity of the time-sharing metrics for discriminating between potentially unsafe display properties from safer properties can be assessed to be at a satisfactory level. The difference between the text types for the favor of the spaced text type was statistically significant ($p < .050$) already with less than 20 participants in both experiments. This is in line with the findings of Horrey and Wickens [8], who concluded that the frequency and percent of over-long glance durations are more sensitive and reliable than the established glance measures of visual load (such as total and average glance durations) for revealing distraction effects of visual secondary tasks.

OCCCLUSION

The occlusion experiment was conducted following the parameters of the ISO occlusion standard 16673 [9]. However, the 20 percent quota of elderly participants was excluded in order to gain data that was comparable to that of the driving simulation experiments with the participants of ages 20 to 34. In addition, the R-ratios were calculated between subjects as explained in detail below.

We expected that the occlusion technique would not provide us with the same information on task interruptability, and on the suitability of the visual secondary tasks to be engaged while driving, that the time-sharing metrics in simulated driving could. In addition, we hypothesized that the occlusion technique could not discriminate reliably between the distraction effects of the text types the same way as in the driving simulation experiments.

Research Method

Participants

Twenty student volunteers were recruited via public university e-mail lists. They included 12 women and 8 men between the ages of 20 and 29 ($M = 23.1$; $SD = 2.6$). They all had a valid driving license and normal or corrected vision. The experiment was conducted in Finnish with fluent Finnish-speakers. The participants were randomly selected into two groups. However, pairs of participants between groups were matched for gender and age (within 2 years). There were 6 female and 4 male

participants in both groups, and their average ages were 23 years ($SD = 2.7$) in the Spaced-group and 23.2 ($SD = 2.5$) in the Compressed-group.

Environment and Tools

The experiment was conducted in the same test environment comprising a vehicle buck as the two driving simulation experiments, but without the driving task. The possible affecting factors, such as lighting in the facility, position of the display in relation to the participant, and display designs, were the same as in the simulation experiments. As a means of occlusion, the 17" display was occluded with a white blank screen in 1.5 second periods after every 1.5 second vision interval in the occluded tasks. Occlusion intervals were system-paced with the aid of the IrfanView presentation software's slideshow functionality. Literature shows similar results for goggles and other means of occlusion [3]. The helmet-mounted eye-tracking system was used for enabling possible later detailed gaze path analyses. The display was recorded with a video camera attached to a laptop.

Design and Procedure

After signing the consent form and the eye-tracking system calibration, a participant was instructed and trained with four tasks, two occluded and two unoccluded. The task was to find the correct answer to the question displayed from the text chapter below, exactly in the same manner as in the driving simulation experiments. The participant was instructed to try again after providing a wrong answer until finding the correct answer. After training and repeated task instructions the participant completed five occluded and five unoccluded tasks in turns. A between-subjects design was utilized to enable comparison between exactly the same tasks with the different text types. Every participant in the Spaced-group completed five occluded tasks with the spaced text type and five tasks unoccluded with the compressed text type, and vice versa in the Compressed-group. This was to ensure the R-ratios were calculated for the exact same tasks occluded and unoccluded between the age- and gender-matched pairs of participants. Due to the nature of the reading tasks, it was impossible to give an exact same task to a participant in the occluded and in the unoccluded condition (see Figure 4 for illustration of the experimental design and procedure). Five different orders for the total of 10 different tasks were randomized to mitigate possible order effects. Six of the tasks were exactly the same as in the driving simulation experiments, the other four were designed to be similar. Finally, the participants were interviewed for their search strategies similarly as the participants in the driving simulation experiments, and the found strategies were classified.

Variables and Analysis

Twenty-five frames per second video from the display was scored frame-by-frame for task times. Tasks were scored from the first frame the text was visible to the frame the participant started to read the correct answer. Wrong answers were included in task times. For calculating TSOT, the occlusion intervals were subtracted from the total task times. Two-tailed t tests and the alpha level of .05 were used in statistical testing.

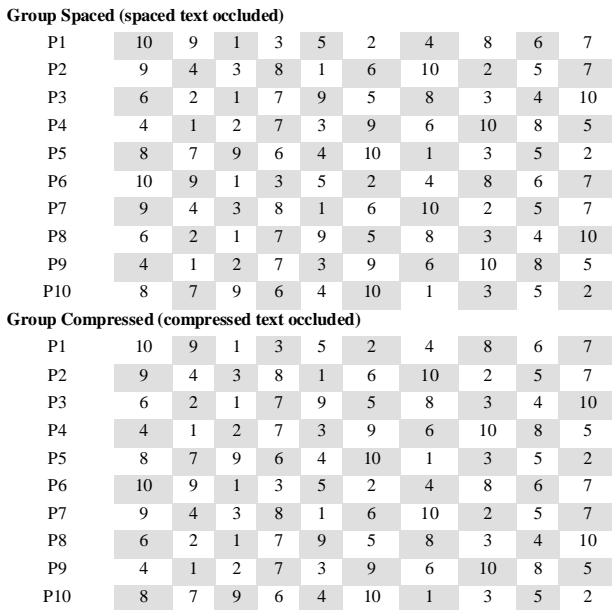


Figure 4. Task orders in the two groups. Note. Occluded tasks are represented in grey. R-ratios were calculated between the pair-matched participants, e.g., between P1s.

Results and Discussion

A qualitative dimension of the occlusion technique suggests that if the participants are not able to complete a task under occlusion or make excessive amount of errors, then it is a strong indication that the task is not compatible with driving and should be redesigned [3]. In this experiment, 199 tasks out of 200 were successfully completed. The one uncompleted task took over five minutes because of a misunderstanding and was interrupted by the experimenter. The task was not included in the results.

Average task completion times (TSOT) were 31.26 seconds (SD = 10.76) for the spaced text type and 34.33 seconds (SD = 9.76) for the compressed text type. Total task times unoccluded (TTTUnoccl) were longer with both text types than TSOTs; for spaced text 39.51 seconds (SD = 19.12) and for the compressed text 45.09 seconds (SD = 15.44).

There were no significant effects of text type on any metrics (see Table 5). The mean R-ratios of the text types are varying near 1, which is typically taken as a dividing line between easily resumable and not as easily resumable visual tasks [3]. This finding is in contradiction to the findings of the driving simulation experiments, which indicated that the spaced text type significantly better supported interruptible and resumable search behaviors than the compressed text.

Overall, besides the relatively long average task completion times [22], the R-ratios and the success in tasks suggest that these types of tasks could be performed quite efficiently while driving. However, in the driving experiments neither the visual behavior, especially in the Compressed-group, or overall driving performance decrements in both groups, can be taken that these types of tasks would be without potential distraction effects, at least for the majority of the participants.

Table 5. Mean values of the occlusion data (standard errors)

	Spaced (n = 10)	Compressed (n = 10)	t(18)	p
Total Shutter Open Time	31.26 (3.40)	34.33 (3.09)	.67	.510
Total Task Time Unoccluded α	39.51 (6.05)	45.09 (4.88)	.72	.480
R-ratio	1.04 (.20)	.84 (.10)	.86	.410

Note. α : Data displayed for text types.

In the post-experiment interviews, all the participants reported they tried to resume reading at the point they were before the occlusion interval with the compressed text. Only 1 participant out of 18 after the second driving simulation experiment reported to have utilized this strategy. Instead, most of the others utilized quick unorganized skimming of the text. Finally, an interesting point was observed when comparing the participants' subjective reports on task demands after the driving simulation and the occlusion experiments. There were considerably more "I could"-answers after the occlusion experiment than after the driving simulation experiments with the same secondary tasks when the participants were asked: "Would you consider reading these types of news or messages while driving?" After the second driving experiment, only 1 participant out of 18 reported that he would consider conducting these types of tasks while driving, while 5 answered "I could" and 3 were not sure out of 20 after the occlusion experiment. This finding can call in to question drivers' abilities to reliably assess their own capabilities and combined task demands without any actual experience of them.

GENERAL DISCUSSION

The results suggest that the external validity and sensitivity of the occlusion technique as metrics for providing information on task interruptability and distraction effects of visual display properties can be questioned. The selected tasks easily passed the qualitative criteria of successful task completion under occlusion although they were found as being potentially dangerous while driving in a simulated environment. At least there are visual secondary tasks that do not suit being tested with the occlusion technique as defined in the standard ISO 16673 [9]. These tasks involve information-filled displays in which the driver has to find textual information while driving. These types of displays are already found in systems providing unstructured and unpredictable e-mail and Internet content on an in-vehicle display or on the display of mobile device.

If the 15 second criterion of maximum duration of visual secondary task [22] is accepted, then task completion times seem to be plausible metrics to bench-test distraction without simulating the driving context. However, the occlusion test is in this case unnecessary. The average task completion times with the two different text types were 39.51 (SD = 19.12, spaced text) and 45.09 seconds (SD = 15.44, compressed text) unoccluded, and thus, significantly greater than 15 seconds, which would suggest that these tasks are potentially dangerous while driving. This is in line with the findings of the driving simulation experiments that the search tasks with either text type significantly degraded driving performance compared to baseline driving with the measures of lane excursions.

Another question is; do task completion times (see [26]) or successful performance while occluded, really say anything about the compatibility of driving and a secondary task? Task duration metrics, such as the occlusion metrics, evaluate distraction on the level of operational dual-task control (see [11]). The basic assumption behind these types of measures is that task completion times correlate with task complexity, or in this case, with the visual demand of the task. Even if this is the case and there is evidence that visual task times correlate fairly well with crash risk [5], as well as TSOTs with total glance times, lane maintenance performance, and speed variations while driving [27], these measures alone do not suit for the evaluation of driver distraction on the levels of tactical and strategic control. They do not provide any information on drivers' capabilities of sharing visual attention between driving and secondary tasks in time and thus, on how the IVIS should be redesigned to enable safer interaction strategies. It is highly possible, that visual tasks completed in less than 15 seconds can still be very distracting [3], [16], [26]. The task completion times occluded or unoccluded could not discriminate reliably between the two different text types, while the measures of time-sharing efficiency could do this in both driving simulation experiments with a comparable number of participants. Total glance times at the display differed significantly only in the second driving simulation experiment. Neither could the R-ratio succeed at this, and for this reason, the measure does not seem to work efficiently as a measure for task resumability. The near-one R-ratios suggest that both text types support equally interruptible and resumable tasks [3], while the time-sharing metrics in the driving simulation experiments clearly indicated that the spaced text type better supported interruptible search strategies.

The observed low R-ratios and the fact that the mean TSOTs were shorter than task times unoccluded with both text types deserve an explanation. Foley [3] speculates that participants probably feel time pressure with the occlusion procedure and therefore invest more effort in the task in the occluded condition than while unoccluded. Another explanation could be that the participant can process the task while occluded. A related question relates to what makes these types of visual tasks unsuitable for testing with the occlusion technique. An appealing explanation is that the task settings are not the same while driving, i.e. the ecological validity of the experimental conditions is low. Monk and Kidd [15] argue that the occlusion period does not impose any additional cognitive demand on the participant in contrast to the actual driving situation. Indeed, it seems that our mental machinery is rarely quite as efficient after task switches as when allowed to concentrate on a single task [18]. Absence of task set switch costs [18], [23] and memory effects of interruptions [17], as well as participants' abilities to utilize their iconic and short-term memory while occluded (see [25]) are plausible candidates for explaining the absent dual-task costs in the occlusion-based experimental settings. One can also question whether the mental representations, and for example, search strategies of the participants are the same in occlusion experiments as while driving. The interviews indicated that there is more variation in time-sharing strategies and also in individual time-sharing skills in the driving context. If our aim is to explore visual task's distraction effects and drivers' time-sharing capabilities on the levels of tactical and strategic control, the time-sharing metrics utilized in a driving simulation environment seem to be a more suitable option than the occlusion technique.

At this point, one could argue that the occlusion technique is intended to simulate visual glances at the road to test interruptibility and resumability of a task, and not for evaluating drivers' time-

sharing efficiency with the task. The occlusion technique is indeed a poor method of studying time-sharing considering that glance times on each task are fixed and not controlled by the participant which eliminates any strategic or tactical component. This was confirmed in our experiments. However, the value of an IVIS task interruptability test method that does not tell us much about drivers' abilities to efficiently interrupt and resume (i.e. time-share) the task under testing when coupled with driving, can be questioned. Underestimations of visual distraction provided by the occlusion technique in the current study indicate that we should be critical when applying these types of "quick and dirty" metrics when assessing distraction potential of our IVIS prototypes. The actual mechanisms of interaction with IVIS are often very different in the driving context than in a bench test.

There were no over-50-year-old participants in the experiments, although the occlusion standard [9] includes a 20 percent quota of elderly people. This quota might have helped also in the case of our visual tasks in showing more negative results given the possibly slower information processing capabilities of the elderly [7], and can be highly recommended. However, a quota of older drivers in the driving simulation experiments would most likely have also had a greater impact on driving performance and time-sharing metrics (see [30]).

The occlusion technique may be still well suitable as a screening tool in early evaluations of certain types of interaction designs, such as simple visual displays and dashboard manual controls (see e.g., [1], [2], [16], [19]). However, dual-task condition with a cognitive task similar to driving during occlusions could be a valuable addition for improving the external validity of the test results (see [15]). Additional evaluations with time-sharing metrics, involving actual or simulated driving, can be highly recommended especially when dealing with longer visual tasks and more informative display designs that provide drivers with more freedom to choose their interaction styles. Conclusions based on successful completion while occluded or low R-ratios can have severe consequences if they are used in deciding whether a design is ready or not to be implemented and made available for drivers.

ACKNOWLEDGMENTS

This research was partly sponsored by Theseus, a research project on human-technology interaction design funded by TEKES (Finnish Funding Agency for Technology and Innovation). The driving simulation environment was developed with a grant from the Henry Ford Foundation.

REFERENCES

- [1] Akamatsu, M. 2008. Japanese approaches to principles, codes, guidelines and checklists for in-vehicle HMI. In: *Driver Distraction: Theory, Effects, and Mitigation*. M.A. Regan, J. D. Lee and K. L. Young, (Eds.). CRC Press, Boca Raton, FL. Pp. 425–443.
- [2] Baumann, M., Keinath, A., Krems, J. F., and Bengler, K. 2004. Evaluation of in-vehicle HMI using occlusion techniques: experimental results and practical implications. *Appl. Ergon.*, 35, 197–205.
- [3] Foley, J. P. 2008. Now you see it, now you don't: Visual occlusion as a surrogate distraction measurement technique. In *Driver Distraction: Theory, Effects, and Mitigation*. M. A. Regan, J. D. Lee, and K. L. Young (Eds.). CRC Press, Boca Raton, FL. Pp. 123–134.

- [4] Gelau, C., Henning, M. J. and Krems, J. F. 2009. On the reliability of the occlusion technique as a tool for the assessment of the HMI of in-vehicle information and communication systems. *Appl. Ergon.*, 40, 181–184.
- [5] Green, P. 1999. Visual and Task Demands of Driver Information Systems (UMTRI Rep. 98-16). Ann Arbor, MI: The University of Michigan Transportation Research Institute.
- [6] Harbluk, J. L., Burns, P. C., Go, E. and Morton, A. 2006. Evaluation of the occlusion procedure for assessing in-vehicle telematics: Tests of current vehicle systems. In: *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. San Francisco, CA, October 16–20, 2006. Human Factors and Ergonomics Society. Pp. 2373–2377.
- [7] Horberry, T., Stevens, A., Burnett, G., Cotter, S. and Robbins, R. 2008. Assessing the visual demand from in-vehicle information systems by means of the occlusion technique: the effects of participant age. *IET Intell. Transp. Sy.* 2, 170–177.
- [8] Horrey, W. J. and Wickens, C. D. 2007. In-vehicle glance duration: Distributions, tails, and model of crash risk. *Transport. Res. Rec.* 2018, 22–28.
- [9] International Organization for Standardization 2007. ISO 16673 Road vehicles-ergonomic aspects of transport information and control systems-Occlusion method to assess visual demand due to the use of in-vehicle systems. Geneva, CH: International Standards Organization.
- [10] Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D. and Ramsey, D. J. 2006. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data (DOT HS Rep. 810 594). Washington DC: U.S. National Highway Traffic Safety Administration.
- [11] Kujala, T. (in preparation). Efficiency of visual time-sharing behavior – The effects of menu structure on POI search tasks while driving. In: *Proceedings of the Automotive UI 2009*. Essen, GER, September 21–22, 2009. Approved for publication.
- [12] Lansdown, T. C., Burns, P. C. and Parkes, A. M. 2004. Perspectives on occlusion and requirements for validation. *Appl. Ergon.*, 35, 225–232.
- [13] Lee, J. D., Regan, M. A. and Young, K. L. 2008. What drives distraction? Distraction as a breakdown of multilevel control. In: *Driver Distraction: Theory, Effects, and Mitigation*. M. A. Regan, J. D. Lee and K. L. Young, (Eds.). CRC Press, Boca Raton, FL. Pp. 41–56.
- [14] Michon, J. A. 1985. A critical view of driver behavior models: What do we know, what should we do? In: *Human Behavior and Traffic Safety*. L. Evans and R. C. Schwing (Eds.). Plenum Press, New York, NY. Pp. 485–520.
- [15] Monk, C. A. and Kidd D. G. 2007. R we fooling ourselves: Does the occlusion technique shortchange R estimates? In: *Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Stevenson, WA, July 9–12, 2007. University of Iowa Public Policy Center. Pp. 2–8.
- [16] Noy, I., Lemoine, L., Klachan, C. and Burns, P. C. 2004. Task interruptability and duration as measures of visual distraction. *Appl. Ergon.*, 35, 207–213.
- [17] Oulasvirta, A. and Saariluoma, P. 2006. Surviving task interruptions: Investigating the implications of long-term working memory theory. *Int. J. Hum.-Comput. St.*, 64(10), 941–961.
- [18] Pashler, H., Johnston, J. C. and Ruthruff, E. 2001. Attention and performance. *Annu. Rev. Psychol.*, 52, 629–651.
- [19] Pettitt, M. A., Burnett, G. A., Bayer, S. and Stevens, A. 2006. Assessment of the occlusion technique as a means for evaluating the distraction potential of driver support systems. *IET Intell. Transp. Sy.* 153, 259–266.
- [20] Regan, M. A., Lee, J. D. and Young, K. L. 2008. *Driver Distraction: Theory, Effects, and Mitigation*. CRC Press, Boca Raton, FL.
- [21] Society of Automotive Engineers 2000. SAE J2396 Surface vehicle recommended practice, definitions and experimental measures related to the specification of driver visual behavior using video based techniques. Warrendale, PA: Society of Automotive Engineers.
- [22] Society of Automotive Engineers 2004. Navigation and route guidance function accessibility while driving, SAE recommended practice J2364. Warrendale, PA: Society of Automotive Engineers.
- [23] Sakai, K. 2008. Task set and prefrontal cortex. *Annu. Rev. Neurosci.*, 31, 219–245.
- [24] Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W. and Ward, J. L. 1967. The attentional demand of automobile driving. *Highway Research Record* 195, 15–33.
- [25] Sperling, G. 1960. The information available in brief visual presentations. *Psychol. Monogr.-Gen. A.*, 74(11), 1–29.
- [26] Tijerina, L., Johnston, S., Parmer, E., Winterbottom, M. D. and Goodman, M. 2000. Driver distraction with wireless telecommunications and route guidance systems (DOT HS Rep. 809-069). Washington DC: U.S. National Highway and Traffic Safety Administration.
- [27] Tijerina, L. and Kochhar, D. 2007. A measurement systems analysis of total shutter open time (TSOT) as a distraction metric for visual-manual tasks. In: *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*. Baltimore, MD, October 1–5, 2007. Human Factors and Ergonomics Society. Pp. 1545–1549.
- [28] Wierwille, W. W. 1993. An initial model of visual sampling of in-car displays and controls. In: *Vision in Vehicles IV*, A. G. Gale, I. D. Brown, C. M. Haslegrave, H. W. Kruyse and S. P. Taylor, (Eds.). Elsevier Science Publishers B.V., Amsterdam, NL. Pp. 271–279.
- [29] Wikman, A. S., Nieminen, T. and Summala, H. 1998. Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics*, 41, 358–372.
- [30] Wikman, A. S. and Summala, H. 2005. Aging and time-sharing in highway driving. *Optometry Vision Sci.*, 82, 716–723.

New Interactions with Workflow Systems

Ingo Wassink, Paul E. van der Vet,
Betsy van Dijk
University of Twente
Enschede, the Netherlands
{wassink, vet, bvdijk}@ewi.utwente.nl

Gerrit C. van der Veer
Open University
Heerlen, the Netherlands
gerrit@acm.org

Marco Roos
University of Amsterdam
Amsterdam, the
Netherlands
m.roos@science.uva.nl

ABSTRACT

This paper describes the evaluation of our early design ideas of an ad-hoc of workflow system. Using the teach-back technique, we have performed a hermeneutic analysis of the mockup implementation named NIWS to get corrective and creative feedback at the functional, dialogue and representation level of the new workflow system.

Keywords

Ad-hoc workflow, teach-back, user evaluation

ACM Classification Keywords

D.3.2 [Language Classifications]; H.5.2 [User Interfaces]; J.3.1 [Life and Medical Sciences].

INTRODUCTION

Bioinformatics is the domain where life science meets computer science. The bioinformatician is a life scientist who uses computer tools and programming to perform biological experiments (known as in-silico experiments). An enormous amount of tools are available today as programs and web services, provided by many different organizations [2]. In a single experiment, multiple services are combined: data produced by one service is used as input for the next service. Bioinformaticians create scripts to connect the services used in an experiment. These experiments can become complex due to the huge amount of data and large number of services involved.

Workflow systems are developed to help bioinformaticians deal with the complexity of designing and running these in-silico experiments. Their chief appeal lies in the fact that they provide easy access to tools and services provided by different groups and using different protocols. A workflow system provides a graphical user interface in which task-labels represent programs and web services. The experiment itself is represented as a graph: the tasks are nodes of the graph and arrows are used to have the output of one task function as the input of another task and to indicate execution order. The user can create an experiment by dragging and dropping new tasks into the graph and connecting them.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Building a workflow is a difficult job. The bioinformatician has to choose the right services and, when services are connected, to deal with data incompatibility problems between services [2, 8]. The situation is even more complicated because in current workflow systems, the complete workflow needs to be designed in advance before it can be run. In practice, however, the complete setup of the experiment is often not known in advance [1, 4]. In such cases, the bioinformatician wants to decide on the next step of the experiment using the outcomes of steps that have been finished.

We propose a new type of workflow system, named NIWS (New Interactions in Workflow Systems). NIWS is an ad-hoc workflow system; it enables the bioinformatician to design and execute partial workflows. This system will better fit the explorative working approach of the bioinformatician. The outputs of the tasks in the partially designed workflow can be inspected to decide how the workflow will be extended and how the current output can be used as input for new tasks. The important question is, of course, will such a system satisfy the bioinformatician? To answer this question, we embarked on a systematic design approach: (1) we analyzed the domain problem; (2) we developed a view on a solution (adaptable workflows); (3) we developed a first draft design; (4) evaluated our envisioning (the current paper); and subsequently, (5) we will iterate on our design, finally build a full blown implementation, and assess its value in a real world setting. For step 4 we investigated the design's relevance and usability with bioinformaticians familiar with workflow systems, by performing a teach-back technique, a hermeneutic method to provoke the users to externalize their mental models [6].

We will first give an overview of existing studies of workflow systems. We will describe NIWS. Next, we will explain the teach-back technique. Then we will describe our empirical investigation with professional participants. After that, we will discuss our results and we will end with a reflection.

WORKFLOW SYSTEMS FOR SCIENTIFIC EXPERIMENTATION

Much research has been done on scientific workflow systems, though, only few consider the usability of these systems. There is often a big gap between the level of detail that is relevant for a life science problem and the level of detail required for the implementation of the experiment as a workflow [1]. Gordon et al. [5] performed a user study to test the usability of the Taverna workflow system. They found functionality problems due to the exploratory nature of life science life scientist need to interact with the workflow during the actual experiment. Direct interaction enables the life scientist to try parameter settings and to debug workflows [1].

Downey [3] performed a user study to test the usability of the Kepler workflow system. One of the main features workflow

users found were missing in this tool is a real-time debugger of the workflow to inspect intermediate results to make further decisions. The workflow system should guide its users to construct the workflow. Additionally, participants request for data directly being visible in the workflow diagram.

Gibson et al. [4] provided a first implementation and evaluation of an ad-hoc workflow system. The workflow designer can design and execute partial workflows and reuse the intermediate results to further design the workflow. The results of the user study were promising; however, the system is not further developed.

NIWS – AN ADAPTIVE WORKFLOW SYSTEM

In prior work, we have discussed our workflow system named e-BioFlow [7]. This workflow system previously only supports the classical approach in which the complete workflow has to be designed in advance. NIWS is mockup implementation of the ad-hoc extension for e-BioFlow to support and stimulate explorative experiment design and execution.

Designing and running workflows in NIWS is intended to be easier than in classical workflow systems. Tasks can be executed in isolation by pressing the play button in the task box. Input ports and output ports of the tasks are present at respectively the top and the bottom of the task box. The data consumed and produced by the tasks is explicitly present in the workflow as circles. The user can inspect these data and use them as sources of inspiration how to further design the workflow. The data can be defined to be input for new tasks. NIWS does not require the user to rerun the complete workflow, but only inserted and modified tasks.

Finding suitable tasks is difficult. NIWS has a search engine to help its users find tasks based on the name, the type of operation it performs, the inputs and outputs, and the authority that hosts the application the task represents.

But NIWS is more: it supports guided analytics. Based on the type of data in the workflow, it suggests tasks that can take these data as input. This helps the user to find compatible tasks in a quick manner, but at the same time it forms a source of inspiration of possible directions in the experiment. In a similar way, NIWS can help to deliver the input required for a certain task by suggesting tasks that can produce the right data.

To connect tasks, users often have to parse and build complex data structures. NIWS helps its users doing this for XML structures. It provides so-called composer and decomposer tasks to build and to parse XML structures. In case of a composer, the user only has to provide the attribute values of the XML; in case of a decomposer, NIWS will return the attribute values.

TEACH-BACK AS A TECHNIQUE FOR HERMENEUTIC ANALYSIS

People working with complex systems need a mental model of the system in order to (1) plan use; (2) actually interact with; (3) understand and assess the effect of the interaction; and (4) understand the meaning of unexpected system actions.

Mental models are knowledge structures inside people's mind, based on learning the semantics of the system and its context ("what-is" knowledge), experiencing the dialogue with the system ("how-to" knowledge), and understanding the representations of the system state, system actions and system feed-back (the "vocabulary" of the interaction).

Mental models actually develop based on a current need (to act, or to explain to a colleague, etc.), in a current context (with or without the system being at hand).

Since mental models are "mental" we cannot directly observe or register them. Hermeneutics is a philosophical method where an analyst develops understanding of the meaning an object (e.g., an artifact) has for a certain person or a certain group of people.

We apply the teach-back technique [6] for our hermeneutic analysis: We introduce prospective users of our design (professional bioinformaticians) to our early design ideas (use cases represented as realistic scenarios by introducing a realistic user persona, a typical context of use and a relevant task).

We then ask these users to *teach back* their understanding of the system to an imaginary colleague. In order to teach back, we pose, both, "what-is" questions and "how-to" questions, the latter in different degrees of similarity with the use cases shown in the scenarios. In order to record the externalized mental representations, we ask our users to write down (scribble, use key words and full text at will) their teach-back.

To interpret these representations, we first develop a scoring schema and fine tune this to a level where independent analysts reach agreement to an acceptable level. We aim at a level comparable with inter-rater reliability accepted for psychological personality measurement techniques.

ASSESSMENT OF A DESIGN ENVISIONING

The aim of the current study is to gain insight into the mental model bioinformaticians have about our early design envisioning, NIWS. Our study focuses on professionals (life scientists with some experience in using workflow systems), to analyse if this new system is an improvement over state of the art existing workflow systems. The study consists of three phases. First, the participants are shown a mockup of NIWS. Second, based on the scenarios four questions are asked to gain insight into the bioinformaticians' mental model of NIWS. Third, the filled protocols are scored in categories to explore the participants' mental models.

Setup

The mockup of NIWS is an animated slideshow presentation containing a narrative of a bioinformatician performing experiments, showing text and sketchy mockups of the system. A voice-over reads the text in the slideshow to make the presentation vivid and realistic. The presentation contains two scenarios that show various features of the envisioned system and suggest new possibilities when using this system. The scenarios are based on real-life situations in bioinformatics, but worked out using our system ideas. The presentation of the scenarios takes about ten minutes.

The four questions consists of a "What is" question, probing a semantic mental model, and three "How to" questions, probing procedural mental models. In the first question, the participant is asked to explain to an imaginary colleague Tom, who is familiar with workflow systems but does not know NIWS, what NIWS is. In the three "How to" questions, the participant is asked to explain to Tom how to perform a particular task using NIWS. These tasks are not explicitly covered by the scenarios, but using NIWS could be inferred from them in relation to the individual participant's mental model.

The questions are distributed on paper. To respond, participants can write, scribble, make drawings, etc. The participants get five minutes to answer each question. They are, however, not allowed to discuss or to ask questions, since we are interested in what they believe the system can do. We do explicitly mention that the questions are not to test the participants' knowledge: there are neither right nor wrong answers. Participation is anonymous and voluntary. All participants are rewarded for participation with a 1 GB USB key.

Participants

In total, there are 50 respondents, originating from different countries, though most of them are Dutch. The participants have different backgrounds (biology, bioinformatics, chemistry, computer science) and their expertise in using workflow in bioinformatics experiments differs from beginner to experienced user. These respondents are recruited during six sessions: during visits at life science research groups, courses in the Taverna workflow systems and a meeting of the BioAssist Group. The size of the groups ranges from 1 to 20 persons.

A strict protocol is handled in these sessions in order to keep the experiment reproducible. In each session, an experimenter is present to start the scenarios, to distribute and collect the protocols and to manage the time. No information about NIWS is given to the participants other than the scenarios. The participants received the reward when they handed in the form.

Scoring the Protocols

The result of teach-back may consist of both creative and corrective feedback. Creative feedback will encompass new features the participants expect to exist based on the scenario. In corrective feedback, the participants mention features they do not like, expect not to work, or want to be improved.

The feedback is analyzed regarding three levels of the system: (1) feedback related to the functionality: what the participants believe the system can do and what its limitations will be; (2) feedback related to the dialogue; (3) feedback related to the representation of the workflow experiment and the system interface.

A scoring scheme is set up to analyze the forms in an unambiguous and reproducible way. This scheme consists of rules and examples how to categorize the feedback. To set up this scheme, two analysts (authors) separately analyzed five forms. They discussed their findings with a third author and built a scoring scheme. The two analysts separately scored another three protocols and compared their scorings to test the agreement on the scoring scheme, which confirmed interpretation and scoring reliability. Consequently a single analyst was sufficient to score the remaining 42 protocols.

RESULTS

The results are grouped along the scoring levels. In 6.1 we show the results (illustrated by examples from the protocols) for functionality:

- Corrective feedback: functionality ("what-is" knowledge) as indicated in the scenarios that we found back in the protocols and that is consistent with the scenarios, as well as functionality understood by the participants that is inconsistent with our scenarios, and indications of functionality aspects not appreciated by the participants.
- Creative feedback: We will show examples of functionality found in the protocols not mentioned in the scenarios, that makes sense as extension of the design.

Based on this we intend to repair, expand and improve the functionality of NIWS in the next phase of this project.

In 6.2 we will report on corrective feedback and creative feedback regarding the dialogue ("how-to" knowledge) of NIWS, and in 6.3 we will do the same for feedback regarding the representations. On this last aspect we need to keep in mind that the scenarios as presented by us are describing our NIWS design ideas at a global level, focusing on the functionality, and hinting the dialogue, but being vague on the actual representation of the system interface and on the users' actions.

Functionality

Most respondents react positively on the system presented. Many of them mention NIWS is like other workflow tools, but then more intuitive, simpler or easier to use. As one said, "a big plus is that you can add additional processing anywhere in the chain, without having to re-run everything as it caches intermediate results". Another respondent mentioned "You don't have to rerun the workflow every time. Therefore you will save a lot of time". It is also easier to use for beginners: "NIWS is this new workflow system that has this cool feature of giving you hints when you don't know what to do. Ideal for beginners like me :-)." However, one respondent said the questions were easier to solve without using a workflow system.

Many respondents picked up the idea of designing workflows step by step. Intermediate results can be used to further design the workflow. "The nice thing is that one can execute every process in isolation and that one can inspect the outputs of the workflows at any moment." NIWS enables one to execute the partial workflow, to test and debug the workflow. One respondent describes this as "kneading" de workflow.

Eight respondents propose a two step approach to design and run workflows in case large data sets are analyzed. First, design a workflow using a small example data set. Second, when the design is finished, run the workflow for the entire data set. Another respondent suggests to create a workflow for one data item, and to embed this one into a larger workflow that runs it for each data item of the complete set in parallel.

To find services, 27 respondents recommend using the search facility of NIWS, though some of them found the use of this facility to be unclear. One respondent expects the search function to be smart: meta-data can be used to further refine the search. For example, the database name can be used to find blast services that have access to that database. Others recommend using external resources, such as Google or colleagues, to find services. NIWS is expected to provide access to many different types of web services, such as BioMOBY, REST, XML-RPC and SOAP/WSDL.

The feature of NIWS to suggest services that can take data available in the workflow as input is picked up by 11 respondents. Three of them explicitly mentioned that they expect the suggested services to be compatible with the data in its current format; so no data conversion should be needed.

NIWS's functionality to automatically compose and decompose XML data is found useful by many respondents. Sixteen respondents even expect these facilities to solve all data format problems and data conversion to be a built-in feature of NIWS. Others, however, were skeptic about the automatic data conversion facilities: "If this went well, e.g. if you would never experience data compatibility issues, is questionable, because the output of one service needs to know what kind of format is expected as input of the other service". Some respondents expect support for scripting facilities, including query languages, to

perform the data conversion. These scripting facilities could be used to perform data transformations, but also to affect the control flow of the workflow. Others recommend searching for external data conversion services that will hopefully transform the data into the right format.

Dialogue

The scenarios show the drag and drop facilities of NIWS. Similar, two respondents expect copy and paste functionality to be available for easily reusing parts of workflows. A few respondents expect the option to embed workflows previously designed or designed by others in larger workflows.

Many respondents have picked up to use the play button in the task box to run a task in isolation. From the scenarios, it is not clear whether tasks upstream in the workflow will be executed automatically. One respondent supposes this to be the case. NIWS will ask its user to enter missing data. In case only a fixed set of options is valid as input, NIWS will list them to let users choose from them. One respondent mentioned it would be nice if users can also choose from data already in the workflow.

Many respondents perceived that composer tasks and decomposer tasks can be added by right-clicking on respectively the input and output ports of a task to feed correct input or parse output. One respondent described this as some magic: "The blast service needs some magic before we can use it, so we must tell [NIWS] to do its magic. The result is two boxes which we can give our [user] name and sequences". He refers to the composition tasks to deliver the correct XML input. Some other respondents expect right-clicking on ports is used to set default input values. Two respondents expect NIWS to add composer tasks and decomposer tasks automatically. Two other respondents declared that building and parsing hierarchical XML structures can be established by a chain of composition or decomposition tasks.

NIWS is expected to warn about the existence of data compatibility problems. "The system will give an error if the outputs don't match the inputs. When that is the case (and yes, this will happen) write a small converter script to process the output." To create these scripts, a good editor with auto-completion and syntax highlighting is desirable.

One respondent recommends limiting the amount of mouse interaction required: "It looks like a lot of clicking is required, for every composer and decomposer and to execute a task, you have to click."

Representation

In total, 28 respondents use drawings to explain Tom how to use NIWS. In many drawings, the tasks boxes have inputs and outputs explicitly visible at top and bottom. In the scenarios, data are presented as circles, which makes the workflow graphs look like colored Petri nets. Data are explicitly present in drawings of 19 respondents. These data are connected by arrows coming from output ports and arrows going to input ports. Some respondents drew data using boxes instead of circles. So, the difference between these two symbols seems to be unclear. One respondent used stacked circles to represent collections of data in case a task returns multiple items.

DISCUSSION

A scenario-based mockup implementation is an easy and a fast way to evaluate design ideas (e.g., of NIWS) in an early stage of the design. Applying the teach-back technique we found that

NIWS is a significant improvement over traditional workflow interfaces. Many respondents put high value on the ability to inspect intermediate results to further design the workflow. Helping with data conversion and finding and suggesting services are other features these respondents put high value on.

Besides positive feedback, respondents gave feedback about desired **functionality** of a workflow system, even of aspects not shown in the scenarios. Furthermore, the respondents gave directions to improve the **interaction** with the system presented and other workflow systems.

The results of this study will be used to develop an interactive implementation of NIWS in our workflow tool e-BioFlow.

ACKNOWLEDGEMENT

Our thanks to L. Packwood for her help to design the mockup of NIWS. We like to thank J. A. M. Leunissen, P. B. T. Neerinx, M. van der Bruggen, H. Rauwerda, T. M. Breit and M. van Driel for their help to find participants for the study.

This work was part of the BioRange programme of the Netherlands Bioinformatics Centre (NBIC), which is supported by a BSIK grant through the Netherlands Genomics Initiative (NGI).

REFERENCES

- [1] Addis, M., Ferris, J., Greenwood, M., Li, P., Marvin, D., Oinn, T. and Wipat, A. Experiences with e-Science workflow specification and enactment in bioinformatics. In: S. J. Cox (Ed.). *e-Science All Hands Meeting 2003*, Nottingham, United Kingdom, 2004. Pp. 459–466.
- [2] Curcin, V., Ghanem, M. and Guo, Y. Web services in the life sciences. *Drug Discovery Today*, 10(12), 865–871, 2005.
- [3] Downey, L. L. Group usability testing: Evolution in usability techniques. *Journal of Usability Studies*, 2(3), 133–144, 2007.
- [4] Gibson, A., Gamble, M., Wolstencroft, K., Oinn, T. and Goble, C. The data playground: An intuitive workflow specification environment. In: S. J. Cox, (Ed.). *E-SCIENCE '07: Proceedings of the Third IEEE International Conference on e-Science and Grid Computing*, Washington, DC, USA, 2007. IEEE Computer Society Press. Pp. 59–68.
- [5] Gordon, P. M. K. and Sensen, C. W. A pilot study into the usability of a scientific workflow construction tool. Technical Report 2007-874-26, Sun Center of Excellence for Visual Genomics, 2007.
- [6] Puerta Melguizo, M. C., Chisalita, C. and Van der Veer, G. C. Assessing users mental models in designing complex systems. In: A. El Kamel, K. Mellouli and P. Borne, (Eds.). *IEEE International Conference on Systems, Man and Cybernetics, Yasmine Hammamet, Tunisia*, 2002. Pp. 1–6,
- [7] Wassink, I., Rauwerda, H., Vet, P. E. van der, Breit, T. M. and Nijholt, A. e-BioFlow: Different perspectives on scientific workflows. In: M. Elloumi, J. Küng, M. Linial, R. F. Murphy, K. Schneider and C. Toma, (Eds.). *2nd International Conference on Bioinformatics Research and Development (BIRD)*, Vienna, Austria, 2008. Pp. 243–257,
- [8] Wassink, I., Vet, P. E. van der, Wolstencroft, K., Neerinx, P. B. T., Roos, M., Rauwerda, H. and Breit, T. M. Analysing scientific workflows: why workflows not only connect web services. In: L. J. Zhang, (Ed.). *SERVICES 2009 (Part I)*, Los Angeles, USA, 2009. Pp. 314–321.

Session 12: Developing Tools for Design Activity

An Approach to Assess the Quality of Collaboration in Technology-mediated Design Situations

Jean-Marie Burkhardt
Paris Descartes University,
Ergonomics–Behavior-Interaction Lab.
45 rue des Saints-Pères
75270 Paris Cedex 06, France
jean-marie.burkhardt@univ-
paris5.fr

Laurence Perron
France Telecom R&D / Orange Labs
Lannion, France
laurence.perron@orange-
ftgroup.com

Françoise Détienne
LTCI- UMR 5141
CNRS – Telecom ParisTech
46 rue Barrault
75634 Paris Cedex 13, France
1 INRIA-Paris France
francoise.dettienne@telecom-
paristech.fr

Stéphane Safin
Lucid Group
Université de Liège –Belgique
stephane.safin@ulg.ac.be

Anne-Marie Hébert
LTCI- UMR 5141
CNRS – Telecom ParisTech
46 rue Barrault
75634 Paris Cedex 13, France
hebert@telecom-paristech.fr

Pierre Leclercq
Lucid Group
Université de Liège –Belgique
pierre.leclercq@ulg.ac.be

ABSTRACT

Our objective is to measure and compare the quality of collaboration in technology-mediated design activities. Our position is to consider collaboration as multidimensional. We present a method to assess quality of collaboration which is composed of seven dimensions concerning communication processes such as grounding, coordination processes, task-related processes, symmetry of individual contributions as well as motivational processes. This method is used in a study aiming to compare the quality of collaboration in architectural design. In this experimental study, design situations vary according to technology-mediation – co-presence with an augmented reality (AR) environment versus distance with AR and visio-conferencing -, and according to number of participants – pairs versus groups of four architects -. Our results show that distinctive dimensions of collaboration are affected by the technology mediation and/or the number of co-designers. We discuss these results with respect to technology affordances such as visibility and group factors.

Keywords

collaboration, design, methodology, cognitive ergonomics, Computer Supported Collaborative Learning

ACM Classification Keywords

C4. Design Studies; H.5.2 User Interface. Ergonomics.

INTRODUCTION

With the growing importance of technology mediation for group work, developing methods for assessing the quality of collaboration should become as central as developing methods for assessing usability of UI in user-centred design. In spite of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

a growing number of methods to evaluate groupware technologies and group work, no measurement method of this facet of collaboration has been proposed as far as we know.

Our objective is to measure and compare the quality of collaboration in technology-mediated design activities. In this context, the term ‘quality’ can be understood in descriptive terms (identifying and discriminating the intrinsic properties of collaboration) and/or in a normative sense (identifying what makes ‘good’ or less good collaboration, considered *sui generis*). We consider these visions of quality as complementary. On the basis of previous work in cognitive ergonomics of design and in computer supported collaborative learning (CSCL), we consider collaboration as multidimensional and propose an evaluation method that covers these specific dimensions. Furthermore, whenever possible we provide norms and explicit qualitative references to support the comparison of measures in various technology-mediated situations.

In the first part of the paper, we provide the rationale of this multidimensional approach on the basis of theoretical arguments and of results from empirical studies. We also aim to elicit references (often implicit in the literature) and relevant standards regarding collaborative activities. A method to assess selected dimensions of activity related to collaboration quality is then presented, followed by the test of its reliability. In the second part, we present a study aiming to compare the quality of collaboration, relatively to our various dimensions, in contrasted technology-mediated design situations. We use our assessment method to compare quality of collaboration in design situations varying according to technology-mediation – co-presence with an augmented reality (AR) environment *versus* distance with AR and visio-conferencing – and according to number of participants – pairs *versus* groups of four architects. The results of this empirical study are presented and discussed.

COLLABORATION IN DESIGN: A MULTI-DIMENSIONAL APPROACH

Empirical studies on the process of collaboration in design teams (for a state of the art, see [1]), in various application domains (e.g., software design, architectural design), have highlighted distinctive collaborative processes most important for successful design. These processes can be taken as a referential of good collaboration with respect to design. They

can be grouped along several dimensions concerning communication processes such as grounding, task-related processes (e.g. exchanges of knowledge relevant for the task at hand; argumentation processes), coordination processes, and motivational processes. Furthermore we consider, through all these dimensions, how symmetric individual contributions are in order to provide a complementary aspect of collaboration and its quality.

Communication Processes

Communication processes are most important to ensure the construction of a common referential within a group of collaborators. The establishment of common ground is a collaborative process [2] in which the co-designers mutually establish what they know so that design activities can proceed. Grounding is linked to sharing of information through the representation of the environment and the artefact, the dialog, and the supposed “pre-existing” shared knowledge. This activity ensures inter-comprehension and construction of shared (or compatible) representations of the current state of the problem, solutions, plans, design rules and more general design knowledge.

Empirical studies of collaborative design (e.g. [3, 4, 5]) found that grounding, although time-consuming, was most important to ensure good design: for instance, Stempfle and Badke-Schaub [5] found that when teams bypassed grounding (referred to as “analysis”), this led them to premature evaluation of design ideas.

Key characteristics of collocated synchronous interactions are assumed to support grounding [6]. Rapid feedbacks allow for rapid corrections when there are misunderstandings or disagreements. Multiple channels (visual, oral, etc.) allow for several ways to convey complex messages and provide redundancy: e.g., gaze and gestures help to easily identify the referent of deictic terms. The shared local context allows for mutual understanding. At distance, characteristics of communication media, such as the lack of visibility or simultaneity [2], may affect grounding and awareness. Using videoconferencing tools can extend the channels by which people communicate.

Task-related Processes

Task-related processes concern the evolution of the design problem and solution: (a) design activities, i.e., elaboration, enhancements of solutions and of alternative solutions; (b) evaluation activities, i.e., evaluation of solutions or alternative solutions, on the basis of criteria. These activities are supported by argumentation and negotiation mechanisms. These content-oriented mechanisms reveal how the group resolves the task at hand by sharing and co-elaborating knowledge concerning the design artefact, by confronting their various perspectives, and by converging toward negotiated solutions.

Whereas studies show evidence that these mechanisms are important for the quality of design products (e.g. [7]), empirical studies show that important drawbacks of observed design teams (e.g. [5]) may be: limitation in solution search; early choice of a solution without exploration of all alternatives; rapid solution evaluation on the basis of just a few criteria; difficulties in taking into account all criteria and their interdependencies (constraint management).

Technology-mediation tends to have few effects on these design processes. For example, a previous study [8] showed a similar distribution of the main categories (based on a coding scheme of interactions) of activities related to design for pairs of architects in co-presence and at distance (with videoconferencing and digital tablets). This absence of effect could be, however, specific to synchronous collaborative situations.

Group Management Processes

Collaboration concerns group management activities such as: (a) project management and coordination activities, e.g., allocation and planning of tasks; (b) meeting management activities, e.g., ordering, postponing of topics in the meeting. These process-oriented mechanisms ensure the management of tasks interdependencies, which is most important in a tightly coupled task such as design. These coordination mechanisms tend to become more central with technology mediation [e.g. 9].

Cooperative Orientation and Motivation

Although less covered in previous studies on design as well as in studies on technology-mediated collaborative design, cooperative orientation and motivation may be considered as important aspects of collaboration. Indeed, recent research on collaboration processes in design [10, 11] considers the participants’ roles in communication, group management and task management. The balance between these roles is considered as a good indicator of collaboration. This aspect is similar to the notion of reciprocal interaction highlighted by Spada et al. [12] and symmetry in the interaction pointed out by Baker [13] or Dillenbourg [14] in CSCL. These authors consider that quality of collaborative learning in small groups of learners is linked to the symmetry of the interaction. We will adopt a similar posture to assess the quality of collaboration in small teams of designers. We will also consider the dimension of motivation as important in so far as it can strongly affect the actual way of collaboration.

CURRENT APPROACHES OF COLLABORATION IN EVALUATION METHODS OF GROUPWARE TECHNOLOGIES

Groupware evaluation methods are of two types: usability inspection methods and usability evaluation methods based on users’ studies. On the inspection side, usability evaluation techniques [e.g., 15] do not rely on the participation of real users of the system. Advocates of these techniques argue that they are less costly than field methods and they can be used earlier in the development process. However their focus remain on individual-centered task models [e.g., 16], i.e. eliciting goals and actions required for users to interact together and not on the collaboration processes and their quality per se. Furthermore, they do not explore effective collaboration processes in context.

Regarding user studies, there is a lot of methods which rely on different data collection and analyses techniques: they can be based on computers logs, interactions between participants (coding methods or ethno-methodological methods), or interviews. In field or experimental studies, the few indicators used to assess usability regarding collaboration processes are focused on quantifying fine-grained interactions. An example, given in a recent review by Hornbaek [17], concerns the measure of “communication effort”: number of speakers’ turns; number of words spoken; number of interruptions; amount of grounding questions. Furthermore, quantitative variations of such indicators are non-univocal: any increase or decrease of them could signify either an interactive-intensive collaboration or evidence of huge difficulties in establishing or maintaining the collaboration. Several other drawbacks of these methods are usually pointed out: they are often difficult (and sometime impossible) to apply with prototypes and they are most time-consuming.

Two additional criticisms are the extent to which existing empirical-based methods cover all the dimensions of collaboration and their generality or *ad hoc* nature. Indeed, user-based methods to assess collaboration usually concentrate only on one or two dimensions among the numerous ones we wish to cover: for example, verbal and gestural communication to assess the grounding processes. Furthermore, motivational aspects as well as the balance/symmetry of individual contributions are rarely considered, although they reflect complementary aspects of collaboration assessment. Assessment methods also vary according to their generality and their explicit/formal characteristics. Ethno-methodological approaches generally remain *ad hoc* to the analyzed situations and they do not rely on any explicit methods. This is based on the adopted theoretical position considering the particular context under study as most important. Other user-based approaches rely on coding schemes making explicit categories of analysis, but they often remain *ad hoc* to the observed situation. Still, in some task application domains, some efforts have been made to construct more generic categories (see for example, [4, 5, 8]). To summarize, none of the user-based methods¹⁸ developed in the Computer-Supported Cooperative Work (CSCW) field are both multidimensional and generic.

In the close field of CSCL, the analysis of the process of collaboration is also a central topic of research. The Spada rating scheme [12, 18, 19] is certainly the most representative of recent effort made in this field to assess collaboration and its quality. It has been developed to compare and assess collaboration in collaborative learning tasks, with respect to various learning methods or technical support. An advantage of this method, beside its low temporal cost¹⁹, is its coverage of a wide range of collaboration dimensions. Indeed, these authors consider nine qualitatively defined dimensions that cover five broad aspects of the collaboration process: communication (sustaining mutual understanding, dialogue management), joint information processing (information pooling, reaching consensus), coordination (task division, time management, technical coordination), interpersonal relationship (reciprocal interaction) and motivation (individual task orientation). A review of the literature [19] on computer-supported collaborative learning and working provides theoretical arguments to consider these five aspects as central for successful collaboration under the conditions of video-mediated communication and complementary expertise. Furthermore, their method is generic enough to be applied to different technology-mediated situations.

However, Spada's method shows some limits from our viewpoint. Indicators exploited by judges (or raters) in order to assess collaboration are underspecified. Indeed, the method relies essentially on the subjective evaluation of each given dimension on a 5-grade Likert-like scale, oriented by a training manual. One consequence is that such an approach hides the reference to any quantifiable events or observables from the collaborative situation, preventing any track back from the assessment values to original data. However, its multidimensional characteristic endows the method with a good basis to further develop a method to assess the quality of collaboration in technology-mediated design. In this objective, we modified the assessment procedure to make the observable indicators underlying

the evaluation more explicit. This is reported in the next section presenting our method.

THE MULTI-DIMENSIONAL METHOD

We propose a multi-dimensional method to evaluate the quality of collaboration in technology-mediated design. Our method is initially (and thus partly) based on Spada's method [12]. It has been modified so as to take into account characteristics of collaborative design and to improve the assessment procedure.

Assessment Procedure, Dimensions and Indicators of Collaboration

In an initial version, our method kept the principle of subjective scale rating by Spada, i.e., the judges were requested to give a score on a 5-grade Likert scale for each of the dimensions. However to generate explicit traces of the rating processes, we modified the scoring method by requesting the judges to give additional explicit answers (Yes, Yes/No, No) to questions related to the specific indicators of each dimension (Table 1). For each indicator, we balanced questions with positive valence and questions with negative valence. These questions distinguish between what we consider as "good" collaboration (question with positive valence) and collaboration with a lowest quality (questions with negative valence) with respect to successful collaborative design. As an illustration, let us consider two examples. For the indicator "mutual understanding of the state of design problem/solutions" of *Dimension 2 (Sustaining mutual understanding)* the judge is requested to answer two questions (by Yes, Yes/No, or No): the question with positive valence is "Do the designers ask questions, give clarifications or complementary information, using verbal or behavioural backchannels, on the state of the design artefact?". The question with negative valence is "Are there misunderstandings on the state of the design artefact during relatively long periods of time?". As another example, the indicator "common decision taking" of *Dimension 4 (Argumentation and reaching consensus)* is splitted up into two questions: the question with positive valence is "Are the individual contributions equal concerning the design choices?"; the question with negative valence is "Is there one contributor who imposes the design choices?".

Consequently, assessment of the quality of collaboration was based on the rating of the seven dimensions plus 46 questions distributed along these dimensions. There were 3 positive + 3 negative questions for dimensions 1, 2, 3, 4, 7 and 4 positive + 4 negative questions for dimensions 5, 6. For dimension 7, both the questions and the scale rating are applied to each participant of the assessed collaborative situation.

We also modified the definition of dimensions to take into account the nature of the design task and the technology mediation characteristics²⁰. For the first aspect, we made explicit design task processes for *Dimension 3 (Information exchanges for problem solving)* and *Dimension 4 (Argumentation and reaching consensus)*: in particular those related to design constraints, design problem, design solutions, design decisions. For the second aspect, we distinguished for *Dimension 2 (Sustaining mutual understanding)* between processes related to the state of the design, the state of the system, and actions in progress.

¹⁸ Modeling techniques such as CUA (Collaboration Usability Analysis) [15] could be considered as multidimensional and generic. However this method does not enter into the category of user-based methods.

¹⁹ The judges can directly apply the rating method on the basis of video recordings. Thus they do not rely on transcriptions which are time-consuming.

²⁰ We also added items to make explicit the collaborative modalities (verbal, gestural, graphical, textual) predominantly associated to each considered dimensions of collaboration. We will not discuss this aspect of the evaluation in this paper.

Table 1. Dimensions and indicators of our method.

Dimensions	Definition	Indicators
1. Fluidity of collaboration	It assesses the management of verbal communication (verbal turns), of actions (tool use) and of attention orientation.	<ul style="list-style-type: none"> - Fluidity of verbal turns - Fluidity of tools use (stylet, menu) - Coherency of attention orientation
2. Sustaining mutual understanding	It assesses the grounding processes concerning the design artefact (problem, solutions), the designers' actions and the state of the AR disposal (e.g., activated functions).	<ul style="list-style-type: none"> - Mutual understanding of the state of design problem/solutions - Mutual understanding of the actions in progress and next actions - Mutual understanding of the state of the system (active functions, open documents)
3. Information exchanges for problem solving	It assesses design ideas pooling, refinement of design ideas and coherency of ideas.	<ul style="list-style-type: none"> - Generation of design ideas (problem, solutions, past cases, constraints) - Refinement of design ideas - Coherency and follow up of ideas
4. Argumentation and reaching consensus	It assesses whether or not there is argumentation and decision taken on common consensus.	<ul style="list-style-type: none"> - Criticisms and argumentation - Checking solutions adequacy with design constraints - Common decision taking
5. Task and time management	It assesses the planning (e.g. task allocation) and time management.	<ul style="list-style-type: none"> - Work planning - Task division - Distribution and management of tasks interdependencies - Time management
6. Cooperative orientation	It assesses the balance of contribution of the actors in design, planning, and in verbal and graphical actions.	<ul style="list-style-type: none"> - Symmetry of verbal contributions - Symmetry of use of graphical tools - Symmetry in task management - Symmetry in design choices
7. Individual task orientation	It assesses, for each contributor, its motivation (marks of interest in the collaboration), implication (actions) and involvement (attention orientation).	<ul style="list-style-type: none"> - Showing up motivation and encouraging others motivation - Constancy of effort put in the task <ul style="list-style-type: none"> • Attention orientation in relation with the design task

Testing the Reliability of the Method

We tested the reliability as well as the usability of an initial version of the method on the basis of inter-raters correlations, interviews and analyses of judges' activity during their application of the method (see [20] for the evaluation approach). Four judges participated in this first test (one of them is the third author of this paper). After an initial training on the method involving its application on one excerpt of 15 minutes, followed by a debriefing, the raters were provided with three excerpts of 10 minutes to be assessed. The test of this early version showed a weak inter-raters agreement of subjective rating, i.e., on 5-grade notes on dimensions. Conversely, we observed a high inter-raters agreement in the way they responded to the specific questions (all kappas > 0.60). Additionally, 17% of the questions were either misunderstood or judged as not applicable to the extracts.

On the basis of these results, the difficult questions were either reformulated or withdrawn and we decided to adopt an explicit scoring algorithm based on the number of positive and negative answers to questions, instead of subjectively rating each dimension. Thus, the score for each dimension is calculated on the basis of the answers given by the judge to the questions. The reliability of this second version was tested by looking again at inter-raters agreement. Three judges (the first three co-authors of this paper) assessed independently the quality of collaboration in an excerpt of collaborative design task. After an initial training on the method involving its application on two excerpts of 15 minutes, each followed by a short debriefing to

elicit and share a set of common rules with the method, the three raters were provided the same (and previously not viewed) excerpt to be assessed. The analysis of all the responses to questions for all dimensions shows an excellent inter-rater agreement (96.15% of agreement between the three raters on all questions; Kappa=.92). Furthermore, specific analyses of responses show an excellent agreement for both Yes (Kappa=.94) et No answers (Kappa=.92). Only Yes/No questions exhibited a lack of agreement (Kappa = 0.00) due to their very low frequency (2/78). This can also be explained by a constructed assessment rule which was to avoid this answer mode as far as possible and to favour a clear Yes or No decision. We used this version of method in the empirical study presented in the following section of this paper.

EMPIRICAL STUDY OF QUALITY OF COLLABORATION IN TECHNOLOGY-MEDIATED DESIGN SITUATIONS

We conducted an empirical study to compare the quality of collaboration with our method across three distinctive technology-mediated design situations. The design domain is architecture. Briefly, groups of two or four architects were asked to collaborate to solve a design task, either in co-presence around a virtual desktop or at distance with a virtual desktop on each site and a videoconferencing set up. The work sessions were entirely videotaped and solutions recorded. They represent about 18h of video records.

Participants

Sixteen last year students in architecture or in building engineering and architecture (from 22 to 26 years old) participated in this study. They had a similar experience in design tasks, to avoid biases due to the diversity of professional practices, particularly important in architecture. They were distributed arbitrarily into three conditions: pairs in co-presence, pairs at distance and 4-designers groups at distance. They were paid for their participation.

The Three Technology-mediated Experimental Conditions

The study is based on an integrated aided design tool, the EsQUIsE software, based on a Virtual Desktop [21, 22]. This environment has been developed to assist architects in preliminary design while keeping the natural and simple characteristics of the pen/paper drawing process. It is composed of a mixed software and hardware solution which offers (i) the natural aspect of digital freehand sketching, (ii) the ability of drawing interpretation and generation of 3D models based on 2D sketches and, (iii) the direct model manipulation and evaluation of performances (presently in building engineering). The system consists in a classical "A0" desk with a suspended ceiling equipped with a double projection system offering a large working surface (approximately 150x60 cm). The electronic stylus allows drawings of virtual sketches on this surface. The designers can manipulate their drawings and are provided with automatically generated models without having to use any usual modeller in the AR environment. To ensure the sharing of data in real time at distance, the software Sketsha was used in the distant conditions. This software is very similar to EsQUIsE except for the sketches interpretation and the generation of 3D models, which are not supported.

We contrasted three technology-mediated situations with two observations for each situation:

- Two pairs were in co-presence (cf. Figure 1): architects of a pair were sat side by side on a virtual desktop.
- Two pairs were at distance (cf. Figure 2): each architect of a pair had at his/her disposal a virtual desktop and a videoconferencing set up.
- Two groups of 4 architects at distance were spatially distributed as follows (cf. Figure 3): each group was composed of two collocated pairs working on a virtual desktop with a videoconferencing set up.

The videoconferencing system (IChat on 17" monitors with integrated webcam) allowed distant designers to collaborate through multiple channels and modalities (video, audio). The two virtual desktops in the distant situations were connected in a completely synchronous way.



Figure 1. A pair of architects collaborating in co-presence.

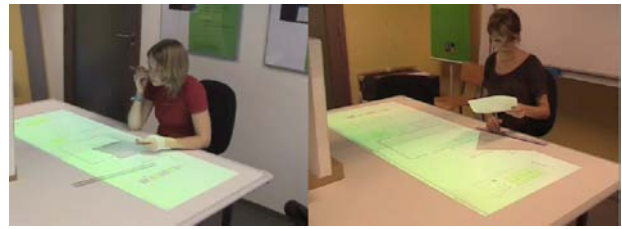


Figure 2. A pair of architects collaborating at distance.



Figure 3. A Group of four architects collaborating at distance.

Design Problems

One version of two similar architectural design problems (rural school and urban school) was given to each pair or group. This problem is a pedagogical exercise representative of design problems commonly faced by architects in their practices [23]. Furthermore, it is assumed that the two versions of the problem, with similar constraints and proportions, do not influence the nature of the activity and the collaboration of designers.

Collected Data

The experimental sessions were recorded between May 2006 and April 2008 in Liège. In the co-presence condition, we used two video-recorders to capture both a detailed view of the workspace and a larger view of interactions between participants. In the distant conditions, pairs or groups of 4 participants, we used two video-recorders per site with the same two views adjusted. The video recordings were synchronized.

Applying the Method for Evaluating the Quality of Collaboration

Two excerpts for each session were selected: one at the end of the first part and one at the beginning of the fourth part of the session. All excerpts concerned generation of solution and collective phases. So we had a total of 12 excerpts of 15 min.

One judge (third author of this paper) evaluated the quality of collaboration in each observed session, on the basis of two excerpts for each session. This evaluation was made on the basis of the second version of our method (Table 1). The judge began by assessing all excerpts from the collocated situations, then the excerpts from the pairs of architects at distance and the excerpts of the groups of architects at distance. The judge took 7 hour 30 minutes to apply the assessment method on the 3 hours of excerpts. We also verified that all design solutions elaborated by participants satisfied the constraints of the design problem.

Results

Sensitivity of Dimensions to Experimental Conditions

Scores on dimensions exhibit different sensitivity to the various cooperative situations involved in this research (Table 2). Whereas we found a minimal difference of 0.5 point (10%) between highest and lowest scores for *Dimensions 1, 3 and 5* (*Fluidity of collaboration, Information exchange for problem solving, Task and Time management*), we observed differences up to 2 points (40%) between some configurations for *Dimension 6* (*Cooperative orientation*). The *Dimensions 2, 4, and 7* show differences that fall within this range: *Sustaining mutual understanding* (1, i.e. 20%), *Argumentation and reaching consensus* (1, i.e. 20%), *Individual task orientation* (0.8, i.e. 16%).

Comparing respective scores in the three conditions for each dimension (Table 2) shows the following patterns of results. For Dimension 4 (*Argumentation and reaching consensus*), the score decreases from pairs in co-presence to groups of four designers at distance. For three dimensions (2, 6, 7), the mixed presence-distance condition with 4 designers exhibited clearly lower scores than pairs at distance as well as collocated pairs: *Sustaining mutual understanding, Cooperative orientation, Individual task orientation*. Here, the number of participants – and the intertwining between distant and collocated collaboration – could have a negative effect on these dimensions. Such a pattern suggests effects like social inhibition and social laziness as classically reported by social psychology in association to the increasing number of people within a collective. Both are affecting participation in terms of effort to contribute to the common task at hand and probability of engagement of participants. The poorest score on *sustaining mutual understanding* is also consistent with empirical results showing impact of technology-mediation on grounding (e.g., [6]).

Finally, we observed a highest score in *Dimension 1* (*Fluidity of collaboration*) for pairs at distance over both collocated pairs and groups of 4 designers. This pattern could reflect a strong management of interactions and use of tools. We will develop this point when discussing the fluidity of verbal turns patterns in the next section.

Table 2. Global scores for the three design situations.

Dimensions	Pairs in collocation	Pairs at distance	Groups of 4 at distance	All conditions mean (sd)
1. Fluidity of collaboration	4,0	4,5	4,0	4,2 (0,4)
2. Sustaining mutual understanding	4,0	4,5	3,5	4,0 (0,6)
3. Information exchanges for problem solving	5,0	4,5	4,5	4,7(0,5)
4. Argumentation and reaching consensus	5	4,5	4	4,5 (0,8)
5. Task and time management	3,5	3	3	3,2(0,8)
6. Cooperative orientation	5	5	3	4,3 (1,2)
7. Individual task orientation	4,5	4,5	3,7	4,1 (0,6)
All dimensions – mean (sd)	4,4 (0,6)	4,4(0,7)	3,7 (0,8)	4,1(0,5)

Obviously, no statistical inference can be made based on these results since our objective is mostly an exploratory test of the sensibility of the method carried out by applying it on clearly contrasted technology-mediated cooperative situations. However, as reported below, these patterns are strongly consistent with results in the literature.

Indicators Reflecting Differences in the Quality of Collaboration at Distance Compared to Co-presence

Based on the questions for each indicator, we found that specific indicators reflected a poorest quality of collaboration at distance compared to co-presence. These indicators are:

- *Coherency of attention orientation (Dimension 1)*
- *Mutual understanding of the actions in progress and next actions (Dimension 2)*
- *Attention orientation in relation with the design task (Dimension 7).*

Firstly, problems pointed out in attention orientation at a group level (*Dimension1*), are also pointed out at the individual level (*Dimension7*). Secondly, this lack of congruency in attention orientation can be related to the problems observed in the construction of mutual understanding of actions. These results reflect difficulties encountered by designers to construct a shared context and to be aligned on the task at distance. We can interpret these difficulties as related to the fractured space²¹ of interaction, in particular for the visibility at distance. Indeed visibility is fractured in two spaces: the visibility of the state of the design artefact on the virtual desktop and the visibility of the gestures (deictics...) and gazes on the videoconferencing display.

In contrast, we found that a specific indicator reflected a best quality of collaboration at distance compared to co-presence: *Fluidity of verbal turns (Dimension 1)*. In fact, there were less verbal overlaps at distance as the management of verbal turns was made verbally, e.g., addressing distant partners by their name. By contrast, in the collocated situation the management of verbal turns was based on visual cues like gazes and verbal overlaps were more frequent.

Finally, no difference was found for task related processes such as *Coherency and follow up of ideas (Dimension 3)* and *criticisms and argumentation (Dimension 4)*. This is coherent with results in our previous study, using a coding scheme [8], showing no effect of technology mediation on the design processes.

Indicators Reflecting Differences in the Quality of Collaboration in Groups of 4 Compared to Pairs

Analysing answers for each indicator, we found that specific indicators reflected a poorest quality of collaboration in the group of 4 architects compared to the pairs. These indicators are *Common decision taking (Dimension 4)*, *Constancy of effort put in the task (Dimension 7)* and all indicators related to *Cooperative orientation* (Dimension 6): *Symmetry of verbal contributions, symmetry of use of graphical tools, symmetry in task management* and *symmetry in design choices*.

Thus the groups of 4 have more difficulties than pairs to reach common decisions which is clearly related to the number factor. The results on all indicators of *Dimension 6* reveal an asymmetry of roles when they were four participants as compared to two. This asymmetry is observed at the general level of verbal and graphical contributions as well as at the

²¹ This refers to the notion of fractured ecology introduced by Luff et al. [24].

design task level (tasks management and design choices). We can wonder to which extent this asymmetry is reinforced by the technology-mediation in the groups of 4. Finally, the indicator about constancy of effort reflects unequal motivation between participants in groups of 4 compared to pairs.

Interestingly, we found that the groups of 4 at distance were rated higher than the two other situations on the following indicators: *distribution and management of task interdependencies (Dimension 5)* and *time management (Dimension 5)*. This suggests that the need for coordination, growing with the number of participants, is particularly well managed by these groups.

CONCLUSIONS AND PERSPECTIVES

The results of our empirical study show that our method reveals interesting differences in the quality of collaboration that can be related to technology mediation or group factors. On one hand, our method produces results which are consistent with those found in previous studies using either ethno-methodological approaches (e.g., [24]) or coding schemes (e.g., [8]): in particular for attention orientation, reflecting lack of alignment on local objectives, and mutual understanding of actions. On the other hand, our method goes further by showing asymmetry of roles and decrease of motivation for the groups of four architects at distance. In particular, indicators of *Dimension 6 (cooperative orientation)*, which cross the other dimensions, seem to be particularly consistent one to each other. However, the lack of a control situation with groups of four designers in co-presence prevents us to conclude that there is a combined effect of group number and technology mediation on the symmetry of roles.

Nevertheless, the actual patterns of results should be considered cautiously due to the limitedness of the sample of situations on which the current data are extracted. Indeed, a recurring problem is the cost of multiplying the experiments with designers, both at the level of required time and the ability to access to a sufficient number of participants. Furthermore we were not able to assess the design solutions and thus discuss the relation between collaboration processes and the performance in terms of design product. This question is not trivial as design product assessment is based on a multiplicity of criteria.

Our perspectives are as follows. To go further on the analysis of collaboration quality, we plan other experiments to vary group composition and technology mediation characteristics. The question of the relationship between quality of collaboration and quality of design as well as group efficacy will be explored. To go further on the development of our method, we plan to compare in a systematic way between the results obtained with this rating method and those obtained with more time-consuming coding methods. We will also explore to which extent this method can be used by judges from the design domain, the architectural design domain in our case. Finally, we will explore to which extent the method could be used as a reflective support for the group itself to improve its collaboration process.

ACKNOWLEDGMENTS

This research has been partly funded by OrangeLabs. We thank Michael Baker for his input and feedbacks on our method.

REFERENCES

- [1] Détienné, F. 2006. Collaborative design: managing task interdependencies and multiple perspectives. *Interacting With Computers*, 18(1), 1–20.
- [2] Clark, H. H. and Brennan, S. E. 1991. Grounding in communication. In: *Perspectives on socially shared cognition*, L. Resnick, J.-M. Levine and S. D. Teasley, (Eds.). Washington, DC: American Psychological Association Press. Pp. 127–149.
- [3] D’Astous, P., Détienné, F., Visser, W. and Robillard, P. N. 2004. Changing our view on design evaluation meetings methodology: a study of software technical review meetings. *Design Studies*, 25, 625–655.
- [4] Olson, G. M., Olson, J. S., Carter, M. R. and Storosten, M. 1992. Small Group Design Meetings: An Analysis of Collaboration. *Human-Computer Interaction*, 7, 347–374.
- [5] Stempfle, J. and Badke-Schaub, P. 2002. Thinking in design teams – an analysis of team communication. *Design Studies*, 23, 473–496.
- [6] Olson, G. M. and Olson, J. S. 2000. Distance matters. *Human-Computer Interaction*, 15, 139–178.
- [7] Pahl, G., Badke-Schaub, P. and Frankenberger, E. 1999. Resume of 12 years interdisciplinary empirical studies of engineering design in Germany. *Design Studies*, 20(5), 481.
- [8] Burkhardt, J.-M., Détienné, F., Moutsingua-Mpaga, L., Perron, L., Leclercq, P. and Safin, S. 2008. Multimodal collaborative activity among architectural designers using an augmented desktop at distance or in collocation. In: *Proceedings of the European Conference of Cognitive Ergonomics, ECCE 2008*. Madeira, Portugal, Sept 16–18, 2008.
- [9] Herbsleb, J. D. and Mockus, A. 2003. An empirical study of speed and communication in globally-distributed software development. *IEEE Transactions on Software Engineering*, 29(6), 481–494.
- [10] Baker, M., Détienné, F., Lund, K. and Séjourné, A. 2003. Articulation entre élaboration de solutions et argumentation polyphonique. In: *Proceedings of Deuxièmes Journées d’Etude en Psychologie ergonomique – EPIQUE*. Boulogne-Billancourt, France, 2–3 October, 2003. Pp. 235–240.
- [11] Barcellini, F., Détienné, F., Burkhardt, J. M. and Sack, W. 2008. A socio-cognitive analysis of online design discussions in an Open Source Software community. *Interacting With Computers*, 20(1), 141–165.
- [12] Spada, H., Meier, A., Rummel, N. and Hauser, S. 2005. A new method to assess the quality of collaborative process in CSCL. In: *Proceedings of Computer Supported Collaborative Learning 2005: The Next 10 Years!*, T. Koschmann, D. Suthers and T. W. Chan (Eds.). Mahwah, NJ: Lawrence Erlbaum Associates. Pp. 622–631.
- [13] Baker, M. 2002. Forms of cooperation in dyadic problem-solving. *Revue d’Intelligence Artificielle*, 16(4–5), 587–620.
- [14] Dillenbourg, P. 1999. What do you mean by collaborative learning? In: *Collaborative learning: cognitive and computational approaches*. P. Dillenbourg, (Ed.). Oxford: Elsevier. Pp. 1–19.

- [15] Pinelle, D., Gutwin, C. and Greenberg, S. 2003. Task analysis for groupware usability evaluation: Modeling shared-workspace tasks with the mechanics of collaboration. *ACM Transactions on Computer-Human Interaction*, 10(4), 281–311.
- [16] Tromp, J., Steed, A. and Wilson, J. R. 2003. Systematic usability evaluation and design issues for collaborative virtual environments. *Presence. Teleoperators and Virtual Environments*, 12, 241–267.
- [17] Hombæk, K. 2006. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2), 79–102.
- [18] Meier, A., Spada, H. and Rummel, N. 2007. A rating scheme for assessing the quality of computer-supported collaboration processes. *Computer-supported Collaborative Learning*, 2(1), 63–86.
- [19] Voyiatzaki, E., Meier, A., Kahrimanis, G., Rummel, N., Spada, H. and Avouris, N. 2008. Rating the quality of collaboration during networked problem solving activities. In: *Proceeding of 6th International Conference on Networked Learning*, Halkidi. Pp. 409–416.
- [20] Détienne, F., Burkhardt, J.-M., Hébert, A.-M. and Perron, L. 2008. Assessing the quality of collaboration in design: bridging cognitive ergonomics and CSCL approaches. Workshop « CSCW and Human Factors », CSCW'2008. San Diego, USA, November 9, 2008.
- [21] Juchmes R. and Leclercq P. 2004. A Multi-Agent System for Architectural Sketches Interpretation. *Eurographics Workshop on Sketch-Based Interfaces*, Grenoble, France.
- [22] Safin S., Boulanger C. and Leclercq P. 2005. Premières évaluations d'un Bureau Virtuel pour un processus de conception augmenté. 17ème Conférence Francophone sur l'Interaction Homme-Machine, Toulouse, France.
- [23] Safin, S., Leclercq, P. and Decortis, F. 2007. Impact d'un environnement d'esquisses virtuelles et d'un modèle 3D précoce sur l'activité de conception architecturale. *Revue d'interaction Homme-Machine*, 8(2), 65–98.
- [24] Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K. and Oyama, S. 2003. Fractured Ecologies: Creating Environments for Collaboration. *Human-Computer Interaction*, 18(1), 51–84.

Some Severe Deficiencies of the Input-output HCI-paradigm and Their Influence on Practical Design

Kai Tuuri

Department of Computer Science
and Information Systems
FI-40014 University of Jyväskylä,
Finland
krtuuri@jyu.fi

Antti Pirhonen

Department of Computer Science and
Information Systems
FI-40014 University of Jyväskylä,
Finland
pianta@jyu.fi

Eve Hoggan

Glasgow Interactive Systems
Group, Department of
Computing Science
University of Glasgow,
Glasgow, G12 8QQ, UK
eve@dcs.gla.ac.uk

ABSTRACT

The Cartesian dichotomy of human mind and body has largely ruled the development of western thought. One effect of that Cartesian legacy is the tendency to conceive interaction between a user and smart device as being composed of different inputs and outputs. In many cases, this is a practical and highly appropriate approach to design interactive technology. We, however, argue that such an approach tends to put too much emphasis on the technical instrumentation that provides information for different senses, thus considering sensory modalities as independent ‘receiver modules’. Perception is not directly created on the basis of the physical origin of the sensation. Rather, we argue that it is based on sensory-motorically integrated gestalts. For instance, a haptic feedback experience can even take place in the presence of only visual or audio cues that become coupled with interaction [1, 2]. If the concept of haptic feedback is merely understood in terms of the sense of touch and its usage with the help of, e.g. force actuator technology, it could be argued that the choice of options in user interface design is severely narrowed, and may result in the inappropriate use of available technology. By discussing the design of haptic feedback for touch screen applications, this paper illustrates the deficiencies of the input-output paradigm. It also stresses the close coupling between perception and action, which is realised in the course of interaction in a way that does not justify splitting them conceptually when striving towards a deeper understanding about human-computer interaction.

Keywords

interaction design, haptics, pseudo-haptics, multimodal interaction, embodied cognition

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *theory and methods*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

INTRODUCTION

To understand the current paradigm of human-computer interaction (HCI), it is beneficial to look at the origin of the discipline. In the era of centralised computers, interaction with them was a highly professional activity, and people dealing with computers could be educated to cope with them. However, as soon as computers were spread to the hands of everyone, it became essential to pay attention to interaction with computers. The aim was to make a computer and its user a seamless whole [3]. In the conceptualisation of HCI, basic concepts were largely derived from a technical context, paralleling the function of the computer and its user. Thus, the whole paradigm of HCI is often seen as constituting inputs and outputs between a user and a computer.

Under the umbrella of the *cognitivist* paradigm, the adoption of the computer metaphor of human cognition [4] in HCI has played an important role in the development of the discipline ever since. The computer metaphor was easily adopted on the basis of the traditional Cartesian dichotomy of mind and body, which has ruled the development of western thought for centuries. Although cognitivism has been gradually rejected by researchers of mainstream cognitive science [5], this recent development of cognitive science and the philosophy of mind has never really overdriven the temptingly simple and easy-to-understand computer metaphor in HCI. In human computer studies, the influence of the computer metaphor is salient. Even in today’s vocabulary of HCI practitioners, interaction is largely conceptualised in terms of technical devices which represent input and output modalities of interaction. The term *multimodality* thus often means the use of different technical devices, as channels of information transmission [6], in the design of HCI applications. For example, visual displays, speakers and motion actuators for corresponding output channels, or keyboards, microphones and motion sensors for corresponding input channels. Such an approach is indeed appealingly practical, and it makes the analysis and development of HCI applications straightforward.

As long as HCI mainly meant working with a desktop computer by entering information with a keyboard, and receiving information through a visual display unit, the input-output paradigm was not questioned. However, the addition of computational power to various products in our everyday life has not only broadened the scope of HCI but also challenged the appropriateness of this model of interaction. New technologies and use cultures opened a new perspective to HCI, whether it is a question of traditional desktop setting or some newer concept of utilising digital technology.

The ongoing shift towards user-centered design merits an investigation of the appropriateness of treating HCI in terms of

the technology employed in interaction. From the human point-of-view, our interaction modalities are not channels of information transmission. None of the contemporary models of human cognition suggest that we have separate sensory and motor systems for each available user-interface technology. Even though we can see screen events with our eyes, our minds do not handle the visual perception in isolation from other cognitive processes. Rather, recent studies in neuroscience suggest that our cognitive processes are based on extensive sensory-motor integrations at a neural level [7].

In this paper, we challenge the prevailing HCI input-output paradigm, by indicating clear deficiencies in its appropriateness for the conceptualisation of multimodal interaction. We present examples which illustrate design issues at a practical level, and relate these to the traditional HCI paradigm. We also discuss alternative ways to conceive multimodal interaction in terms of embodied cognition.

SAMPLE DESIGN: A VIRTUAL SLIDER

For about five years, we have conducted research projects in which the potential of non-speech sounds in HCI has been studied²². Recently, we have broadened the scope to cover haptics as well. The projects have been carried out in close cooperation with the Finnish manufacturing industry. The application areas have been extremely diverse; from wrist computers to control room technology, from tiny portable applications to heavy industrial machinery. The problems and focus have varied, but there have also been a great deal of common factors among the different applications and contexts of use. In this paper, we do not present any of these industrial cases in detail, but present an illustrative sample about the issues of conceptual design for a ‘virtual slider’ for touch screen applications. Although the discussed sample design is closely related to an ongoing case study, it also draws upon past experiences concerning many kinds of products of our partners.

Background

Touch screens are an interesting piece of interaction technology, in many senses. Their obvious advantage over mediated pointing devices, like the mouse and roller-ball, is that users can physically point at objects in graphical user-interface (GUI). Touch screens have also been successfully used without visual displays, utilising their touch sensitivity in mobile applications by substituting the visual display with an auditory one (see, e.g. [8]).

GUIs, especially when implemented in touch screen, provide a tempting opportunity to substitute tangible hardware with similar looking simulations. From the very beginning of GUIs, the appearance of basic tools has imitated their real counterparts. Word processing applications, for instance, have inherited much of what they look like and even the vocabulary, from mechanical typewriters. Sometimes, this strategy of substituting a real-world object with a somewhat similar digital application, are referred to as metaphors. It is more appropriate, however, to talk about simulation or imitation [9].

In terms of usability, the rule-of-thumb is that each function has a dedicated, hardware based control [10]. Virtualised controls, or virtual substitutes, can hardly ever be better than the real-world counterparts. However, when reasoning beyond

pure substitution, virtualised controls have advantages. First of all, in mass products in particular, their production is practically free of cost. This is probably the foremost motivation for switching from hardware based UIs to software based. Second, the virtual control can provide levels of flexibility that a real thing cannot. For instance, the UI is no longer dependent on controlling schemes based on hardware controllers. It can also be changed in terms of personal preferences or use mode.

The clearest disadvantage of virtual control elements, compared with hardware-based ones, is the lack of physical interaction with the ongoing process. For instance, with a physical slider users can feel the current position and the resistance of the slider as well as adjust it accurately. An ideal slider would combine the strengths of the physical and virtual. The design challenge of a virtual slider would thus be how to reproduce the physical experience of moving a slider. In the application which was the foremost motivation for this discussion, it was also necessary to provide such an unambiguous experience, that the slider could be adjusted without looking at it.

Desired Feedback Technology vs. Desired Interaction Experience

In interaction design, expertise from various disciplines is necessary. We have been in many brainstorming sessions, in which novel interaction ideas have been discussed among the experts of technology, human behaviour and the experts of context of use. In such sessions, visions are encouraged to be freely expressed. However, in concrete visions, technical issues tend to be the driving force.

The traditional input-output paradigm is a perfect match to the technical approach to design. In order to cause a haptic experience, some tangible, force-emitting elements are needed in the user interface. Such apparatus can thus represent the channel through which haptic information is provided to user. In the case of touch screens, a screen with an enormous number of up and down moving pins might represent the ideal scenario in enabling the haptic feedback. The issues which an engineer would be mostly interested in, would be the number of pins per inch, the available force, range and speed of movement of each pin, and the latency in milliseconds. On the basis of these specifications, the appropriateness of the technology in the given context could then be assessed.

Since it is not currently economically realistic to implement the display described above, designers must usually conform to some kind of ‘light’ version of the haptic feedback technology, such as simple vibro-tactile actuators. However, the motivation for going for such a technology, which has very limited resolution for presenting haptic information, has not been any model of human perception, but purely economical. The initial idea is just reduced to a level which is feasible to implement. That underlying idea is to physically cause events which can be perceived by the sense of touch.

In the case of the virtual slider, feedback which could somehow simulate the touch-related properties of real physical interaction would certainly be beneficial. But must we confine our design considerations to the seemingly obvious ‘haptic technology’ for information presentation? While not having the credible technology to implement physical events other than visual and audio, we were thus forced to consider our design task the other way round. We can either strive to use motorised tactile/haptic effects as part of the user-interface, or we can try to provide an appropriate overall experience for interaction with the available means. If we choose the first one, we will need to apply the given technology as appropriately as

²² GEAR, GEAR2 and GEAR3 projects funded by the Finnish Funding Agency for Technology and Innovation (TEKES).

possible. It cannot be assumed, however, that using e.g. vibrotactile actuators would result in desired interaction-relevant haptic experience. Therefore, the preferable option is to start the considerations from the other end, the desired experience. Only after having specified the qualities of the desired experience, should the technical issues be considered.

Designing Pseudo-haptics

There are a number of studies which show that designers do not need to be limited to existing, fairly primitive actuator technologies when aiming to provide a haptic experience. For instance, sounds as well as visuals have the potential to produce a haptic perception, as the coupling between a user's motor-movements in touch screen interaction and the related sonic/visual feedback can result in *pseudo-haptic* [1] illusions. Thus, auditory or visual sensations, as direct feedback for actions of the user, can modulate touch-related perceptions about the physical features of an object, like surface properties [11, 12, 13] or inertial properties such as resistance to motion [2]. The designer has the option to consider, for example, what kind of sonic feedback would modulate the haptic perception towards the desired experience, i.e., how it should sound when the screen surface is touched.

Touch screen interaction is an example of direct manipulation, in which feedback techniques can give touch-related impressions that some kind of *surrogate object* is interacted with instead of touch screen surface itself. We started conceptualising the virtual slider by exploring the propositional flow of interaction: what is being controlled with the slider and what kind of feedback experiences would be desired for such activity? In other words, we needed to determine what kind of surrogate object the user would perceive while using the virtual slider – how it should feel, and how that 'object' would function as a controller. Using the ecological approach to perception [14], we can thus utilise the directly meaningful attributes of the active object – meanings which are engaged with the user's purposeful actions, and which are based on the experiential history of interacting with the environment.

The surrogate virtual object can be directly based on the subject matter which is being manipulated via the controller. When controlling operations remotely, such as rock drilling, physical invariants of the actual drilling operation can be used as feedback cues – thus supporting feedback experiences analogous to the non-remote drilling. However, as a surrogate object, we can also utilise many kinds of virtual tool-objects which are based on a suitable metaphor that matches the real situation of manipulation. Moreover, we do not have to fully simulate any real object. We can just use certain *action-relevant* attributes of active objects. These attributions are action-relevant because they propose the occurrence or afforded potential of some activity, and are also indexed to contextual activity.

Let us assume that the function of a virtual slider is related to the manipulation of a parameter, which is to be increased until the predetermined threshold-value is reached. For such a scheme, we can use, for example, the action 'rotating an increasingly resisting wheel clockwise until it snaps into a retainer that holds it' as a model for determining what kind of feedback, mapped to the circular movement of a finger on touch screen, would relate to sensations of appropriate resistance and finally the removal of the restraint. We do not have to accurately re-create such a wheel, but use it as a propositional mental model for conceptualising the functionality of the controller. Based on that functionality of the 'wheel' and the adequate modelling of its inertial and material properties, we can discover invariant attributes that can be utilised as

action-relevant feedback cues. As physical invariants, such action-relevant cues can be potentially encoded into various modes of presentation; vibrations, sound or visuals.

In the case of the virtual slider, the general design intention was to provide a feedback experience that would indicate functional 'resisting friction' during the control process. This is possible, for example, by attributing surface roughness as a continuous, action-relevant parameter to the feedback of virtual slider. The current design is still in a conceptual phase, in which different touch screen widgets and corresponding interaction schemes or surrogate tool-objects are being considered. Pseudo-haptic feedback has offered a viable substitution to the use of force actuator technology. Because the use context of the virtual slider does not allow the gaze of the user to be focused on display all the time, audio is clearly the preferred form of feedback over visual. Audio feedback has been already proved to be effective in providing and modulating haptic experiences [2, 11, 12].

Haptics as a Perceptual System

Although the construction of the virtual slider is still in an early stage, we have already learned a lesson or two about interaction when going through different options for design approaches and technical issues. First of all, we needed to critically examine the current interaction paradigm, in which interaction is usually seen as a combination of inputs and outputs. If the hypothesis about pseudo-haptic perception holds and perceived haptic feedback can be based on sounds only, it is difficult to justify the current divide to input and output. From the point of view of the user, what is 'input' if it is physically a sound but results in a haptic experience? Moreover, in the case of haptic experiences, the user's exploratory activity is the key element in 'picking up' the information about, for example, the roughness of the surface. In other words, this kind of haptic experience is *obtained* through the active role of the user in interaction. This issue further undermines the input-output dichotomy. The finger, which is moved around the touch screen surface, could be seen as a *perceptual subsystem* because it participates in the exploratory effort of obtaining the information [15]. So, when the perception of the information also requires muscle activity and body movement, it is not easy anymore, or even appropriate, to make a distinction of which is 'input' and which is 'output'.

Of course, these presented issues, are not new findings at all. In the mid 1960's, J. J. Gibson [15] discussed them in his book 'The senses considered as perceptual systems'. He asserts that inputs for perception are not equivalent to inputs for sensation. A haptic experience is not merely based on the sense of skin pressure, or even the sum of the sensations of skin pressure and kinaesthesia. Gibson sees haptic systems as a much broader apparatus, by which the subject picks up information about both the environment and his body. He admitted that 'the simple, neat easily-remembered contrast between *receptors* and *effectors*, between *sensory* and *motor*, will have to be abandoned.' Despite this, he did not completely reject the input-output dichotomy, but tried to reformulate it to take the complexities of sensory-motor integration into an account. However, more recent, related accounts of *enactive perception* [16] and *embodied simulations* [7] have further blurred the boundaries between input and output (or perception and action), by arguing that all perception is being intrinsically 'acted out', i.e., the acquiring of perceptual experience requires the perceiver's skilful activity. Recent findings in neuroscience support these approaches by highlighting the integration of perception and motor action at the neural level [7].

The concept of interaction modalities is another issue, closely related to the discussion of this section. When talking about multimodal interaction, the design example of a virtual slider forces the question, what kind of interaction is *not* multimodal? We next consider the themes which became topical in the design of virtual slider: the nature of multimodal interaction, and multimodal approach to interaction design.

THE NATURE OF MULTIMODAL INTERACTION

In the bulk of HCI literature, interaction modalities are handled as ‘channels’ between human cognition and the environment. Whether we talk about channels or pipelines, which is another popular, related metaphor, it is undeniable, that the underlying concept is closely related to the computer metaphor (see introduction). In it, our senses and motor-systems are conceived as being directly and independently connected to the metaphorical central processing unit.

In the simple pipeline metaphor, different modalities are handled as independent resources. This basic assumption can be seen in numerous HCI studies. For example, the problem of human computer interaction in mobile contexts is typically ‘solved’ on the basis of pipeline metaphor. While on the move, users are often unable to fully focus their visual sense on the mobile device when using it. Also, mobile devices are usually so small in size that the interaction cannot be based on a tiny screen alone. Due to these limitations, more ‘bandwidth’ for human computer interaction is searched for from other modalities than vision. Hearing and touch, for instance, have been identified as free resources and therefore excellent opportunities for enhancing interaction with a mobile device.

However, it has been shown for quite some time, that interaction modalities are far from independent from each other. Extensive research work in this area has taken place in attention studies since 1950’s. The development of theories of human attention has started from simple pipeline models [17] gradually to more and more complicated theories.

The development process of the models of attention is typical in the history of science: we first have a simple theory, which is easy to adopt. When we gradually get results of empirical studies which are in conflict with the model, the model is revised to match the empirical evidence. As a result, the revised model matches better with the existing evidence, but is probably more complicated than the previous version. After a number of this kind of revisions, the theory is so complicated that it is hard to communicate to the scientific community which is supposed to apply and further develop it [18].

In the models of multimodal human-computer interaction, the idea of independent interaction channels has been challenged; not only by previous attention studies, but also by user studies with e.g. mobile applications. A prominent example is the use of mobile phones while driving. The channel metaphor suggested that the use of hands free equipment would solve the safety problem of talking on the phone on the road; since the driver does not need to talk or listen in order to control a car, those modalities were declared as free resources which could be used for phone calls. While using hands free equipment instead of a handset, the same resources for driving were supposed to be available whether there was a phone call on or not. However, the observed deteriorating of driving performance while talking in the phone, whether using a handset or hands free equipment [19], questions the channel model. Apparently,

concentrating in conversation is distracting regardless of the available telephone user-interface.

The above example is obviously closely related to the studies about human attention, the paradigm of divided attention in particular. It appears, however, that the cognitivist approach used in such studies has quite limited power to support the design of multimodal interaction. The recent studies in *embodied cognition* are a promising option for conceptualising interaction. In the embodied approach [5], the human mind is inseparable from the sensory-motor experiencing of the physical world and cognition is thus best described in terms of embodied interaction with the world. Human understanding is thus seen as arising inherently from embodied couplings with the environment. The emergence of these couplings is based on an experiential background of constant encounters and interaction with the world by using our bodies. Quite in the same line with the ideas of ecological perception, the embodied approach stresses an action-based understanding of the world, (see e.g. [14, 5, 7, 20]). From such an action-orientated perspective, multimodality and the multimodal experience appear as an inseparable characteristics of interaction. That is because the understanding of actions, as a subjective experience, is not characterised necessarily by sonic, visual or tactile qualities. In fact, actions presented in different modalities seem to produce a very similar neural basis for the understanding of an action [21]. This agrees with other findings of extensive sensory-motor integrations in neural mechanisms, which are hypothesised to contribute as a basis for action understanding [7].

The viewpoint of embodied action understanding further challenges the existence of isolated domains for sensory input and motor output. We must presume that perception is not directly created on the basis of the physical origin of the sensation. Rather, we argue that it is based on *sensory-motorically integrated gestalts*²³, which are based on intrinsic human capabilities of action understanding and recurrent patterns of embodied experiences of interacting with the environment. Such multimodal gestalts would indeed explain the phenomenon of pseudo-haptics, i.e., why a stimulus in auditory or visual sensory modality has the potential to effectively modulate haptic perception. It would also explain why concurrent stimuli in different modes of presentation result in a fused perceptual content [23], or why even stimuli of a single type of modality, for example music, can trigger multimodal completions that include imagery of body kinaesthesia, touch and even visuals [24]. Such a gestalt completion results in multimodal experiences, even when the stimulus is unimodal. In a sense this could be called *amodal* completion. But unlike the traditional cognitivist view [25], we propose that amodality is *not* symbolic in nature but inseparably bound with our sensory-motor system. Thus the resulting meaningful experience is not seen here as modality independent, but essentially as engaged with the human body and all of its modalities of interaction. Therefore it might make sense to call these mental completions multimodal, rather than ‘amodal’.

On the basis of what we have discussed so far, it appears that all interaction is inherently multimodal. Therefore multimodality is not something that designers can implement in applications. Rather, multimodality is the nature of interaction that designers must take as a starting point and acknowledge the embodied situation as a whole. Hence, such a multimodal approach to interaction design [26] should not focus on different presentation (or input) modalities or on any communication technologies in themselves.

²³ Such gestalts refer to the theories of, for instance, embodied image schemas [22, 20] and embodied simulations [7].

THE CONCEPT OF CROSSMODAL INTERCHANGEABILITY

One traditional rationale for using multiple modalities in user interfaces is to provide choices for the user. It has been proposed that when a message is coded in the user interface in multiple forms, the user should be given the opportunity to receive the message in a form most suitable for her given the situation or context [27]. It has been argued that, by allowing the user to choose from a selection of modalities, different capabilities or cognitive styles can be taken into account. In this kind of design strategy, redundancy is a central concept: information is presented redundantly in various communication channels in order to assure a correct reception of a message.

The strategy described above, often presupposes an unambiguous segmentation of form and the content. In other words, this presupposition means that there would be interchangeable interaction modalities, i.e., a given piece of information should be able to be presented equally in different modalities. For example, rhythms can be created identically in both audio and vibrotactile feedback. Previous work in crossmodal transfer has shown that, if trained to understand multidimensional audio alerts, a user can then also understand the corresponding tactile alerts with no additional training and *vice versa* [27]. In this case the mappings between information and feedback are abstract and have to be learned.

In some other real-world situations, crossmodal presentation of information is also somewhat successful; for instance, in pedestrian traffic lights in which the message is either ‘go’ or ‘don’t go’, this simple message can probably quite appropriately be argued to be communicated equally by a colour, and an icon, frequently by a sound as well (at least in Scandinavian style lights). However, it is as well common sense as a result of theoretical reasoning, that different forms of presentation deliver qualitatively different kind of information [28]. Therefore, the idea that modalities would be universally interchangeable is perhaps not so appropriate in certain situations. In the example of traffic lights, the encoding and decoding of the messages requires a great deal of conventionality and habit-based knowledge. Without that backlog of knowledge, in the case of the traffic lights, it could be hard to describe what a ‘red colour’, a ‘figure of a standing man’ and ‘short, repetitive beeps’ have in common. As mentioned in the previous work, to use crossmodal feedback effectively, some user training is required [27]. This, in particular, applies to symbolic/arbitrary mappings between information and feedback. However, the utilisation of *ecologically valid* mappings, designed to agree with our existing experiential knowledge, should diminish the need for training.

The underlying idea of interchangeable modalities is that there first exists an amodal meaning. Communication of such a meaning would then involve the implementation of certain amodal, i.e., modality independent attributes in the chosen presentation modality. The traditional cognitivist approach considers amodal information as essentially symbolic [25], thus implying an arbitrary relationship between amodal meaning and its form of expression. However, in the light of embodied approach to human cognition and the recent neuroscientific findings, such a view on amodality has been strongly criticised [7]. Designers can successfully utilise symbolic attributes (such as information encoded in rhythms) in the design of crossmodal feedback [27], but it does not mean that crossmodal interchangeability (e.g., crossmodal recognition of rhythms) would be based on symbolic or abstract processes. As discussed briefly in the previous section, we propose that amodality

should be seen in relation to embodied sensory-motoric integrations that takes place in the course of perception and thinking. These integrations are realised in neural mechanisms – such as mirror neurons [29] – which have a role in contributing action understanding. For example, crossmodal rhythm perception can arguably be based on motoric attuning to the observed rhythmic impulses [30]. As these enactive perceptual processes should occur in a sensory-motorically integrated substrate [7], it does not matter whether it was the sonic, tactile or visual features of the observed feedback that allowed the ‘motor-mirrored’ amodal understanding of the rhythm. But can we consider such ‘amodal’ understanding modality independent then? Not entirely, because it would still be dependent on the motoric perception and its intergration to the senses.

There is no any reason why we should not consider ‘amodality’ as being engaged to the body and its means to interact with the environment. For example, in our case of designing pseudo-haptics, we could call the perception of surface roughness or friction amodal. Indeed it seems that the attribution of surface texture can be equally based on touch, sound or visuals. We argue, however, that even in this case the desired meaning is highly dependent on interaction, in particular, the ecological validity of the presented artificial feedback events. Therefore the key element in the haptic experience is the intentional activity of the perceiver, who acquires the information by moving her finger on the touch screen surface. It is still true, though, that the certain invariant properties of a modelled object, which indicate the roughness during the interaction, can be equally manifested through a variety of modalities. But perhaps the next step in crossmodal interaction research is to acknowledge the inevitable dependencies of such crossmodal interchanges on the nature of multimodal interaction. As Gibsonian approaches have already underlined, it is almost impossible to keep perception fully separated from interaction. Thus, the perceptual content is unavoidably affected by our intentional activity.

In terms of embodied cognition, it is difficult to argue that content can be fully separated from its form. According to the embodied perspective, the conception of mind is bound to the physical body, or even beyond [31]. For instance, writing with a pen on paper is not just a reflection of human thought, resulting in words. Rather, the act of writing itself is essential part of both expression and thinking. As an articulatory process, writing also results in nuances of form that cannot be understood in terms of verbal semantics. Some theorists go even further by conceptualising the physical result of e.g. writing, as part of the human mind [31].

CONCLUSION

Interaction between users and devices is traditionally viewed as a combination of inputs and outputs, in accordance with the computer metaphor of the human mind. This paper has suggested that such an approach leads to an emphasis on technical instrumentation with each sensory modality simply viewed as a receiver module. The argument presented in this paper focuses on sensory-motorically integrated gestalts. In other words, there is a close relationship between perception and action, which occurs in the course of interaction and therefore, there is no justification in splitting these into two separate concepts. The case of pseudo haptics highlights this issue because it has been shown that a haptic feedback experience can take place without the presence of physical haptic cues but in the presence of only visual or audio cues that become coupled with interaction. It may even be more

appropriate to replace the term *pseudo-haptics* simply with *haptics*, given that the main difference between this and ‘real’ haptics is related to the available sensory stimuli, not the resulting experience.

In order to demonstrate that our arguments are not just academic ‘hairsplitting’ about concepts, we wanted to show that the prevailing input-output paradigm has impact on practical designs. Established dispositions towards conceptualising interaction in oversimplified technical terms can effectively hinder the design potential for multimodal interaction, as the sample design of virtual slider revealed.

The underlying idea of this paper is to stress the human-centeredness in the cost of technology. The strength of the input-output paradigm is its consistency with the structure of technical devices. By conceptualising interaction in technical terms, it is easy to analyse a system as a whole consisting of a technical system and its user. From this point of view, the aim of user interface design is easily biased to view the user as a part of the system, and make her behave as effectively as possible in terms of the task. However, different products which have very similar interfaces in terms of usability and functionality can still differ dramatically in terms of use experience [32]. Perhaps, a design approach genuinely focusing on the users point of view, would lead to the successful design of a desired experience – which, of course, does not rule out effective task completion.

The discrepancy between the application of traditional information-centered models of human cognition and how users experience the use of technology has already been widely acknowledged. Recent trends in HCI research stress the role of user experience in understanding HCI, and the importance of the embodied interaction in understanding the human mind [33]. The shift of focus from the design of efficient HCI onto sketching experiences should inevitably change the way we conceptualise interaction. By questioning the boundaries between input and output and boundaries among different modalities, we hope to contribute to the efforts of adopting more appropriate concepts for interaction design.

ACKNOWLEDGMENTS

This work is funded by Finnish Funding Agency for Technology and Innovation, and the following partners: GE Healthcare Finland Ltd., Suunto Ltd., Sandvik Mining and Construction Ltd. and Bronto Skylift Ltd.

REFERENCES

- [1] Lécuyer, A., Coquillart, S., Kheddar, A., Richard, P. and Coiffet, P. 2000. Pseudo-haptic feedback: Can isometric input devices simulate force feedback? In Proc. IEEE Int. Conf. on Virtual Reality, 2000. Pp. 83–90.
- [2] Avanzini, F. Serafin S. and Rocchesso, D. 2004. Friction sounds for sensory substitution. In: Proceedings of ICAD 2004, Sydney.
- [3] Card, S. K., Moran, T. P. and Newell, A. 1983. The psychology of human-computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.
- [4] Gardner, H. G. 1985. The mind’s new science. New York: Howard Gardner.
- [5] Varela, F., Thompson, E. and Rosch, E. 1991. The Embodied Mind. Cambridge, MA: MIT Press.
- [6] Shannon, C. E. and Weaver, W. 1949. The Mathematical Theory of Communication. Urbana, IL: The University of Illinois Press.
- [7] Gallese, V. and Lakoff, G. 2005. The Brain’s Concepts: The Role of the Sensory-Motor System in Reason and Language. Cognitive Neuropsychology, 22, 455–479.
- [8] Pirhonen, A., Brewster, S. and Holguin, C. 2002. Gestural and audio metaphors as a means of control for mobile devices. In: Proceedings of CHI2002, 20-25 April 2002, Minneapolis, Minnesota. Pp. 291–298.
- [9] Pirhonen, A. 2005. To simulate or to stimulate? In search of the power of metaphor in design. In Pirhonen, A., Isomäki, H., Roast, C. and Saariluoma, P. Future Interaction Design. London: Springer Verlag. Pp. 105–123.
- [10] Norman, D. A. 2002. The design of everyday things. New York: Basic Books.
- [11] Jousmäki, V. and Hari, R. 1998. Parchment-skin illusion: sound-biased touch, Curr. Biol. 8, 190.
- [12] Guest, S., Catmur, C., Lloyd, D. and Spence, C. 2002. Audiotactile interactions in roughness perception, Exp. Brain Res. 146, 161–171.
- [13] Mensvoort van, K., Hermes, D. J. and Montfort van, M. 2008. Usability of optically simulated haptic feedback. Int. J. Human-Computer Studies, 66, 438 – 451.
- [14] Gibson, J. J. 1979. The ecological approach to visual perception. Boston, MA: Houghton Mifflin.
- [15] Gibson, J. J. 1966. The senses considered as perceptual systems. Boston, MA: Houghton Mifflin.
- [16] Noë, A. 2004. Action in Perception. Cambridge, MA: MIT Press.
- [17] Broadbent, D. E. 1958. Perception and Communication. London: Pergamon.
- [18] Kuhn, T. S. 1970. The structure of scientific revolution. Chicago, IL: University of Chicago Press.
- [19] Horrey, W. J. and Wickens, C. D. 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Human Factors 48(1), 196 – 205.
- [20] Johnson, M. and Rohrer, T. 2007. We Are Live Creatures: Embodiment, American Pragmatism and the Cognitive Organism. In: Zlatev, J., Ziemke, T., Frank, R. and Dirven, R. (Eds.). Body, Language, and Mind, Vol. 1, Berlin: Mouton de Gruyter. Pp. 17–54.
- [21] Kohler E., Keysers C., Umiltà M.A., Fogassi L., Gallese V. and Rizzolatti G. 2002. Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846–848.
- [22] Johnson, M. 1987. The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason. Chicago, IL: University of Chicago.
- [23] McGurk, H., and MacDonald, J. 1976. Hearing lips and seeing voices. Nature, 264, 746–748.
- [24] Godøy, R. I. 2003. Motor-Mimetic Music Cognition. Leonardo, 36(4), 317–319.
- [25] Fodor, J. A. 1975. The Language of Thought. Cambridge, MA: Harvard University Press.

- [26] Tuuri, K., Pirhonen, A. and Väkkinen, P. Bodily Engagement in Multimodal Interaction: A Basis For a New Design Paradigm? In: Kurkovsky, S. (Ed.). *Multimodality in Mobile Computing and Mobile Devices: Methods for Adaptable Usability*. Hershey, PA: IGI Global. (In press).
- [27] Hoggan, E. and Brewster, S. A. 2007. Designing audio and tactile crossmodal icons for mobile devices. In: *Proceedings of ACM International Conference on Multimodal Interfaces 2007 (Nagoya, Japan)*. ACM Press. Pp. 162 – 169.
- [28] McLuhan, M. 1966. *Understanding Media: the Extensions of Man*. New York, NY: New American Library.
- [29] Rizzolatti, G. and Craighero, L. 2004. The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169 – 192.
- [30] Leman, M. 2008. *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.
- [31] Clark, A. *Supersizing the mind*. 2008. New York: Oxford University Press.
- [32] Buxton, B. 2007. *Sketching User Experiences*. San Francisco, CA: Elsevier.
- [33] Dourish, P. 2001. *Where the Action is: The Foundations of Embodied Interaction*. Cambridge, MA: MIT Press.

Designers, Users and a Shared Understanding of the Product: A Case Study with Product Designers

Anna Mieczakowski, Patrick Langdon, P. John Clarkson
Engineering Design Centre
University of Cambridge
Trumpington Street, Cambridge
CB2 1PZ, United Kingdom
{akm51 | pml24 | pjc10}@eng.cam.ac.uk

ABSTRACT

This paper describes the results of a detailed study which included a literature review and twenty semi-structured interviews with product designers from a large telecommunications company. The company studied has a lot of experience in product design and consistently produces high quality, usable and competitive products. In essence, the study investigated the nature and structure of users' cognitive representations of products and the ways in which designers currently go about matching their intended design of products with the users' understanding of those products. The findings from the study indicate that designers have often very little time, limited financial resources, and not enough support to take notice of users' understanding of products as much as they would like to. Moreover, no appropriate tool for predicting inclusive interaction between products and users is currently used across the organisation. However, the interviewed designers expressed high interest in using such a tool. Further research will evaluate existing tools for modelling the match between the conceptual models of designers and users and find an appropriate tool for facilitating inclusive interaction.

Keywords

inclusive design, universal design, mental models, product-user interaction, interaction design

ACM Classification Keywords

H.5.2 [User Interfaces]: User Centered Design.

INTRODUCTION

Understanding users' capabilities, needs and expectations is a central tenet of the inclusive design philosophy. Inclusive design aims to optimise design for maximum accessibility and minimise the effort required in using products, and focuses on achieving "the balance between the marketability of the product and the level of population exclusion" [12]. Research shows that by considering the full diversity of users, irrespective of age or ability, designers can create more accessible and usable

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

products for everyone [20]. Although, the most valid method for gathering user information and feedback will always be through having users interact with the product in question. A variety of restrictions, including time and cost constraints, are often imposed on product designers, and as a result, designers do not include heterogeneous users in designs, or involve them too little or too late [8]. Consequently, in order to design products for as many people as reasonably possible, designers require tools for predicting difficulty and exclusion which users with a full range of capabilities experience during interaction with products [18]. A wide range of methods have already been developed in order to increase designers' comprehension of users' capabilities and needs [21]. However, thus far many of those methods have had limited uptake in design practice due to, among other things, a poor fit between the structure of many methods and the ways in which designers think and work, as well as a lack of quick-to-use and understandable guidance that would raise designers' awareness on how people interpret and use different interface features [8]. Accordingly, the new tools need to go beyond the standardised checklists and textual guidelines and be represented in a visual form [7]. A review of the literature on design practice and interviews with product designers were conducted in order to investigate how the match between the intended design of products and the users' understanding of how those products work is currently predicted and how it can be facilitated. The use of a couple of research methods allowed the study findings to be cross-checked against each other and provided more objective results [5].

LITERATURE REVIEW

A review of the literature on the nature and structure of users' cognitive representations of products and designers' work practices was conducted. This review provided valuable information about the experiences of many more designers and users and gave useful background to a qualitative study.

Norman [16] postulates that there are three fundamental elements of every product: (1) the intended design of the product; (2) the user's understanding of the product; and (3) the product's appearance and operation. Based on this assumption, it is further argued that in order to design more usable and accessible products, a better match between the intended design of products and the users' understanding of products is needed [6]. Before moving on to the results of 20 semi-structured interviews with product designers, it is first necessary to refer to the wide and varied literature on mental models and design practice to investigate the formation of people's internal representations of products and the ways in which user information and feedback is currently used in the design process.

Mental Models

Research shows that design has a significant effect on people's ability to understand the functionality, appearance and behaviour of different product features and thereby inhibit effective interaction [14]. Norman [17] believes that an accessible and usable product can be designed by matching the designer's conceptual model with the mental model of the user. Accordingly, to create more inclusive products designers need to have a better understanding of the ways in which people "approach, explore, or interact" with products [13]. Two of the early precursors of intuitive interaction, Craik [2] and Johnson-Laird [11], postulate that each individual constructs a mental model of every object in environment in order to understand it. Norman [16] agrees with these assertions and adds that mental models are constructed through observations, feedback provided by products after manipulation. Consequently the structure of such models is dependent on the amount of prior experience with products. In addition, it appears to be difficult to capture the structure of mental models as they tend to be incomplete, limited and unstable. Although Freudenthal [6] does not propose any specific techniques for representing mental models of users, she believes that information about mental models could be potentially used to derive appropriate design properties that would contribute to the design of more accessible and usable products. Persad et al. [19] suggest that an effective representation of users' mental models should include information on different features of a given product and how they work, as well as information about the sequences of actions between an initial state of a given product and the goal state of the user. It is beyond the bounds of this paper to thoroughly discuss the different ways in which the mental models of users can be represented and matched with the designers' conceptual models of products; as more information about this can be found elsewhere [15].

Design Practice

There is a tendency among designers to design products for people like themselves by relying on their own intuition, experience and self observation [9]. Many designers also rely heavily on the user information supplied by the client in spite of that information being often of limited and dubious quality. Furthermore, poor consideration of users' capabilities and needs during design process can be attributed to the fear that some inclusive design methods may constrain creativity [8]. Crilly et al. [3] lists a number of organisational, technical, financial and legislative constraints, which prevent products from being designed and manufactured according to initial intentions, or to a standard that meets accessibility and usability guidelines. While, Krippendorff [13] notes that although designers are very unlikely to "see the world with the same eyes" as users, they should be able to design products that "make sense" to the users and self-contain instructions as to their use. For example, Langdon et al. [14] suggest that products can be made more inclusive by way of decreasing the cognitive demand they place on users, which in turn can be achieved by basing the products' function, appearance and behaviour on well-known and well-learned designs. Likewise, Blackler [1] believes that prior experience plays an essential role in facilitating intuitive interaction with products and so designers should bear that in mind when they sit before their drawing boards or computer screens. The researchers of the continuum of knowledge sources, Hurtienne and Blessing [10], add that user interfaces that tap on subconscious use of primitive linguistic schema in prior knowledge are more intuitive to use. However, Docampo-Rama [4] stresses that different generations of users have varied frequency and level of exposure to technology and

the range of skills they have available to deploy. For instance, the results of Docampo-Rama's [4] study on users' technological familiarity show that modern symbols and layered (multi-window and menu) computer interfaces are more familiar and suited to the interactional processes of those 25 years and younger.

METHODOLOGY AND ANALYSIS

A qualitative study was conducted at a large telecommunications company located in the UK in order to investigate whether designers transfer users' feedback into the design of product features, and find out whether they are making use of different types of supportive materials to help them match their understanding of the product with the users' understanding of the product in question. The interview is a well-established method, particularly within the social sciences, for the elicitation of information from a study participant. Semi-structure style of interviewing was found to be the most suitable method for acquiring first hand information about the consideration of users and use of methods and tools throughout the design process, as well as the practical constraints on the design process.

Interview Procedure

Prior to the interviews, consent forms that complied with ethical guidelines and non disclosure agreements were signed between the researchers and the telecommunications company. A list of topics was generated before each interview. The interviews with designers centred on three main areas:

1. The design process, in order to find out more about the workflow at each of the design stages and the different stakeholders involved in the process;
2. The designers, so as to investigate if designers transfer users' feedback into design and whether they use any supportive tools to match their understanding of products with the users' understanding of those products;
3. The users, with the view of finding out when and how data about user capabilities is collected and how much of that data is fed back into the design of products.

The semi-structured style of interviewing allowed the interviewer to focus upon areas of special interest within a particular session, but at the same time, it gave interviewees the opportunity to discuss other areas which they strongly felt were relevant to their work.

Sample

Access to designers from the telecommunications company was opportunistic as it was granted on the basis of an ongoing collaboration between the researchers and the telecommunications company. Out of twenty interviewees seventeen were males aged between 23 and 62 and three were females aged between 35 and 57. The interviewees had different levels of education and over five years of experience in product design. The interviews were approximately one hour long and were conducted in different parts of the UK. Each interview was recorded using an mp3 recorder and subsequently transcribed.

Data Analysis

The data analysis activities began with the preparation of detailed interview transcripts. This allowed the application of a general inductive approach for devising main codes inherent in the collected raw data. Once the spreadsheet with the codes

was compiled, it was analysed for emerging patterns using hierarchical cluster analysis. Sections 3.3.1 and 3.3.2 provide details of both analyses.

General Inductive Analysis

The general inductive analysis provides “straightforward” and “non-technical sets of data analysis procedures” and, unlike other qualitative methods, it bases its theory building on “the most important categories” [22]. Although a range of different qualitative methods, including grounded theory, phenomenology and discourse analysis, were considered for the analysis, the general inductive approach was found to be the most appropriate for the merit of this study as it focuses on finding the dominant and repetitive codes inherent in the collected raw data. Accordingly, general inductive approach is considered as very reliable as it uses straightforward inspection and data reduction coupled with revision and refinement of categories whenever new material emerges.

Transcripts from the interviews with designers were closely read several times to consider the multiple meanings that were inherent in the text. Subsequently, text segments containing meaningful units were identified and a label for the general category of that segment was assigned. After a thorough analysis, the existing segments were subdivided into more detailed codes. For example, a segment initially assigned with the ‘cost-related’ label was later subdivided into ‘low cost’, ‘design reuse’ and ‘feature and cost comparison’ codes etc. The analysis also allowed the data to be annotated with researchers’ own interpretations and comments. The adequacy of the chosen segments and codes was revisited and refined after multiple readings of the transcripts and whenever any new material emerged [22]. Overall, forty-six overarching codes were identified and organised in a tabular form in a spreadsheet. Some examples of codes included ‘time restrictions’, ‘low cost’ and ‘capability assessment’ etc. The spreadsheet with the codes was subsequently run through statistical analysis software (SPSS) and checked for patterns using hierarchical agglomerative cluster analysis.

Hierarchical Cluster Analysis

Hierarchical agglomerative cluster analysis was performed on the aforementioned codes in order to organise them into meaningful structures [23]. In particular, the codes were analysed using Ward’s and Complete Linkage methods. Ward’s method is commonly used for minimising the increase variance caused by linking, whereas Complete Linkage method causes linking with the furthest object from a cluster. The hierarchical agglomerative analysis dendrogram compiled using Ward’s method is shown in Figure 1. This dendrogram shows the distances at which linkage occurs for all the variables on the basis of the participants’ scorings. This was calculated from cases that represent the designers that took part in the study and the variables that signify numbers for codes identified using general inductive analysis. The diagram has a detailed hierarchical structure showing how the clusters are related to each other at different distances. The following clusters were identified: (1) cost-effective reuse of previous designs; (2) strict adherence to company guidelines and legislation; (3) prioritisation of technical integration before usability; (4) a significant lack of one unified tool that helps designers to match their understanding of products with those of users; and (5) ad-hoc use of external resources. Both the Ward’s and Complete Linkage clusters and the relationships between them were found to be extremely stable indicating that the interviewed designers held a strong, common view of the company’s design activities.

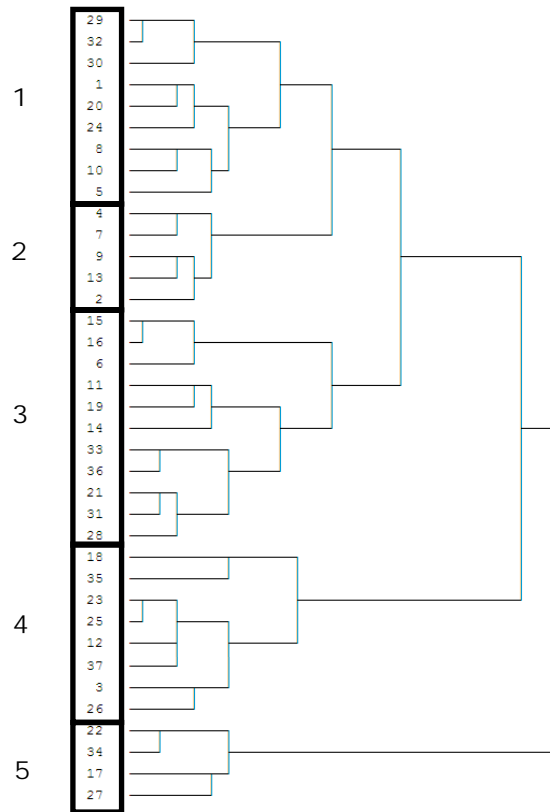


Figure 1. Hierarchical agglomerative analysis dendrogram compiled using Ward’s Method.

Results

The findings from the industry study indicate that the company follows the ‘70: 20:10’ principle during the design process, with 70% of the product being reused from previous designs, 20% being configured during current project and 10% having a new capability added, and as a result a limited amount of user capability data is channelled into the design of products. The study also found that designers need to strictly adhere to the company’s design guidelines and legislation, which means that the consistent use of any support tools has to be approved across the whole company structure. There was a strong, consistent opinion among the interviewed designers that the company champions user centred design and user data is collected during the concept and the test phases through focus groups and a usability testing. However, the designers also said that, due to time and cost restrictions, the main priority of the company is to firstly make sure that all the functional parts of products are well integrated and work as intended before checking whether products are accessible and usable to people. Although the company uses visual means in the form of scenarios (imagined sequences of events describing users and user situations) and prototypes (both computer-generated 3-D models and made by hand using paper, foam and wood) to communicate ideas and themes among all stakeholders, and occasionally uses the advice from external experts and field trialists; there is need for one unified tool that enables all the stakeholders to match their intended designs of products with users’ cognitive representations of products. Most of the interviewed designers expressed high interest in using a tool that would enable them to better understand users’ goals and actions and evaluate the accessibility and usability of product

features. In addition, many designers thought that such a tool would be useful if it had a cost factor embedded in it because as one of the interviewees noted “cost drives everything”. The cost factor was said to be valuable if it provided a quick breakdown of how many users would be excluded from product usage if the cost of product materials was decreased.

CONCLUSION AND FURTHER WORK

The findings from this study suggest that designers have often very little time, limited financial resources, and not enough support to take notice of users’ understanding of products as much as they would like to. Though, they consistently admitted that they would like to have a better understanding of the nature and structure of users’ internal understanding of products and match those representations as closely as possible with the conceptual design of products. The results of 20 interviews show that currently designers do not use any tools that would support them in matching their intended design of products with the users’ understanding of those products. However, most of the interviewed designers admitted that they would be highly interested in using such a tool as long as it was visual, quick to use and in some form facilitated the comparison between the choice of usable features and a cost effective analysis. Further research will evaluate existing tools for modelling the match between the conceptual models of designers and users and finding an appropriate tool for facilitating inclusive interaction.

ACKNOWLEDGMENTS

This work has been funded by the UK Engineering and Physical Sciences Research Council (EPSRC). The authors would also like to thank all designers that took part in the study.

REFERENCES

- [1] Blackler, A. 2006. *Intuitive Interaction with Complex Artefacts*. Doctoral Thesis. School of Design, Queensland University of Technology.
- [2] Craik, K. J. W. 1943. *The Nature of Explanation*. Cambridge University Press.
- [3] Crilly, N. 2005. *Product Aesthetics: Representing Designer Intent and Consumer Response*. Doctoral Thesis, The University of Cambridge.
- [4] Docampo-Rama, R. M. 2001. *Technology Generations handling complex User Interfaces*. Doctoral Thesis, TU Eindhoven.
- [5] Flick, U. 2006. *An Introduction to Qualitative Research*. Sage.
- [6] Freudenthal, A. 1999. *The Design of Home Appliances for Young and Old Consumers*. Doctoral Thesis, Delft University Press.
- [7] Goodman, J., Langdon, P. M. and Clarkson, P. J. 2006. *Providing Strategic User Information for Designers: Methods and Initial Findings*. In: Clarkson, P. J., Langdon, P. and Robinson, P. (Eds.). *Designing Accessible Technology*. Springer.
- [8] Goodman-Deane, J. Langdon, P. and Clarkson, P. J. *Key Influences on the User-Centred Design Process*. *Journal of Engineering Design*, 2009, in press.
- [9] Hasdoğan, G. 1996. *The role of user models in product design for assessment of user needs*. *Design Studies*, 17(1), 19–33.
- [10] Hurtienne, J. and Blessing, L. 2007. *Metaphors as tools for intuitive interaction with technology*. *metaphorik.de*, 12, 21–52.
- [11] Johnson-Laird, P. N. 1981. *Mental Models in Cognitive Science*. In: Norman, D. A. (Ed). *Perspectives on Cognitive Science*. Erlbaum.
- [12] Keates, S. and Clarkson, J. 2003. *Countering Design Exclusion: An Introduction to Inclusive Design*. Springer-Verlag.
- [13] Krippendorff, K. 1989. *On the essential contexts of artifacts or on the proposition that design is making sense (of things)*. *Design Issues*, 5(2), 9–39.
- [14] Langdon, P. M., Lewis, T. and Clarkson, P. J. 2007. *The effects of prior experience on the use of consumer products*. *Universal Access in the Information Society*, 6(2), 179–191.
- [15] Mieczakowski, A., Langdon, P. and Clarkson, P. J. 2009. *Modelling Product-User Interaction for Inclusive Design*. In: Stephanidis, C. (Ed.). *Universal Access in HCI, Part I*. Springer-Verlag.
- [16] Norman D. A. 1986. *Cognitive Engineering*. In: Norman, D. A. and Draper, S. W. (Eds.). *User Centered System Design: New Perspectives on Human Computer Interaction*. Erlbaum Associates.
- [17] Norman, D. A. 2002. *The Design of Everyday Things*. Basic Books.
- [18] Persad, U., Langdon, P. and Clarkson, J. 2007. *Characterising user capabilities to support inclusive design evaluation*. *Universal Access in the Information Society*, 6, 2, 119–135.
- [19] Persad, U., Langdon, P., Clarkson, P. J. and Brown, D. 2007. *Cognitive Scales and Mental Models for Inclusive Design*. In: Stephanidis, C. (Ed.). *Universal Access in HCI, Part I*. Springer-Verlag.
- [20] Ricability. 2001. *Easier living: a guide for older and disabled people living in London*. <http://www.ricability.org.uk/reports/pdfs/easierliving.pdf>.
- [21] Stanton, N., Hedge, A., Brookhuis, K., Salas, E. and Hendrick, H. (Eds.). 2005. *Handbook of Human Factors and Ergonomics Methods*. CRC Press.
- [22] Thomas, D. R. 2006. *A general inductive approach for analyzing qualitative evaluation data*. *American Journal of Evaluation*, 27(2), 237–246.
- [23] Tryon, R. C. 1939. *Cluster analysis*. McGraw-Hill.

The Expert Community Staff: An Innovative Method for Capturing End-users' Needs

J. M. Christian Bastien, Éric Brangier, Jérôme Dinet,
Javier Barcenilla, Gabriel Michel, Robin Vivian
Université Paul Verlaine – Metz

2LP-ETIC, EA 4165, Expériences utilisateurs dans le Traitement des Interactions technologiques
et des Conduites humaines et sociales
UFR Sciences Humaines et Arts – BP 30309 – Île du Saulcy
57006 Metz, France
Brangier@univ-metz.fr

ABSTRACT

The aim of this paper is to present the “Expert Community Staff” method for capturing end-users’ needs. The elaboration of the method was based on theories of social construction of technology and on social construction of users’ needs. This method was applied to the redefinition/redesign of the European Navigator (www.ena.lu), a digital library that provides research and educational material on the history of European integration. More specifically, the present paper describes the different steps of the method and provides illustrations and results of its application.

Keywords

digital libraries, expert community staff, user needs, user inputs, user-centered design, user profiles

ACM Classification Keywords

H.1.2 [User/Machine Systems – Human Factors]; H.3.7 [Digital Libraries – User issues]; H.5.2 [Information Interfaces and Presentation – User centered design]; I.3.6 [Methodology and Techniques – Ergonomics].

INTRODUCTION

To design digital libraries (DL) so as to satisfy the end-users implies that you center the design on those users. But how do you center the design on the users when you don’t know who they are? What user centered methods do you use in order to integrate them in the design of the DL? When designing an information system or application for a specific population (employees of a specific company, sales representatives, air traffic controllers, etc.), users are generally accessible. In theory, it would be possible to integrate all of them at different moments of the design life cycle. But when the content of a DL focuses on a specific topic that may be of interest to a large population of users, things become very difficult. This is the kind of situation the authors of the present paper faced when

trying to improve the European Navigator (www.ena.lu), a DL that provides research and educational material on the history of European integration.

These difficulties are probably responsible for the way DL have been design so far. For some, just building DL would encourage users to come and to use them [18]. But designing DL in such a way has negative consequences: digital libraries and digital work environment are underused and users encounter many difficulties [16]. In a recent study, results showed that almost 76% of users did not accomplish correctly all the simple tasks they were invited to accomplish on a general public DL [10]. These kind of problems have been identified many years ago and some authors have emphasized the importance of meeting the goals and objectives of end-users [e.g., 9]. Instead of being concerned with end-users, many studies have been concerned with librarians and archivists [1, 2, 6, 7, 8].

However, some methodologies have been used in order to gather information on the end-users so as to identify their needs. Users logs, for instance, have been used in this regard [e.g., 5, 12]. This technique allows the collection of large volumes of interaction traces that are then analyzed in such a way as to allow the identification of users’ interests [e.g., 13, 14]. But, even if those studies are undeniably interesting, they only provide us with indirect user needs. We are still far from real user needs. Only a few studies combine log analysis with users interviews [e.g., 15].

In the information sciences, researchers have been interested in user needs, but essentially through questionnaires and interviews [2, 6, 7].

The usability problems highlighted in many studies echo the utility problems of DL. Those problems are directly related to the misunderstanding of users’ needs. But again, how can we have a better understanding of those needs? From a methodological point of view, three strategies are possible: (1) to turn users’ comments and verbalizations into users’ needs. This strategy implies that users have needs and that they know how to verbalize them; (2) to deduce users’ needs from task and usage analysis; (3) and another approach that assumes that users’ needs can’t exist without a social context that triggers their expression. For proponents of the latter approach the aim is to produce methods that allow the definition of social situations in which users’ needs are socially constructed, presented, shared and discussed.

Besides the techniques that have been briefly presented previously and that we have criticized [3], other techniques offer a more important place to the expression of end-users. These participatory techniques are numerous: brainstorming,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

focus groups, storytelling, and consensus meetings... These techniques are based on the idea that to design a good product, users must be integrated in the design process. To do so, designer: (a) work together with end-users to find a design solution, (b) give users an important place in the design process, and (c) allow everyone to express their ideas. By doing so (d) designers increase the probability that the design solution will be usable and useful, and (e) increase the probability that the design solution will be accepted and adopted. The designers will also: (f) meet with other potential end-users, (g) provide a forum for the expression of problems, (h) reduce the time necessary for the development of the services, (i) ensure the compatibility of the services because the end-users are already taking part in the design of the technology, (j) maintain the motivation through out the project by associating the end-users, (k) acknowledge the essential role of the user/consumer/expert and finally (l) facilitate organizational changes when new technologies modify our or the end-users' lifestyles.

However, calling together end-users is not sufficient for producing significant input for design. These meetings participatory methods must be instrumented and conducted carefully. The question of how to manage this participation becomes crucial. What follows is a presentation of a user-centered method allowing the expression of user needs based on its social construction.

METHODOLOGY

With participatory and creative methods, the construction of the users' needs is envisaged as a process and a result of a complex collective activity produced within a social situation coordinated by a moderator/facilitator [11, 17]. Here we propose the expert community staff method. It takes the form of several groups of experts, legitimately recognized as being representative of communities of practices. The experts are involved in the project and speak together on the subject of the project. The discussions within the groups are lead by a facilitator which tries to have every member of the group verbalize ideas, discuss the ideas in order to refute or validate the group's productions. To do so, the facilitator uses different tools and techniques such as paperboards, computer screens, mockups, and storyboards. The expert staff method aims at confronting the inter-subjectivity of the group experts in order to generate, as much as possible, ideas, needs, functions, and representations of the systems. The general organization of the gathering process that lies within the staff method rests on five steps described bellow.

The Definition of the Community of Practices

Instead of making up focus groups with representatives of the population at large or the general public, expert community staffs begin with an overview of the communities of practices. The aim of this task is to gather a maximum of information on the people likely to be concerned by the project and to determine the future system or product users' profiles. This definition is based on brainstorming meetings with the stakeholders, and on state of the art reviews.

The Identification of the Experts of the Community

The identification of the communities of practices helps find the experts of those communities. These experts must be recognized as such by members of the community. They will talk "on behalf" of the community. These people are identified by social networks.

Organizing and Conducting the Discussion Groups

Each community of practices is composed of 4 to 6 experts. The sessions last between 3 to 3,5 hours and each session is videotaped. Each discussion session takes place in three phases: (1) the experts are invited to express themselves freely on the project {what they think about it? what do they do in relation to the project? How do they see the future?}; (2) the experts are invited to talk about the project while using different supports related to the project such as a computer screen, mockups, and storyboards; (3) the experts are finally invited to organize the ideas and information elicited during the discussions using the card sorting technique.

The Analysis of the Results

After each discussion session, the video recordings are analyzed in order to identify the ideas expressed by the experts that may translate into characteristics of the future system, service or product. All the comments made by the experts are transcribed and ideas extracted from the transcriptions.

Consensus Meetings

During the consensus meetings, the stakeholders and the design team are invited to discuss the results of the expert community staff discussions. The aim of the meeting is to arrive at a consensus on the specifications that are drawn from the experts' comments and ideas. The design team and the stakeholders must take over the results of the method and plan for the re-design of the information system.

CASE STUDY

Context of the Study

The European NAvigator (ENA, www.ena.lu) provides research and educational material on the history of European integration. The website was developed by the Virtual Resource Centre for Knowledge about Europe (CVCE). ENA is a multilingual, multisource and multimedia knowledge base that contains more than 15,000 documents on the historical and institutional development of a united Europe from 1945 to the present day. In this knowledge base, pupils and students, teachers, researchers, and anyone interested in the European integration process can find original material such as photos, audio and video clips, press articles and cartoons, together with explanatory synopses, tables and interactive maps and diagrams. The material included in ENA's vast and varied documentary resources is selected, created, processed and validated by a multidisciplinary team of specialists in European integration.

The initial demand of the stakeholder was to make the digital library evolve so as to be more in line with the users in order to provides them with richer services and more personalized ones through the possibilities of the Web 2.0.

Procedure

The identification of the communities of practices was essentially based on brainstorming techniques. The stakeholders were invited to participate in this task. A total of 14 groups were defined and for each group, experts were identified. These 14 groups were organized in two categories: one group of experts in the European integration (Group 1 to 8), and one group of experts in digital contents (Group 9 to 14).

Group 1: Historian-researcher, expert in contemporary history, European integration specialists. This group was composed of 4 senior university researchers, and laboratory directors in France, Luxembourg and United Kingdom.

Group 2: Historian-teachers at the college and secondary school. These people are experts in the teaching of the European history. This group was composed of 4 teachers in history, geography and civil instruction.

Group 3: Jurists specialized in European law. This group was composed of 2 lawyers from Luxembourg (European law), 1 university teacher, and 1 official representative from Luxembourg.

Group 4: Politicians. This group was composed of 2 members of the French Parliament, and 2 people from the Lorraine Region in charge of the international relations with Germany.

Group 5: Journalists. This group was composed of 3 journalists (from press, radio and web) and 2 persons in charge of communication. These people were experts in the dissemination of the knowledge on Europe.

Group 6: Amateurs and activists of history. This group was composed of 5 representatives of local or regional history association. Among these experts, 2 had written books on history.

Group 7: Schoolteachers. This group was composed of 5 schoolteachers from Luxembourg, France and Belgium. Four teachers were teaching French and history. They were experts in teaching Europe history.

Group 8: Doctoral student. This group was composed of 5 PhD students in history, computer science and in communication.

Group 9: This group was composed of 3 graduate students in history and law. These students were participating in an international exchange program between the university of Luxembourg and the university of Metz.

Group 10: This group was composed of 5 experts in cross-cultural studies (3 researchers, 1 company director expert in diversity management, 1 scientific advisor in a European institution).

Group 11: This group was composed of 5 librarian and archivist. They had expertise in cross-cultural differences between users.

Group 12: This group was composed of 3 German computer scientists. One of them was the manager of a German enterprise and another one was the Director of the Franco-German institute.

Group 13: Web usability specialists. This group was composed of 4 ergonomists, 3 of which were specialized in software and Web usability.

Group 14: Sociologist/communication/didactics. This group was composed of 3 researchers working on cultural practices in three disciplines: sociology, communication and didactics.

A total of 58 people (Belgian, Canadian, French, German, from Luxembourg, Spanish) participated in these 14 expert community staffs. They were invited (a) to produce knowledge about their needs (content experts) or about their representations of others' needs, (b) to express the needs, expectations, and requirements of the target users, (c) to comment on the existing services, and (d) to make explicit the useful and necessary knowledge for the DL. Each of the 14 groups was invited to participate in a working session of 3 to 3,5 hours that was video recorded. As indicated previously, facilitators who used the same procedure animated each group. After each working session, the video recordings were analyzed and a report was written.

RESULTS

The significance or importance of the expert community staff method may be judged on the basis of the quality of the data produced, in other words, on its capacity to produce new functionalities, to express new needs, and to generate new ideas for the DL. Thus, one of the issues was whether the expert community staff was a method allowing the production of novel, and innovative ideas able to guide the re-design of the DL.

Our work was first guided by the fact that the identification of users' needs results from the complex interactions between users and designers, in a context where imitation, learning and co-construction of knowledge as well as the sharing of the representations play an essential role. In addition, the needs expressed and shaped by the users and the designers are cross-validated. The users' needs emerge within and thanks to social interactions that are mediated by language. It is this basic premise that justifies the idea that users' needs elaborate from collaborative work where end-users and facilitators mutually enrich their knowledge through confrontation. The knowledge, which shapes users' needs, is either (a) clearly identified by the end-users who will indicate that they need such and such functionality, or (b) identified but not yet mature enough, or (c) completely ignored. In this latter case, data analysis will try to reformulate the expressions and translate them into user needs. As such, needs analysis corresponds to a form of cooperative production where many people will negotiate and validate shared representations of what they do, what they declare they do and what they would like to do with DL.

The participants produced 134 ideas about the DL. The ideas were classified into 52 topics having between 1 to 4 ideas for the improvement of the DL. Some of them were easy to devise (e.g., "better understand European politics based on European current affairs"). On the other hand, other ideas expressed were specific to some communities (e.g., "to archive the knowledge on the twinning of European towns").

The traditional functions expected from DL (e.g., data storage, information updating, credibility of the sources) were completely reconsidered by the staff of experts in 7 main functions. Although some traditional functions were stated again, some new functions were identified. In the first category, participants indicated that the DL should: (1) **Archive** the knowledge on the European integration in a reasonable, and reliable and to allow the users an easy access to this information while informing them of their rights to use this information; (2) to have the DL **Accredited** by credible institution and experts so as to give the it authority on the knowledge it manages; (3) to **update** the information (e.g., culture, values, etc.) on European integration. In the second category, 4 new functions emerged. Participants expressed ideas that could be grouped around the following functions. For the expert community staffs the DL should: (1) **claim** its culture and its specificity with other DL, while trying to differentiate between its identity and that of other DL; (2) **associate** different authorities (individual as well as legal, public as well as private) to the elaboration of the content of the DL; (3) **analyze** the content of the DL instead of simply store and present it. This would facilitate the understanding of the information by presenting contextual historical, geographical, cultural, artistic, social, psychological and political facts; (4) **lead** the people involved in the production of knowledge on the European culture so as to drive them to share it.

As it appeared, the expert community staffs did not limit themselves to express functional properties associated to online encyclopedia. For them, the DL should have other functions such as the animation of the community by having people

interact on topics of interest, and share data. These functions will represent new issues for the DL.

CONCLUSION

Although the design sciences give more and more place to participatory and creative approaches and methods, a prime importance should be given to the expert community staffs when dealing with digital contents. As this study as demonstrated, the expert community staff method provides new ways to bring relevant and important input from end-users and experts to design. This method starts from the idea that a user need is not given per se. It is a rather long process of collaboration between peoples in groups representing communities of practice having things to say, share, and debate on a specific topic. A need does not exist in itself outside of individuals' histories, outside of the society in which individuals live. Rather users' needs emerge and evolve as a consequence of social interactions. In this respect, expert community staffs allowed the construction of social situations in which each community of practice could express needs on the European integration.

The first aim of this article was to show what the expert community staff might bring to the re-design of digital libraries. The second aim of this paper was to propose a way to conduct the expert community staff based on the constructivist approach. This constitutes a new methodology that will need further development and validation.

REFERENCES

- [1] Allen, T. J. 1988. Distinguishing engineers from scientists. In: R. Katz (Ed.), *Managing professionals in innovative organizations: A collection of readings* New York: Harper Business. Pp. 3–18.
- [2] Anderson, C. J., Glassman, M., McAfee, R. B. and Pinelli, T. 2001. An investigation of factors affecting how engineers and scientists seek information. *Journal of Engineering and Technology Management*, 18, 131–155.
- [3] Brangier, É. and Bastien, J. M. C. 2006. L'analyse de l'activité est-elle suffisante et/ou pertinente pour innover dans le domaine des nouvelles technologies? In: Valléry, G. and Amalberti, R. (Eds.). *L'analyse du travail en perspective – Influences et évolutions*. Toulouse: Octarès. Pp. 143–156.
- [4] Comité technique AIPCR C3.1 Sécurité routière. 2008. *Guide Facteurs Humains pour des Infrastructures Routières plus sûres*. La Défense, France: Association mondiale de la route (AIPCR).
- [5] Del Fiol, G. and Haug, P. J. 2009. Classification models for the prediction of clinicians' information needs. *Journal of Biomedical Informatics*, 42, 82–89.
- [6] Fidel, R. 1986. The possible effects of abstracting guidelines on retrieval performance of free-text searching. *Information Processing & Management*, 22, 309–316.
- [7] Fidel, R. and Green, M. 2004. The many faces of accessibility: engineers' perception of information sources. *Information Processing & Management*, 40, 563–581.
- [8] Ghosh, M. 2009. Digital infrastructure and attitudes towards access and sharing: A case study of selected engineering libraries in the Maharashtra state of India. *The International Information & Library Review*, 41, 89–120.
- [9] Jones, M. L. W., Gay, G. K. and Rieger, R. H. 1999. Project Soup: Comparing evaluations of digital collection efforts [Electronic Version]. *D-Lib Magazine*, 5. Retrieved June 2009 from <http://www.dlib.org/dlib/november99/11jones.html>.
- [10] Kani-Zabihi, E., Ghinea, G. and Chen, S. Y. 2009. User perceptions of online public library catalogues. *International Journal of Information Management*, 28, 492–502.
- [11] Kitzinger, J. 2004. Le sable dans l'huitre: analyser des discussions de focus group (the sand in the oyster: analysing discussions from focus groups). *Bulletin de Psychologie*, 57, 299–307.
- [12] Moreira, B. L., Gonçalves, M. A., Laender, A. H. F. and Fox, E. A. 2009. Automatic evaluation of digital libraries with SSQual. *Journal of Informetrics*, 3, 102–123.
- [13] Nakatsuji, M., Yoshida, M. and Ishida, T. 2009. Detecting innovative topics based on user-interest ontology. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7, 107–120.
- [14] Park, P. 2009. Analysis of characteristics and trends of Web queries submitted to NAVER, a major Korean search engine. *Library & Information Science Research* (31), 126–133.
- [15] Sliman, L., Biennier, F. and Badr, Y. 2009. A security policy framework for context-aware and user preferences in e-services. *Journal of Systems Architecture*, 55(4), 275–288.
- [16] Warwick, C., Terras, M., Huntington, P. and Pappa, N. 2006. If you build it will they come? The Lairah study: quantifying the use of online resources in the Arts and Humanities through statistical analysis of user log data. In: Sun, C., Menasri, S. and Ventura, J. (Eds.). *Proceedings of Digital Humanities 2006*. Paris, 5–9 July. Paris-Sorbonne: ADHO. Pp. 225–228.
- [17] Wibeck, V., Adelsward, V. and Linell, P. 2004. Comprendre la complexité: les focus groups comme espace de pensée et d'argumentation à propos des aliments génétiquement modifiés (understanding complexity: focus groups as a thinking and arguing space about genetically modified food). *Bulletin de Psychologie*, 57, 253–261.
- [18] Wilson, L. A. 2003. If we build it, will they come? Library users in a digital world. In: Lee, H. (Ed.). *Improved access to information: portals, content selection, and digital information*. New York: Haworth Information Press. Pp. 19–28.

Session 13: Supporting Usage of Information Systems

User Interface Sketching to Improve Low and High Literacy User Information Seeking on Government and Social Service Websites

Neesha Kodagoda

N.Kodagoda@mdx.ac.uk

B. L. William Wong

Interaction Design Centre

School of Engineering and

Computer Sciences

Middlesex University, London, UK

W.Wong@mdx.ac.uk

Nawaz Khan

N.Khan@mdx.ac.uk

ABSTRACT

One of the challenges in designing online information systems for Government and Social Service systems is to understand their target ordinance and the level of literacy. The users' literacy is a fundamental factor in interface development, allowing designers to take advantage target users' information seeking behaviour. The paper will discuss how three ideas transformed the existing "Adviceguide" website. The ideas were based on previous research into information seeking behaviour characteristics. We used Jakob Nielsen's ten usability heuristics and Sweller's Cognitive Load Theory to focus on interaction of information and cognitive structures to guide instructional design. We further used different display techniques to solve and address focus plus context problems. The ideas were sketched following the above theories. Sketching, a rapid drawing technique to capture, represent and formulate outcomes visually was used to conceptualise the ideas. We will discuss the differences in the three user interface design sketches. Finally we will discuss how the new user interface designs might assist the low literacy users in their quest for information seeking in Government and Social Service websites.

Keywords

low and high literacy, sketching, behaviour characteristics, cognitive load theory, focus plus context

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)] User Interfaces – *user-centered design*.

INTRODUCTION

According to the National Skills for Life Survey carried out in 2003, 16% of the UK population – 5.2 million people – presented low levels of literacy [1]. As information technologies expand far beyond the traditional personal computers into public information systems, designers of these systems must consider their target audience's level of literacy. Poor English or few years of schooling could often result in limited ability to read,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

write or spell adequately for the demands of the daily life [2].

As Government and Social Services information move to online distribution, digital inclusion will require designers to consider the problems faced by low literacy users [3]. Many on-line social-service systems such as the "Directgov", "Adviceguide" and Barnet Citizens Advice Bureau websites require a high-school level of reading [2] and present information that often spans many different government "silos" [4]. Socially disadvantaged people are more likely to require those services and are more likely to have low levels of literacy [5].

In this paper we describe three interface sketches of the Adviceguide website. The modifications are based on information seeking behavioural models. We believe these changes might assist low literacy users in finding information.

PROBLEM UNDER STUDY

A previous study indicated that low literacy users performed significantly worse than high literacy users when they search for Social Service information online. [2].

Another study looked at low literacy users' information seeking behaviour strategies using a pharmaceutical company website. They found six main strategies: Low literacy users read word by word trying to make sense of information and do not present the ability to scan. They have a narrow field of view and are not likely to notice content above, below or to the side of their focus. They were likely to be satisfied and abandon the search early assuming they found relevant information. When confronted by long, dense pages of text, some low-literacy users simply skip chunks of text. They tried to minimize the amount of reading they would have to do by focusing on finding links instead of reading content and were easily distracted and often derailed by content and links that were pulled out into the right margin. They usually avoid searching because it requires spelling and typing, but if forced, they will click on the link that appeared first or that looked simpler [6].

Kodagoda et al. [3] found similar navigational behaviours except for the tendency to skip chunks of text when faced with dense pages. The results also indicated that low literacy users did not verify the information found for correctness. They were unable to recover from a mistake even if they identified wrong or irrelevant content. They did not share similar clues that lead to very different trajectories during their search paths. Low literacy users' mental representations of the categories were not matched to the actual categories of the system. Finally, low literacy users abandoned a search task if they were (a) unable to find the information, (b) unable to recover from a mistake, (c) unable to match their mental representation of the categories

with the systems representation, (d) satisfied with information they found regardless of its correctness.

Low literacy users demonstrated a critically different strategy to high literacy users when searching for information using the “Adviceguide” website. They spent a lot of time reading instead of scanning, usually terminating the search before finding the right information. Verification was inexistent and a recurrent attitude to give up and terminate the search was presented. Their ability to recover from encountering wrong information was very low and they demonstrated a very narrow focus in all the cases. These behavioural patterns provoked low literacy users to use different search paths or trajectories.

To better support the low and high literacy users with information seeking, we looked at information seeking behaviour models as a theoretical lenses to analyse their behaviour from the identified characteristics [7].

Our aim is to design interfaces that will improve online information seeking by reducing abandonment rates, increasing focus, improving recovery and verification, and lowering the working memory load for low literacy users.

In the next section we will look at theories that could be applied to resolve the identified problems of low literacy users information seeking.

THEORIES USED TO ADDRESS THE PROBLEM

In this section we will look at the theories that were used to address the identified eight information seeking behaviour characteristics.

In all our sketches we used the ten general principles on good interface design identified by Nielsen [8].

Focus + Context

Over the past fifteen years focus plus context problems have been addressed using different display techniques to solve and address information structures [9]. This concept was used to

address the visualization of large amounts of information on a screen, while helping users to maintain focus, avoid disorientation and prevent users becoming lost, all while maintaining a simple navigation system.

In some designs distortion is used to magnify certain parts of the screen while some parts are de-emphasised.

Card, Mackinlay et al. [10] argue that focus + context is based on three premises: “First, the user needs both overview (context) and detail information (focus) simultaneously. Second, information needed in the overview may be different from that needed in detail. Third, these two types of information can be combined within a single (dynamic) display, much as in human vision”.

Cognitive Load Theory

Cognitive Load Theory is proposed by Sweller and his co-workers [11–13]. This theory is built upon the research of Miller [14] which determined that the capacity of working memory is limited to seven (plus or minus two) chunks of information at a given time. This is widely accepted by instructional designers, and acts as a design guideline to enhance development of learning material. We further plan to use three techniques Split-Attention, Redundancy Effect and Modality Effect [12] to address the identified problems by low literacy users during their information seeking.

Sketching User Interfaces

Freehand sketching assists in early design process allowing a natural way of thinking about ideas and communication. The rapid drawing technique helps capture, represent and formulate ideas to visual representations.

PROBLEM WITH THE CURRENT DESIGN

For the purpose of this study we used the “Adviceguide” website to address the identified problems as discussed below. A snapshot of the ‘Adviceguide’ website is shown in Figure 1.

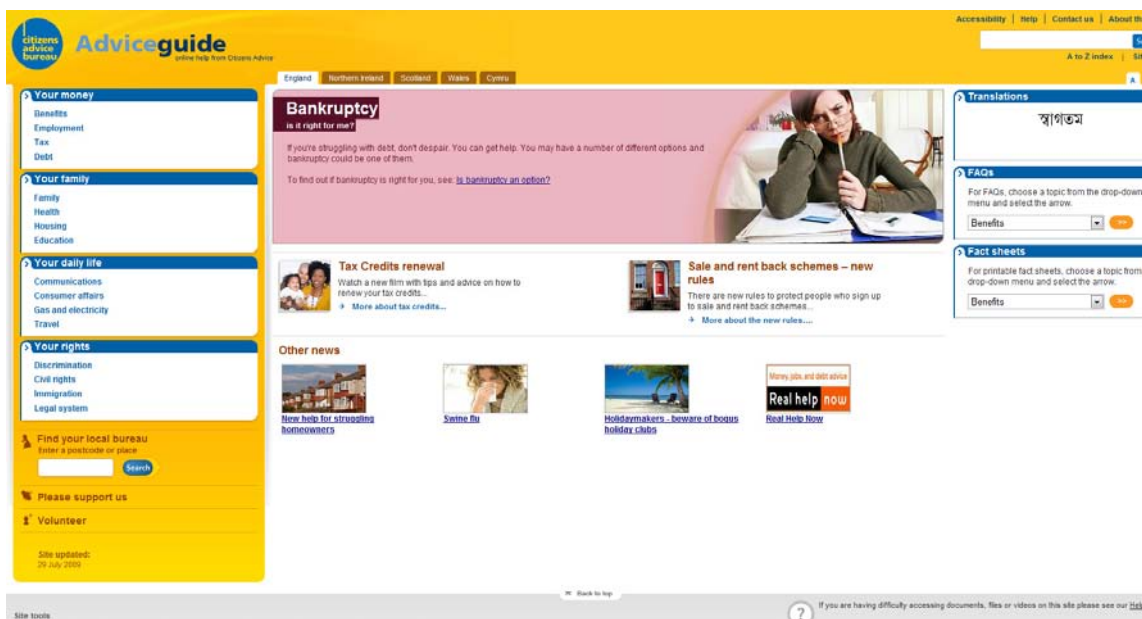


Figure 1. A snapshot of the ‘Adviceguide’ website home pageThe identified eight information seeking behaviour characteristics.

Table 1. Information seeking behaviour characteristics.

Behaviour Characteristic	Definition	Users behaviour	Measures taken to improve the information seeking
1. Reading/ Scanning	For the purpose of this study, Reading behaviour took place when users read word by word, while Scanning behaviour referred to, users glancing through headings and subheadings or start, middle of a paragraph until they found something relevant or interesting.	Low literacy users read word by word trying to make sense of the information they read, while high literacy users glance through headings and paragraphs until they find relevant or interesting information.	Reduce the text content and simplify the reading at entry level. Main and sub level classification to be broken down in to more manageable chunks (+ or – 7 chunks). Dense pages were reduced and anchor links were removed. Information in detail pages were shown in bullet format. Audio was made available for the text.
2. Focus	Focus when users are not likely to notice content above, below or to the side of their focus, results a narrow field of focus.	Low literacy users did not notice content above; below or to the side of their focus creating narrow field of view.	Dense pages were reduced and anchor links were removed. Duplicated or replicated information within the current site was removed. Scrolling down pages was minimized. Only main classifications were shown on the main page and all sub level classifications were removed. Main classifications were colour-coded. Sub level classifications were kept to a + or – 7 chunks. Use of path breadcrumbs. Suggestions of (semantic and other users' visits).
3. Satisfied	Satisfied as soon as the user assumes they have sufficient information and abandon the search task at an inappropriate place, due to being satisfied quickly.	Low literacy users are likely to abandon a search assuming they have found relevant information.	Suggestions of (semantic and other users' visits). The suggestions will encourage the user to navigate and explore other related links without being satisfied quickly.
4. Verification	Verification when users find information they need and examine other related links to support the information found for correctness.	Low literacy users do not examine other related links to verify the information found for correctness.	All duplicated or replicated information from the current site were removed. Information available in audio format. Suggestions of (semantic and other users' visits).
5. Recovery	Recovery refers to recuperate from a wrong or irrelevant information search to a more focused or relevant one resulting in finding the required information.	Low literacy users were unable to recuperate from a wrong or irrelevant information search to a more focused or relevant search.	The colour coded classifications (main and sub levels). Path breadcrumbs To maintain focus on the selected distortion and smooth exploration transitions are used. Top menu which contains the icons home, forward and backward.

6. Trajectories	The trajectories are information search paths taken by users.	Low literacy user information seeking paths were dissimilar to other low literate users.	The colour coded classifications (main and sub levels). Main classifications with pictures. Information available in audio format.
7. Abandon	Abandon: when users show a higher tendency to give up their search due to many reasons.	Low literacy users had a high tendency to abandon the search assuming they are unable to find information.	Suggestions of (semantic and other users' visits).
8. Representation	Representation users' mental representation of information categories becomes a mismatch to system.	Low literacy users mental representation of website categories differs from the system.	Main classifications with pictures.

THE THREE PROPOSED INTERFACE DESIGN SUGGESTIONS USING SKETCHING

We found three interface designs that could reduce the difficulties faced by low literate users. We will describe their commonalities first, and their details later. The designs are based on the theories abovementioned, but further evaluation is required.

GENERAL PRINCIPLES AND GUIDELINES AND THEORIES USED FOR THE THREE INTERFACE DESIGN

- All the three interface design sketches followed Nielsen's ten heuristics.
- The "Adviceguide" main and sub menu classifications presented as text links on the left hand side were removed and only the four main classifications were used in the home page. They were: your money, your family, your daily life and your rights. The main menu text links were replaced by four large colour coded shapes.
- All sub level classifications followed the main classifications colour code to show relationships between.
- Path breadcrumbs were used to help users keep focus and reduce the level of abandoning a search task. These will maintain the same colour of the main classification
- Top menu will contain the following navigation icons: home, forward and backward.
- Reduce the text content and simplify the reading at an entry level.
- Main and sub level classification to be broken down in to more manageable chunks (+ or - 7 chunks).
- Dense pages were reduced and anchor links were removed.
- Information in detail pages were shown in bullet-point format.
- To avoid split-attention, no pictures were available in the specific detail sections.
- All designs show clear outer boundaries.

User Interface Sketch A

User interface design A is quite similar to the original "Adviceguide" website.

- Sketch A keeps the main menu on the left-hand side, and in the middle of the screen but with images that depict the main topics (Figure 2).
- Once a main classification selection is made, the appropriate sub level classification is shown in the middle of the screen as shown in Figure 3. The main classification menu on the left which are not related will be distorted to maintain the focus on the selected sub level classifications.
- When a sub level classification is selected the distortion will keep the focus on the selected sub level and the immediate sub levels (Figure 4).
- If the user chooses the audio modality in the "detail" page (the page which contains the most information), only the audio file will be played and no text will be shown (Figure 5).

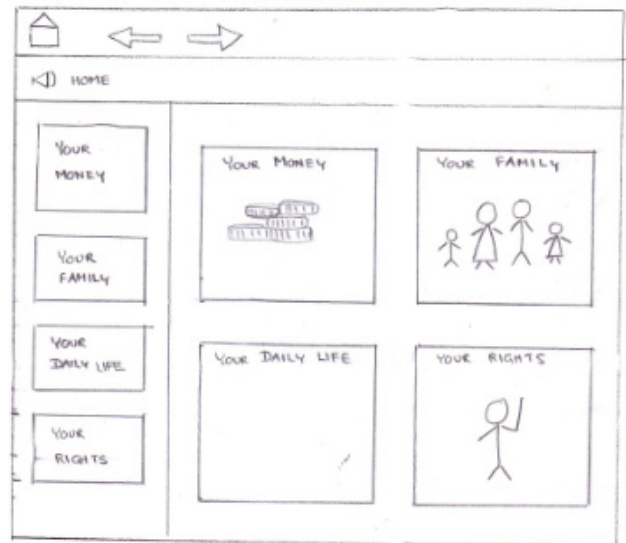


Figure 2. Sketch A – home page.

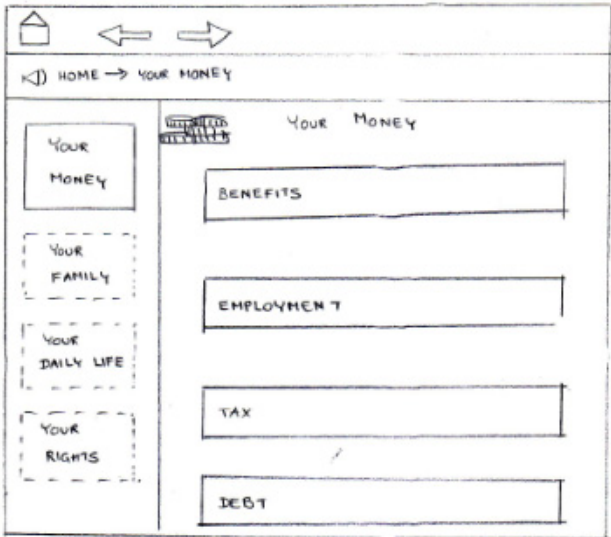


Figure 3. Sketch A – first level.

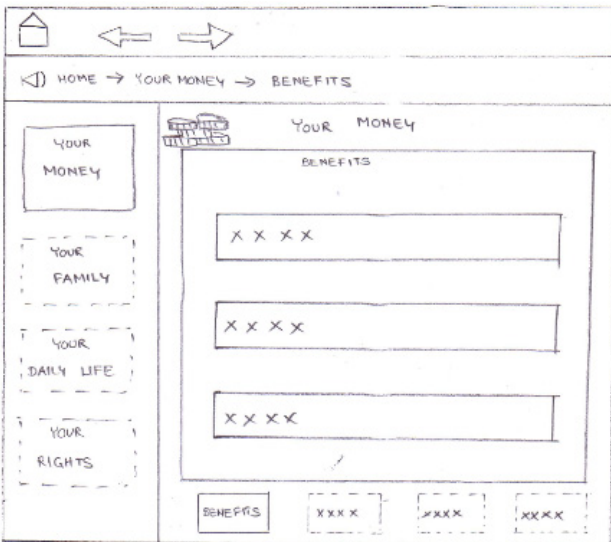


Figure 4. Sketch A – second level.

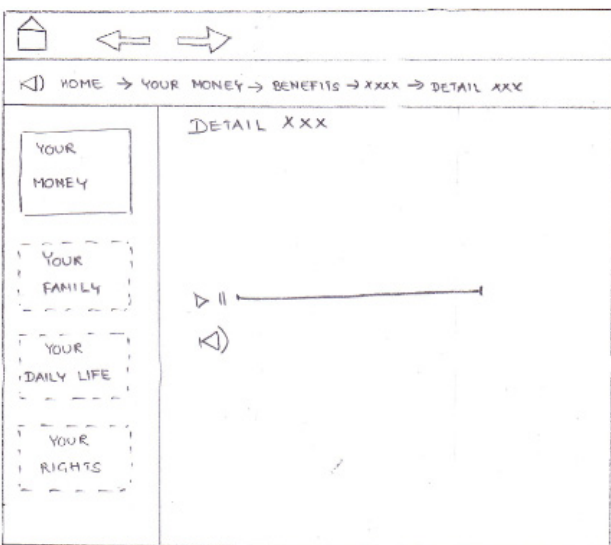


Figure 5. Sketch A – detail page level

USER INTERFACE SKETCH B

The main menu classifications will be presented in the middle of the screen accompanied by related images for easy identification of the topic (Figure 6).

When a main classification is selected, the focus is enforced on the selected main or sub level classification. The remaining classifications move towards the most right and bottom corners of the screen as shown in Figure 7 and 8. The movement is smooth.

In the “detail” page, the system will show text but no audio facility will be available. The system will present suggestions by semantic closeness to other detail pages and by frequency of visited pages (Figure 9).

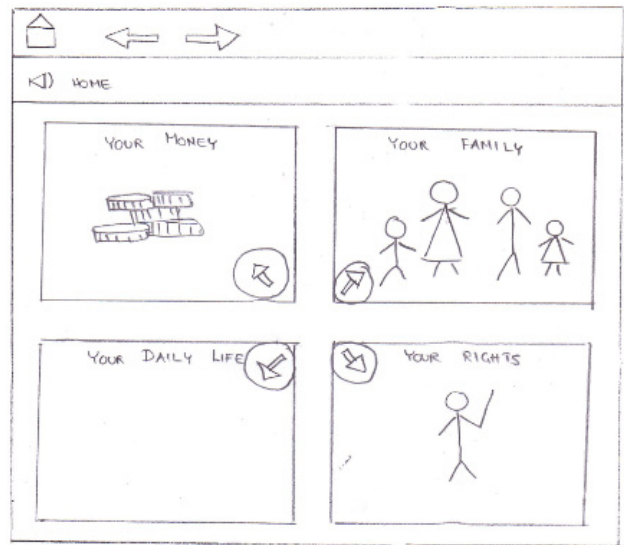


Figure 6. Sketch B – home page.

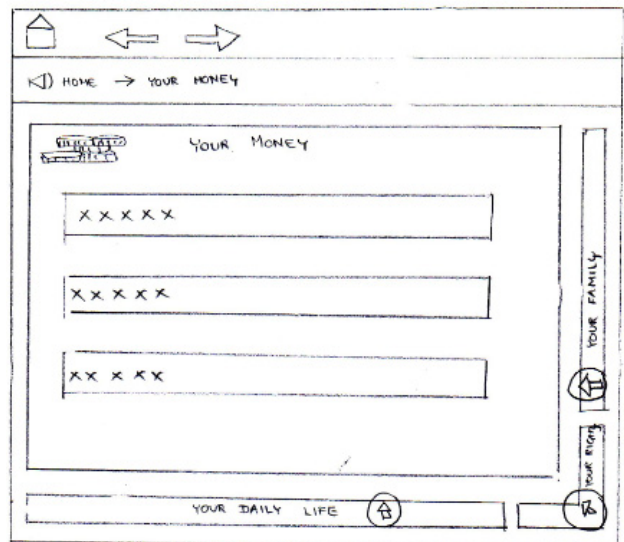


Figure 7. Sketch B – first level.

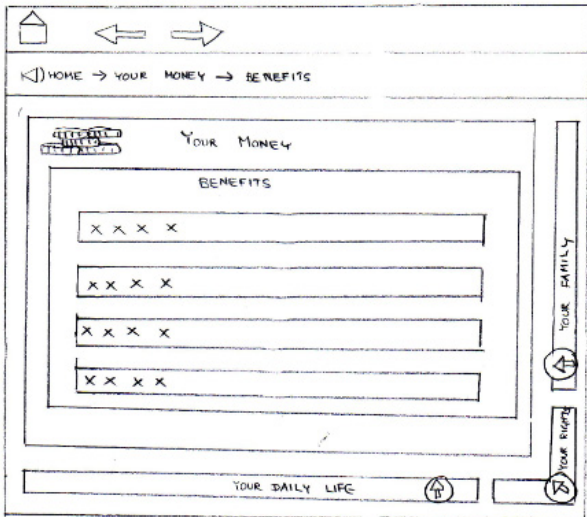


Figure 8. Sketch B – second level.

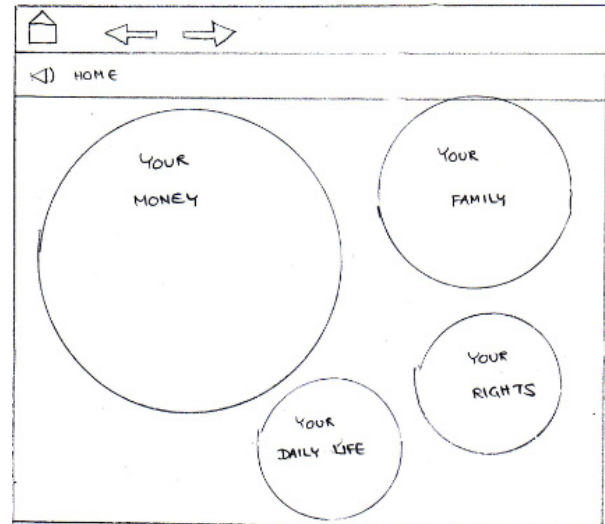


Figure 10. Sketch C – home page.

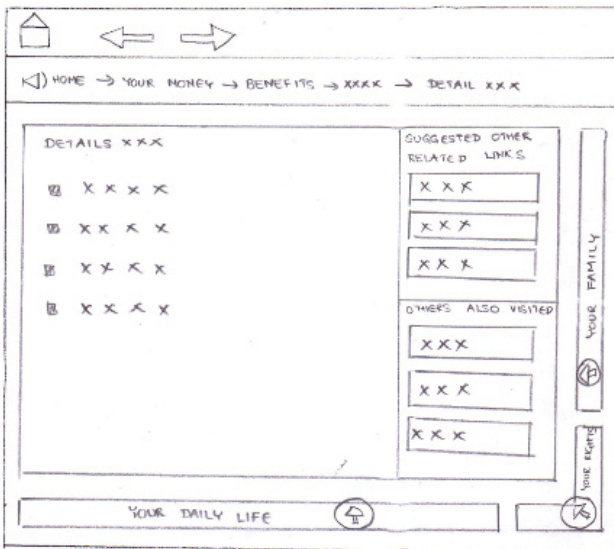


Figure 9. Sketch B – detail page level.

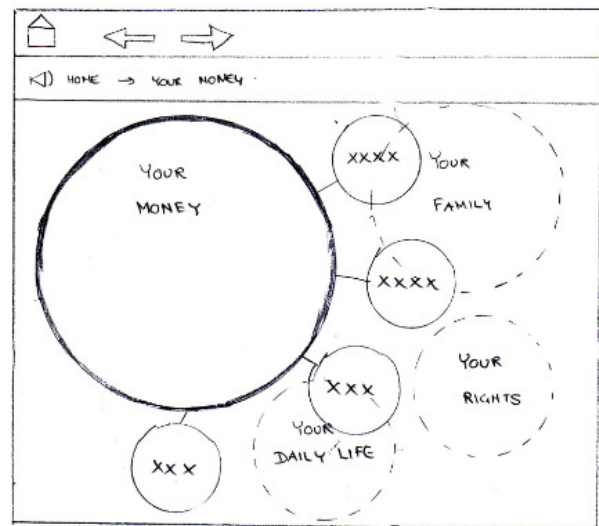


Figure 11. Sketch C – first level.

USER INTERFACE SKETCH C

- The main classifications are shown in the middle of the screen.
- The size of the classifications shape will depend on the number of times it has been accessed, therefore, the more popular a link is; the bigger it will become (Figure 10, 11 and 12).
- There are no images presented with the main menus.
- When a user selects a main classification the related sub level classifications is shown in an outer circle to the selected classification (Figure 11).
- The selected classification outer line will be thicker than normal to gain focus.
- The results of the distortion involved, allows the user to focus on the selected region while keeping a sight of the remaining main classifications. This will follow to the sub level classifications (Figure 11 and 12).
- In the “detail” page, text and audio facilities will be available. The system will give suggestions similar to Sketch B, however, the user has control on the number of suggestions presented by using a sliding bar (Figure 13).

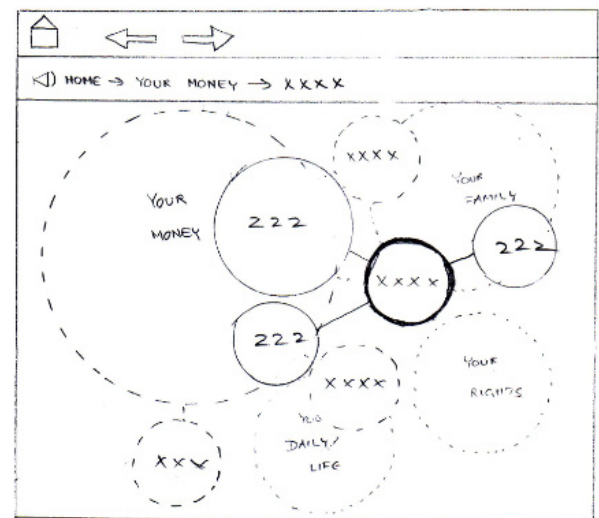


Figure 12. Sketch C – second level.

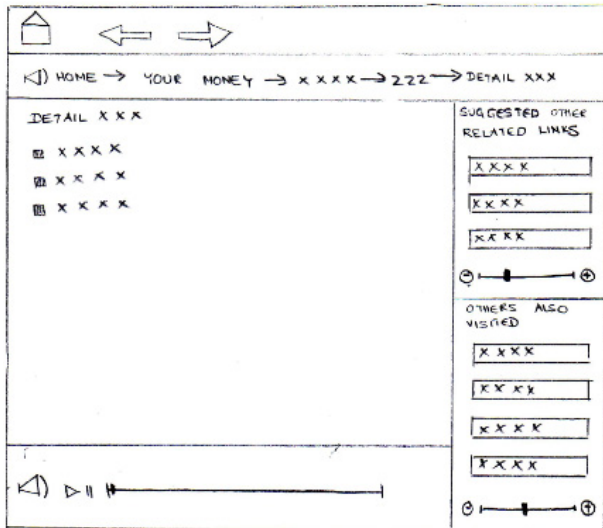


Figure 13. Sketch C – detail page level.

FROM SKETCHES TO THE DESIGN

The above sketches were used to develop three prototypes for user evaluation and testing. The “Adviceguide” data were reorganised. The modified classifications were fed into a MySQL database. All three designs used the same database to display information.

The details information on classification were simplified for ease of readability and where ever possible were shown in point form. The detail information is stored in XML files and the file names were fed to the main database.

The current “Adviceguide” websites duplicated or replicated information were removed.

Clear path breadcrumbs were provided to make users aware of their current position and to facilitate navigation using the path breadcrumbs or the main home, back and forward buttons.

All main and sub level classification to detail information has separate audio files which will be stored in the database. The user can enable or disable this function.

In the ‘detail’ page, the suggested links will be presented depending on the semantic distances to other links: the lower the semantic distance, the higher the relevance. The measurements will be added on to the existing database. Suggestions will also include most frequent visits.

The UI designs are Adobe Flex based. Java is used to access the MySQL database which runs on Apache Tomcat.

These prototypes are currently under development.

DISCUSSION AND CONCLUSION

Previous research identified several problems that low literate people face when searching for information on the Internet. We found eight information seeking behaviour characteristics that needed to be addressed.

As Government and Social Services information move to online distribution, digital inclusion will require designers to consider the problems faced by low literacy users. We proposed three designs to alleviate the lack of focus, the difficulty in reading and the high abandon rates, that low literates present.

These designs used robust theories and concepts like Nielsen’s heuristics, Focus + Context and Cognitive Load Theory. The three main ideas were conceptualize using the sketching technique which is a rapid drawing technique, to capture, represent and formulate the outcomes visually.

The main problem identified for low literacy user was the long amount of time spent on reading without understanding. Reducing the text readability level to an entry reading level and using an audio modality, we hope to reduce the working memory load. Reducing dense pages to bullet points and removal of anchor links will improve the focus + context problems.

To improve the low literacy users lack of focus and reduce the abandon rates, we remove duplicated and replicated versions of information. Scrolling down lengthy pages was minimized. Path breadcrumbs and semantic suggestions were used to improve the focus + context problem. Sub level classifications shown on the home page were removed. to reduce the cognitive load and increase the working memory capacity, we colour coded the main menu classifications and kept that coding thorough the sub levels. To reduce using being satisfied to quickly suggested links will encourage the user to navigate and explore further additional links.

To encourage low literacy users to verify the information, we include suggested links.

We believe the colour coded classification will improve the focus + context problem. While introducing pictures to the main classifications and introducing audio to the text will reduce the cognitive load and assist in finding the information. But we believe this problem is a direct result of the representation problem which needs to be researched and addressed.

Representation becomes a very challenging issue due to low literacy users’ mental representation of the system being different. We believe this could be solved by the suggested semantic distances.

We need to evaluate these prototypes in order to find out which will be the best to support low literate seeking behaviour on the internet. We hope that the results of that evaluation will provide us with guidelines for policymakers and designers of Government and other Social Services websites.

REFERENCES

- [1] Williams, J. et al. The Skills for Life survey: A national needs and impact survey of literacy, numeracy and ICT skills. 2003. London: BMRB Social Research for the Department for education and Skills, 2003.
- [2] Kodagoda, N., Wong, B. L. W and Khan, N. Effects of Low & High Literacy on User Performance in Information Search and Retrieval. In: The Proceedings of the 22nd Annual Conference of Interaction a Specialist Group of the BCS 2008. HCI Liverpool.
- [3] Kodagoda, N., Wong, B. L. W. and Khan, N. Identifying Information Seeking Behaviours of Low and High Literacy Users: Combined Cognitive Task Analysis. 2009. NDM9 2009.
- [4] Wong, W., Keith, S. and Springett, M. Fit for Purpose Evaluation: The Case of a Public Information Kiosk for the Socially Disadvantaged in HCI 2005. 2006: Springer, 2006.

- [5] Shalla, V. and Schellenberg, G. The Value of Words: Literacy and Economic Security in Canada. Catalogue No. 89-552-MPE. 1998, Ottawa: Statistics Canada.
- [6] Summers, K. and Summers, M. Reading and Navigational Strategies of Web Users with Lower Literacy Skills. In: 68th Annual Meeting of the American Society for Information Science and Technology (ASIST). 2005.
- [7] Makri, S., Blandford, A. and Cox, A. L. Using Information Behaviours to Evaluate the Functionality and Usability of Electronic Resources: From Ellis's Model to Resource Evaluation. 2008. Journal of the American Society for Information Science and Technology, 44(2), 613–634.
- [8] Nielsen, J. and Molich, R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people. 1990, ACM: Seattle, Washington, United States.
- [9] Leung, Y. K. and Apperley M. D. A review and taxonomy of distortion-oriented presentation techniques. In: CHI. 1994: ACM.
- [10] Card, S., Mackinlay, J. and Shneiderman, B. Readings in Information Visualization: Using Vision to Think. 1999: Morgan-Kaufmann.
- [11] Sweller, J. et al. Cognitive Load as a Factor in the Structuring of Technical Material. Journal of Experimental Psychology: General, 1990, 119, 176–192.
- [12] Chandler, P. and Sweller, J. Cognitive Load Theory and the Format of Instruction. Cognition and Instruction, 1991, 8, 293–332.
- [13] Sweller, J. Cognitive Load Theory, Learning Difficulty and Instructional Design. Learning and Instruction, 1994, 4, 295–312.
- [14] Miller, G. A. The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review, 1956. 63, 81–97.

A Comparative Field Study of Four Crane Control Interfaces

Joanna Bergström-Lehtovirta¹, Jussi Rämänen², Tiia Suomalainen¹,
Antti Oulasvirta¹, Tuomo Kujala³

¹Helsinki Institute for Information Technology HIIT, Finland

²Helsinki University of Technology TKK, Finland

³Agora Human Technology Center, University of Jyväskylä, Finland

ABSTRACT

This paper reports on a comparative field study looking at the relationship between crane operators' perception of directions in the working environment and the design of the control interface. Moving a load with a crane is a common task on factory floors, and human error in this may lead to accidents. In this paper, we show that the demands of this complex spatial task can be decreased with a simple design solution that helps the operator to understand the mapping between controls and movement directions of the crane more rapidly. We report on a field experiment that was conducted in a real use environment with experienced crane operators ($N = 6$). We conclude with a discussion of how safety on factory floors can be improved with simple design solutions that decrease the attentional demands of the operating task.

Keywords

crane operator, user interface, field study

ACM Classification Keywords

H.1.2 [User/Machine Systems] Human factors.

INTRODUCTION

Our analysis of occupational accident reports (from the Advisory Committee on Occupational Safety and Health, Finland) indicates that inattention was involved as a factor in over 50 percent of investigated accidents in a random sample of 100 occupational accident reports in Finland made during the year 2005. Any tool design solution or work rearrangement that can decrease the level of attention required for operating on factory floors could have a positive effect on occupational safety.

The floor-operated overhead bridge crane with two rails and a girder ("bridge") is commonly used for indoor lifting purposes (Figure 1). The United States Department of Labor has defined the "overhead crane" as a "crane with a movable bridge carrying a movable or fixed hoisting mechanism and traveling

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

on an overhead fixed runway structure." "Bridge travel" is defined as the "crane movement in a direction parallel to the crane runway" and the "trolley" is "the unit which travels on the bridge rails and carries the hoisting mechanism." Finally, "hoist" refers to an apparatus exerting a force for lifting or lowering and "hoist motion" means "that motion of a crane, which raises and lowers a load." [1]

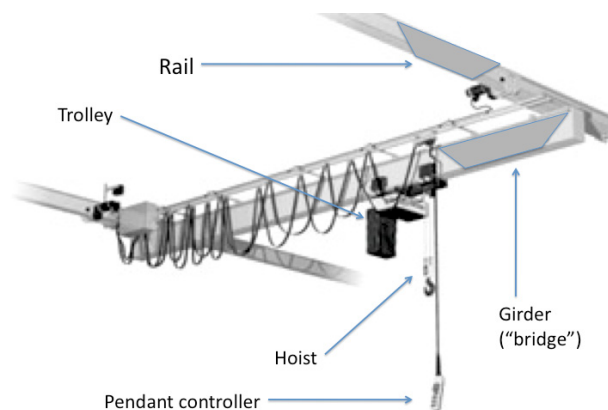


Figure 1. Floor-operated overhead bridge crane.

The operator of the overhead crane stands on the floor with the crane controller. Two rails form a rectangle in a horizontal two-dimensional plane, and a hoist adds the third dimension. Hence, the space in which the user is moving the hoist is a cube-like volume, limited by the rails and the floor. In three-dimensional, indoor space, human perceive the vertical directions (up and down) easily, but the problem is how we could help the user to know which wall of the horizontal plane maps to which buttons on the controller.

Knowing which buttons to press to move a crane from position A to position B requires taking into account the coordinates of the operator, the current position of the load, its target position, and the girder to which the hoist is attached. In the standard controller design (see Figure 2, CN), four symbolic cues are placed in two rows and up/down on a third row; the first four symbols can be seen in the girder. To move the hoist to the correct direction, the operator needs to "map" the direction symbols on the buttons to the desired movement direction of the girder. This requires shifting the gaze back and forth between the crane and the controller to infer the mapping.

Fitts and Deininger [2] reported first the phenomenon of spatial stimulus-response (S-R) compatibility. According to Umiltá and Nicoletti [3], stimulus-response improves with a mapping in which the left stimulus is mapped to a left key press and right to a right key press. Also, if there is a rule-based relationship between the stimulus and response, the mapping is easier – for example, if it can be mapped as a mirror-opposite location.

With overhead cranes, the position of the operator can vary, so the direction from which he is looking at the symbols under the girder varies also. The direction button that now moves to the left will move to the right when the operator is positioned to the other side of the girder. Thus when the buttons are aligned in rows, the “left”/“right” buttons are on the same side as the stimulus only from the other side of the girder.

The direction in which the various machine controls move should be compatible with the resultant movements produced [4]. However, in most cranes of this type, no control–movement compatibility has been maintained, since there are a number of manufacturers, using different designs or standards for their products. The control layout can also vary within the same work area. Often, operators have to switch crane during work. Thus, the operator might move the crane in a non-desired direction if the controller’s layout was different in the crane used previously, resulting in safety risks. If the layout cannot be made consistent within any given factory, one option could be consistent color marking for the controls, as Sen and Das suggest [4]. Also a tactile discrimination of the controls, using shape and size, was suggested. However, consistent alignment supports task performance better than color marks do, when the operator has learned the layout, thus decreasing the need to look and map the marks of the controls.

When stimuli vary along the vertical dimension (up/down) and responses along the horizontal (left/right) dimension, an up–right/down–left mapping typically yields better performance than the reverse order [5]. In the standard controller, the up/down buttons are aligned like this. However, front/back alignment would be more consistent.

In this paper, we describe results of a study done in collaboration with our industrial partner. Our partner’s internal observation studies have indicated that the crane operators do not really use the direction symbols enough. Sometimes the symbols have become ablated from the buttons, or sometimes it is just too inconvenient to map the symbols to the symbols on the girder. Thus, instead of using the symbols, the users first check the direction in which the press of a button will move the hoist. This unnecessary behavior is not only a waste of time; it raises concerns of the possible consequences of moving the hoist without knowing where it is going to move.

The aim of this study was to find out whether better alignment of buttons decreases uncertainty at the beginning of the moving task of the crane. We also examined which factors have an influence on the performance time when the task is to choose a button to press to move the hoist to the assigned target.

The hypotheses for the study were as follows:

- H1. Compatible alignment between the buttons on the controller and the symbols on the girder decreases the performance time.
- H2. Choosing the buttons for the vertical directions is faster than choosing the horizontal directions.
- H3. When one operates from a position where the buttons on the controller and the symbolic cues on the girder are aligned consistently, performance is faster than from other positions.

METHOD

We carried out a comparative field evaluation with four different crane controller interfaces. The tests were conducted in a real usage environment that, because of its special characteristics, is technically difficult to replicate in a laboratory setup.

Participants

The experiment was conducted with six middle-aged experienced male crane operators. We recruited two people who were working in the factory hall where the tests were held and four who were not. The participants received a Maglite® flashlight as a gift for their participation.

MATERIALS

Controller Interfaces

For the tests, four different crane controller interfaces for a push-button pendant controller were prepared (see Figure 2). These interfaces were attached to the real controllers.

We used two different alignments of controls:

- Controls nonaligned (CN)
- Controls aligned (CA)

and two different designs of symbols:

- Symbols-standard (SS)
- Symbols-cues (SC).

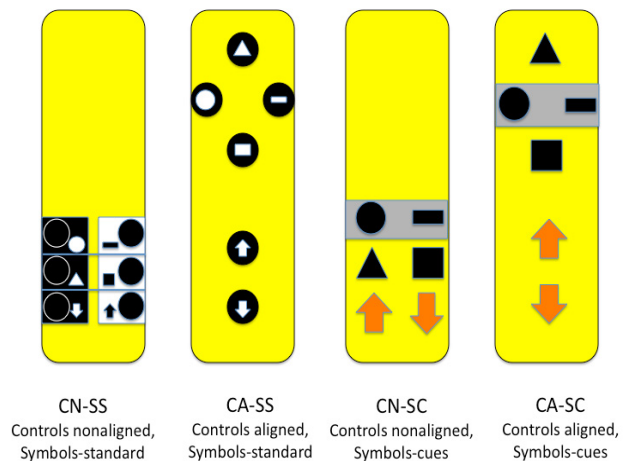


Figure 2. Crane controller interface layouts used in the tests.

CN is a widely used typical controller button layout. CA is a prototype controller button layout that has a consistent alignment of controls to the directions in the environment. SS employs standardized symbols used in Finland. SC has the buttons shaped as symbols, and a color mark is added to discriminate up/down movements from vertical movements and also under the buttons moving the bridge.

Test Setup

The design for the test setup is presented in Figure 3. The setup included three spots from which the tasks were executed (polygons with numbers) and four piles of colored cardboard boxes. Two piles were at the level of the hoist (1.5 meters high), one pile near floor level (0.5 meters high), and one pile at a higher level than the hoist (2.0 meters high).

Experiment consisted of two types of tasks, each of which consists of two subtasks that are directions. Both directions are either in horizontal plane (1.5 m piles), or other is in horizontal plane and other in vertical (either 2.0 m or 0.5 meter pile).

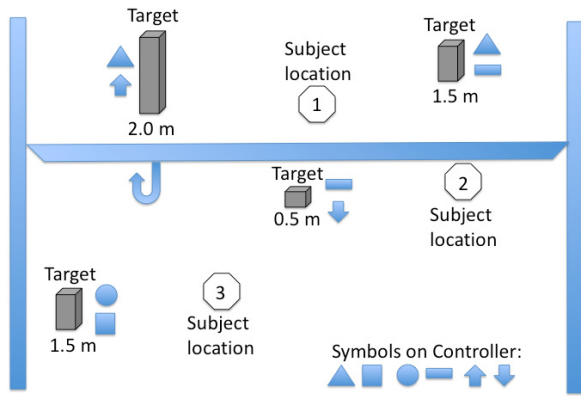


Figure 3. Test setup design. Next to the piles are presented the symbols required for accomplishing the task.

Design

The experimental design was a 4 x 2 x 3 = 24 within-subjects design with controller–symbols combination as the first factor, task target as the second factor, and the coordinates of the operator as the third factor. To mitigate order effects, the order of controller interfaces and tasks were rotated over participants.

During the test, the moderator guided the participant between the three locations on the factory floor and asked him to press as quickly as possible the buttons required for moving the hoist above the cardboard box pile indicated by the moderator. To move the hoist in the correct direction, the user has to compare the symbols on the controller to the ones printed under the girder. However, users who have been working with a specific crane for a long while memorize the button functions and thus do not necessarily have to make the comparison.

The tasks were prepared in such a way that always two and only two buttons were required for the task execution, to examine how the different directions needed to accomplish a task were influencing the performance.

Procedure

Each participant carried out six tasks with each of the four crane controller interfaces. In each of the 24 tasks, the test moderator guided the participant to one of the three spots marked on the factory floor (Figure 3) and asked the subject to choose and show as quickly as possible the two buttons required to move the hoist from its current location to one of the four target locations indicated by the moderator.

The execution times for the tasks were measured with a stopwatch and documented. After the task part of the test, a brief semi-structured interview was conducted. The interview served as a way to collect qualitative data to complement and support the quantitative data collected from the test tasks.

RESULTS

According to the data, task performance was fastest (3.60 seconds on average) with the CA controller with symbol set SS (controls aligned, symbols-standard) (see Figure 4). To test for differences among conditions, we rely on calculated 95% confidence intervals.

From task performance times, CA–SS was considered the only controller to differ from two others (from CN–SS at a borderline level and from CA–SC) significantly. When only CA and CN

(controls nonaligned) were compared, regardless of symbolic cues, CA had the best average task completion time (4.08 s); however, the difference was not significant. CA provided the best results specifically in locations where the buttons and signs on the girder were aligned consistently.

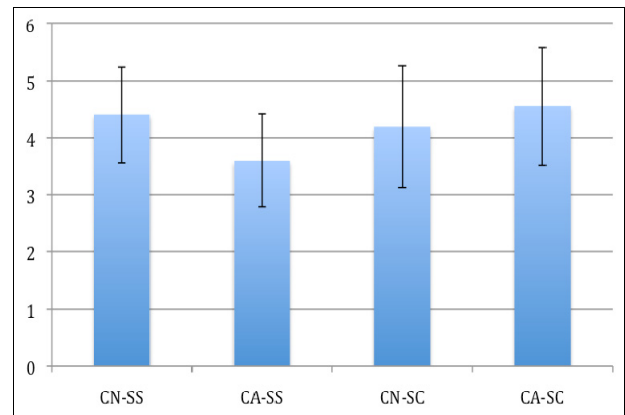


Figure 4. Average performance times for the controller interfaces. Vertical bars denote 95% confidence intervals.

In comparison with symbol set SC (symbols-cues) (4.37 s), task performance was faster (4.00 s) with SS (symbols-standard); however, the difference was not significant. Performance times with controller design CA were better with SS, but CN had better timings with SC.

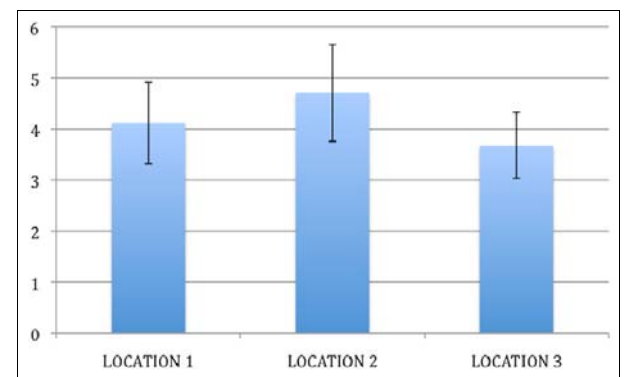


Figure 5. Performance times in different operators' locations. See Figure 3 for locations.

Compared to in location 2, task completion was significantly (3.68 s ± 0.64 s, N = 48) faster for location 3 (see Figures 3 and 5), where buttons on the controller and the symbolic cues on the girder were aligned consistently. Performance was second fastest for location 1 when the location was exactly opposite location 3. Task completion was slowest in location 2 with a 90° difference between the symbolic cues on the girder and on the controller.

On average, tasks that consisted of only one horizontal-level subtask combined with a vertical-level subtask were significantly faster (3.44 s ± 0.59 s, N = 48) than tasks with two vertical directions (4.53 s ± 0.62 s, N = 96).

The data also show that the two participants who had worked in the hall before were, on average, significantly faster (2.02 s ± 0.3 s, N = 48) than the remaining four participants were (5.27 s ± 0.58 s, N = 96). These subjects performed significantly

faster with CA-SS and CA-SC than with CN-SS and CN-SC (see Figure 6).

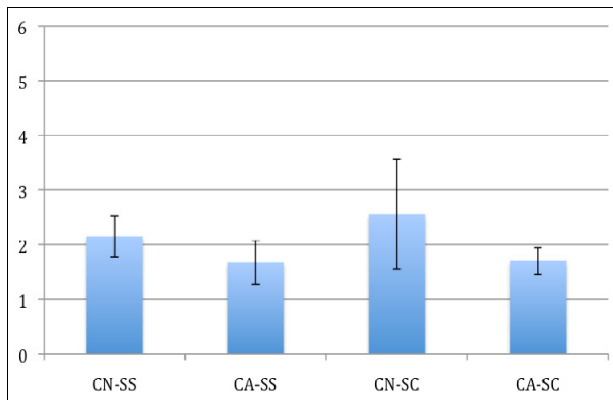


Figure 6. Experienced subjects' (N = 2) average performance times for the controller interfaces. Vertical bars denote 95% confidence intervals.

DISCUSSION AND CONCLUSIONS

The results support our hypothesis H1, that compatible alignment (in controls aligned design) between the buttons on the controller and the symbols on the girder improves performance when compared to the standard version, in which the buttons are in rows (controls nonaligned).

Supporting hypothesis H3, the difference was statistically significant and emphasized especially when the operator was positioned on the side of the crane where the symbols on the girder could be seen in the same layout as on the controller. A rule-based relation makes the mapping easier. Our results indicate that from both consistent and opposite directions it is easier to map the directions than from non-orthogonal position.

The results indicated also that hypothesis H2 is correct, which suggested that choosing the buttons for the vertical directions is faster than choosing the horizontal directions. Tasks that consisted of only one horizontal-level subtask combined with a vertical-level subtask were significantly faster. This indicates that if directions can be perceived naturally, not by mapping, performance is faster.

In the experiments, we observed that participants who had worked in that particular hall before (two persons) had learned the horizontal directions so well that they did not have to look at the symbols on the girder at all. These participants just searched for the right symbol on the controller. The factor most influencing their performance was probably the discrimination of up/down buttons in the controls-aligned design, which may have decreased the time the searching of the symbols took.

The results showed that the added symbolic cues did not have the desired influence on performance time. This was probably because the cues were new to the participants, and the interview revealed that the design of the figures was not coherent enough. Only one of the participants perceived the meaning of the girder figure.

Increased size of symbolic cues (SC) was associated with a faster performance time when the symbols on the CN controller (controls nonaligned) were larger and the standardized symbolic cues (SS) were relatively small. The smallness derives from ablating of the symbols; when symbols are placed on the buttons, they are more likely to get dirty or ablate than are those in a design where they are placed next to the buttons.

In our experiment, we compared widely used pendant push-button controllers. When a pendant controller is used, the movement area of the operator is limited by the wire where the controller is hanging, and which is attached to the bridge. Though it is more common to operate a crane with a pendant controller while positioned on only one side of the bridge, in this study the subjects performed tasks from both sides of it. Thus, the coordinates of the subjects used in the test setup are applicable with both radio and pendant controllers. Our experiment design renders the results applicable also for push-button radio controllers.

We conclude that, of the four crane control interfaces, the prototype controller button layout performed statistically better than the commonly used controller. Consistent alignment of symbols aids in directional inference and decreases the need to shift one's gaze between the controller and the girder, thus releasing more resources for observing events on the factory floor. Consistent alignment also decreases inconvenience caused by mapping the symbols on the controller to the symbols on the girder. Safety on factory floors could be improved with such design solutions, which decrease the attentional demands of the operating task.

ACKNOWLEDGMENTS

This research was sponsored by Theseus, a research project on human-technology interaction design, funded by Tekes (the Finnish Funding Agency for Technology and Innovation).

REFERENCES

- [1] United States Department of Labor, Occupational Safety and Health Administration, Occupational Safety and Health Standards, Materials Handling and Storage, Standard Number: 1910.179, Overhead and gantry cranes.
- [2] Fitts, P. M. and Deininger, R. L. 1954. S-R compatibility: correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology*, 48, 483–492.
- [3] Umiltá, C. and Nicoletti, R. 1990. Spatial S-R compatibility. In: Proctor R. W. and Reeve, T. G., *Stimulus-response compatibility: an integrated perspective*. Pp. 89–116.
- [4] Sen, R. N. and Das, S. 1999. An ergonomics study on compatibility of controls of overhead cranes in a heavy engineering factory in West Bengal. *Applied Ergonomics*, 31, (2000), 179–184.
- [5] Adam, J. J., Boon, B., Paas, F. G. W. C. and Umiltá, C. 1998. The up-right/down-left advantage for vertically oriented stimuli and horizontally oriented responses: a dual-strategy hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1582–1595.

Reasons for Under-use in Professional Software

Anna Kämäräinen
Agora Center
PL 35
40014 University of Jyväskylä, Finland
Anna.k.kamarainen@jyu.fi

Pertti Saariluoma
Department of Computer Science
PL 35
40014 University of Jyväskylä, Finland
Pertti.saariluoma@jyu.fi

ABSTRACT

Aim of this study was to find aspects affecting use difficulties and analyze user psychological problems in the use of professional software.

Keywords

user psychology

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *user-centered design*.

INTRODUCTION

In the last decades, digital revolution has changed the society. It has had an effect how we work, how we organize interpersonal communication and in general lead to the establishment of an advanced service-based Information Society. Computers have been crucial in this transformation, though in the early phase they affect mainly in working practices. In the early days, access to computers was restricted only to skilled expert users. In order to use computers, users must command the machine with low-level cryptic language. The development of computers leads to smaller machines and they broaden to work places and homes.

The digital revolution has increased the amount of different kinds of electronic documents. In nowadays the world of the desktop is the world of managing documents as emails, spreadsheets, images, textual data and business documents just to name few. In the working environment, many electronic document management systems have been developed for management, sharing and retrieval. In the point of of users', document management systems should act as a service point to storing and retrieving information easily.

In the business field, information management has the key role in success. That is the reason, why organizations invest billions to IT infrastructure. In the document process management, majority of executives consider improvement of operational efficiency and customer satisfaction as two key advances of effective

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

document management (Neal 2008). However it seems that document management and moreover information retrieval are not seen as important as other parts of service or production chain. Reliable and useful information management contains also information retrieval. In most systems documents are organized for convenience of the systems rather their users (Dourish, et. al. 2000). Same document can have different user groups and different information needs of users. The purpose of document management is to rationalize use of documents. Though, document use cannot be effective if retrieval is not effective.

The information retrieval from large document management systems relies heavily on the search engines. Two alternative perspectives of effectiveness exist. Effectiveness can be measured in the point of view of system or in the point of view of users. The system-based approach covers amongs others such aspects system performance, system speed and precision.

Human system interaction is important research area when the goal is to ease usage of systems, smooth adoption cycle and intensify work performance. Poor interaction design typically cause poor quality, inefficent performance, slow or negative adoption cycle, emotional aspects and pure safety risks. All these aspects are affecting under-use and can cause economical losses. Interaction design is challenging area, because it deals with two totally different elements, the human and the system. Designers usually do not know and understand well enough the principles which users are driven by. Folk psychological intuitions and technical knowledge is not usually enough. Instead, design decisions should be based on psychological knowledge.

Although people's ability to use computers has become better, many people agonize over use difficulties. A lot of literature and studies on interface design exists, but user difficulties remain. This means that traditional usability studies, usability engineering and user-centered methods do not cover human-technology interaction whole. What we need is psychologically analyzable explanations on human behavior. User psychology use psychologically analyzable explanations expounding interaction situations (Moran, 1981, Oulasvirta & Saariluoma 2004, Saariluoma, 2004; 2005; Saariluoma et. at. 2008). In this paper, we are interested to find reasons for user difficulties affecting under-use.

By under-use, we refer to the situations where people need some tools or services, but for some reason they do not use them (Kämäräinen & Saariluoma 2007; Kämäräinen, Saariluoma 2008). In constructing new forms of interaction, a challenge is how to get people to effectively use available possibilities, services, computational function, just to name few. Under-use is serious challenge for service industry, but also in work places. In the case of service industry, under-use phenomena may expose the service industry to a danger of negative adoption cycle, as has happened in the case of WAP. In the case of work places, the purpose of applications is to help, rationalize and make working more effective, but if employees suffer user difficulties, use of systems cannot be effective.

Employees may have possibilities to contribute and develop their work processes, but thanks to use difficulties, they are not able to gain their performance level. This is why it is essential to investigate the use difficulties in work environment. From user psychology point of view the crucial question to ask is what the possible sources of under-use are. Another important question is, how people learn to use computational capabilities of applications and what aspects cause learning obstacles.

METHOD

The focus on this was finding major reasons for user difficulties. Our study concentrates on user psychological issues related to enterprise information system. We did not attempt to find all difficulties users encountered. Our focus is on typical problems and user psychological explanations behind those problems.

Research design

We conducted a questionnaire with three parts: a) demographic questions, b) experiences of using computer and Internet, c) questions of usability and user problems related to system. The first two parts were to sort out do participants have basic knowledge to use computers and system under study. Experience questions contain background information including weekly computer use, Internet browser and Intranet use experiences and real use of studied system.

The most important part of this study is naturally the questionnaire of usability and user problems. This part consisted of quantitative and qualitative questions dealing with clarity, system speed and information retrieval. The questionnaire was available in organizations Intranet for employees who use the system. In total amount of respondents were 205.

Participants

In total, data consisted of 205 respondents, 29 women, 162 men and four did not answer. Gender distribution has a direct analogy to the real gender distribution in the business field. The biggest age group amongst respondents was 41–50 years (34.6%). More detailed demographics are shown in the Table 1. Occupations include sales (38%), engineering (16%), marketing (6%), management (6%), RTD (5%) and other occupations (19%). This means that questionnaire reached to variety of user groups with varying use contexts and needs.

5% of the respondents used computers minimally at work and spare time and 95% used daily at least at work, although 81% used computer at work and spare time. 18,5 % evaluated themselves as inexperienced users of Internet browsers and 81.5% consider themselves at least somewhat experiences. 83.4% of the respondents used organizations Intranet quite often or more and 84.4% used document management system under study somewhat often or more frequently. Based on data, it seems that respondents are experienced computer and Internet users. Daily computer use and experience of using Internet suggests that respondents have basic computing skills. Internet use is highly dependent of users' ability to seek information and use search engines. Consequently it can be assumed that somewhat experienced browser users have sufficient skills to use search engines and information searching methods.

DATA ANALYSIS

In working environment all devices and systems are seen as tools for achieving human goals (Carroll & Cambell, 1986;

Saariluoma, 2005). This means that interaction activities can be seen from the goal-oriented perspective. In any interaction, people have some goal and therefore it is vital to understand what the obstacles to reach these goals are and what psychological phenomena cause the problems.

Starting point in analyzing user psychological problems was to identify all tasks and problems users described and why they consider these tasks problematic. We collected all problems which employees had mentioned in the open answer questions. By using user psychological framework, we classified problems into four categories.

RESULTS

Aim of the document management or knowledge management systems is to act as a service point for storing and retrieving information easily. This should be the main objective in development of knowledge sharing systems. Knowledge and document sharing has a key role in the business environment and therefore we are interested all aspects of the system that real users consider problematic. Unavailable important information and time spent for searching have negative effects to operational efficiency.

Our goal in this research was to find aspects affecting use difficulties. We categorize answers to groups which manifest similar user difficulties. In the analysis, we found 4 types of difficulties, which are common in this system.

Found categories are 1) continuous failures 2) system speed 3) lack of support 4) contents. Further, these categories manifest two more general main aspects of user difficulties: a) system do not meet users' requirements and b) users' skills do not meet system available.

Continues failures covers all situations where user did not manage to get information needed or finding information relies on heavy effort. Typical examples of failures were: information was not found, although it is in the system, search results consisted of huge amount irrelevant items, users need to remember exact product name or code. The common feature of all these difficulties is that user did not succeed to reach his or her goal. Moreover, the origin of these situations is that user is trying to do some action, but it fails.

System speed is typical user difficulty which has been studied a lot. Also in this data, users felt that system speed is too slow, updates take too long and these hamper working. It is well known fact that slow system speed causes frustration and irritation. It also has direct effect to effective work practices while users have to wait accomplishment of search tasks.

Lack of support consists or all those aspects which origin is on inadequate user support. This category includes lack of manuals, poor help pages and need for training sessions and a contact person assisting when problems arise.

Contents – category covers all aspects where users are not satisfied with the offered contents. This meant that people did not get information they needed in their attempts to solve their problems and continue the effective use.

Continuous failures and system speed are difficulties which origin is in the systems technical functionality. Most of the document and knowledge management systems rely on some sort of document properties or metadata. Documents' metadata typically consists of owner, date, title, last modified, just to name few. The purpose of using metadata is to speed up and enrich searching procedures. This data suggests that documents

in this system are organized for the convenience of the system and according to technical requirements. The data shows that it is hard, time consuming and at times impossible for users to find information and documentations needed.

Lack of support, manuals and contents are difficulties which origin is the gap between system and users' requirements and skills.

Table 2.

Category	Frequency
Continuous failures	50
Contents	18
Lack of support	13
System Speed	12

DISCUSSION

Using a system is a holistic process with many possible obstacles. As all know a program may fail thanks to one missing or mistaken line. Even a type can prevent a big program from functioning in a proper manner. Similarly, one problem in interaction process may also have fatal consequences for proper operation of a human machine system.

When people use a machine they are half Turing machines. They are given a set of instructions which they should use to reach their goals. These instructions define accurately how they can operate and on one level they are nothing but a part of program information flow. If they miss one single command, the operation of the system stops.

Often people can use a part of the given command set and thus they can participate to operation of a part of the machine. This is not a problem as long as they can reach the action goals they need to reach but it is often also the case that people cannot reach everything they could when they use an information system. This latter case we call under-use. However, immediate interaction is not the only aspect of under-use.

Under use is a problem of different nature compared to problem of proper use. The problems of underuse arise from the personal action goals and not from the use of a system directly. People use information systems always to reach their personal action goals. This means that these systems are instruments or tools in they need to reach their goals [7]. This is why we cannot look the problems of under-use only from the narrow perspective of how to use, but we have to consider also the issues of why to use.

If a person does not know what the commands, in text-based or graphical interface are, it is not possible for her to reach the life-goals, which are the ultimate motivation for using the system. If a person does not know what the important commands in cutting a paper roll are, it means that the operation cannot be done. In this paper, we have illustrated some major explanations for why, one cannot use a system and consequently, we know how designers could react to problems of under-use.

However, thinking under-use from the wider perspective we can also argue that users, who cannot reach the action goals they have set for themselves may lose their interest to the system. Even more difficult is an additional problem. People

do not necessarily know that the system could support operations they need to do as they do not know that the system could help them in that. In this latter case they do not even actively search for operations they could benefit, when using the system.

In terms of user psychology, we can thus suggest several content-based explanations for under-use [6]. By content-based explanation we refer to the mental contents of the users, i.e., to the information contents in users' mental representations. In content-based explaining, we investigate mental contents and base the explanation of the critical behavior on the analysis of the properties of mental contents.

Here, we can argue that first type or explanation for under-use is that people do not know the command sets, i.e., and how to reach this knowledge. Second issue is that they do not know that they could reach their personal goals by using the system, if they would train themselves to do it. This means that it is important both to teach people how to use a system as well as what are the action goals they could reach by using the system.

Our data illustrates both types of problems. If manuals do not clearly tell what the users could use the systems for and how they could use it when they understand their goal, it does not provide sufficient support form people, who should mentally represent how to use the system in their everyday work life.

The consequence of such failures in defining accurately, how people should use a system, and what they should use it for, is easily poor motivation and under use for that reason also. This is why it is essential that designers do not look systems only from the Turing machine perspective, but also from the perspective of what people are supposed to do with the system in their work life.

REFERENCES

- [1] Neal, 2008.
- [2] Dourish, P., Edwards, W. K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D. B. and Thornton, J. 2000. Extending Document Management Systems with User-Specific Active Properties. *ACM Transactions on Information Systems* 18(2), 140–170.
- [3] Moran, T. 1981. The applied psychology of users. *ACM Computing Surveys*, 13(1), 1–11.
- [4] Oulasvirta, A. and Saariluoma, P. 2004. Long-term working memory and interrupting messages in human-computer interaction. *Behavior & Information Technology*, 23(1), 53–64.
- [5] Saariluoma, P. 2005a. Explanatory frameworks for interaction design. In: Pirhonen, A., Isomäki, H., Roast C. and Saariluoma, P. (Eds.) *Future Interaction Design*. Pp. 67–84.
- [6] Saariluoma, P. 2005b. Mitä on käyttäjäpsykologia? [What is user psychology?] *Psykologia* [The Journal of the Finnish Psychological Society], 40, 181–186.
- [7] Saariluoma, P., Parkkola, H., Honkaranta, A., Leppänen, M. and Lamminen, J. 2009. User Psychology in Interaction Design – The Role of Design Ontologies. In: Saariluoma, P. and Isomäki H. (Eds.). *Future interaction design II*. Springer: London.
- [8] Kämäräinen, A. and Saariluoma, P. 2007. Under-Use of Mobile Services: How advertising Space is Used. In: Evers, V., Sturm, C., Rocha, M. A. M., Matinez, E. C. and Mandl, T. (Eds.). *Designing for Global Markets* 8,

- Proceedings of the Eight International Workshop on Internationalisation of Products and Systems. Pp. 19–29.
- [9] Kämäräinen, A. and Saariluoma, P. 2008. User Psychological problems in implementing production control system. In: Abascal, J., Fajardo, I. and Oakley, O. (Eds.). Proceedings of European conference on cognitive ergonomics. Pp. 174–177.
- [10] Carrol, J. M. and Campbell, R. L. 1986. Softening up hard science: Reply to Newell and Card. *Human-Computer Interaction*, 2(3), 227–249.
- [11] Carrol, J. M. 2003. *HCI Models, Theories and Frameworks: Toward a multidisciplinary science*, Morgan Kaufmann.
- [12] Fitts, P. M. and Posner, M. I. 1967. *Human Performance*, Brooks Cole.
- [13] Nielsen, J. 2000. Why you only need to test with 5 users. Available as: <http://www.useit.com/alertbox/20000319.html>. 13.4.2008.
- [14] Oulasvirta, A. and Saariluoma, P. 2006. Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies*, 64(10), 941–961.
- [15] Norman, D. 2004. *Emotional design: Why we love (or hate) everyday things*. Basic Books.
- [16] Shiffrin, R. 1988. Attention. In: Atkinson, R., Herrnstein, J., Lindzey, G. and Luce, R. (Eds.). *Stevens' handbook of experimental psychology, Learning and cognition* (2nd ed.). Oxford, England: John Wiley & Sons, 1988. Pp. 739–811. [Chapter]
- [17] Shneiderman, B. 1980. *Software psychology: Human factors in computer and information systems*. Winthrop Publishers.

Anatomy of an Adaptive Scaffold for Analysts

Stephanie Santosa¹, Judi McCuaig², Joseph MacInnes¹, Nathan Kronenfeld¹, William Wright¹
¹Oculus Info Inc.

2 Berkeley Street, Suite 600, Toronto, Ontario, Canada
 {ssantosa, jmacinnes, nkronenfeld, bwright}@oculusinfo.com

²University of Guelph
 Guelph, Ontario, Canada
 judi@uoguelph.ca

ABSTRACT

Adaptive user interfaces offer the potential to improve the learnability of software tools and analytic methodologies by tailoring the operation and experience to a user's needs. Scaffolding is an instructional strategy that can be applied by adaptive interfaces to achieve this. Scaffolding theory suggests that the level of guidance should be adjusted to optimize learning and performance levels. This paper explores the use of adaptive techniques to scaffold user interaction and presents a taxonomy of techniques for adaptive scaffolds within complex software systems. The techniques identified in the proposed taxonomy can help software scaffolds select appropriate adaptations in response to the user's learning and operating needs. A scaffold called nAble was implemented to explore the application of adaptive techniques from the taxonomy to support an analysis methodology called the Analysis of Competing Hypotheses (ACH).

Keywords

adaptive interfaces, scaffolding

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
 User Interfaces – training, help, and documentation, user-centered design.

INTRODUCTION

Introducing new software tools or methods to users in a workplace typically requires either a classroom based training program with simplified toy problems or computer-based training during which students are led through a set of lessons and exercises. In-house trainers may follow-up to reinforce lessons and to help the users perform actual tasks with the new tools and methods. The time required for someone to learn new software and associated techniques can be days to several weeks, which represents time away from normal work duties. Formal training can be a barrier to the use and exploitation of new capabilities. Adaptive scaffolding can assist users to learn both the methodology and the software tool "buttonology" and so reduce the need for formal training.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

Scaffolding is a training strategy used by human instructors to help learners achieve more than they could independently. A good scaffold helps the student to perform beyond current skill levels and fades away as the student gains expertise in the task [10]. The effectiveness of scaffolding lies in the ongoing diagnosis of the student's understanding and the tailoring of support given accordingly. An adaptive interface uses models of users, tasks, interface components and domains along with inference techniques to personalize interaction with the user [3], [7]. A software scaffold is a type of adaptive interface that can, like a human instructor, dynamically adapt interactions with the user [2], [6]. Users of any expertise level can interact successfully with a scaffolded application whether they are acquiring new skills or using well understood procedures.

This paper presents a taxonomy of techniques for adaptive scaffolding within complex software systems such as those used for visual analytics. Section: "Taxonomy of Adaptation Techniques for Scaffolding" begins with a description of each of the adaptive techniques in the taxonomy. Section: "nABLE IN nSPACE2" follows with a description of "nAble", an adaptive scaffold implementation within a tool for intelligence analysis. Section: "Conclusions" concludes with a description of how the taxonomy fits into potential future research.

TAXONOMY OF ADAPTATION TECHNIQUES FOR SCAFFOLDING

This taxonomy was derived from an iterative process involving the examination of previously implemented scaffolding systems along with data from human tutors in an initial scaffolding experiment [9]. The categories in the taxonomy are shown in Table 1 below.

Table 1. Categories of adaptive scaffolding techniques.

1. Recommend Information	6. Scale Automation
2. Support Bootstrapping	7. Adjust Error Recognition
3. Clarify Meaning	8. Change Feedback
4. Manage Attention	9. Scale Complexity
5. Support Sub-Task	

The techniques identified in the proposed taxonomy help software scaffolds select appropriate adaptations in response to the user's learning and operating needs. These techniques can be performed by adapting different aspects of the interface, including GUI elements such as menus and toolbars, interaction methods such as gestures and hotkeys, navigational paths such as links, instructional content, and methods of

content delivery such as popups and sounds. Models of user expertise, along with learning style, preferences, interests, tasks, interruptability, and other properties can drive the scaffold's adaptation decisions. Categories in this taxonomy are not exclusive. For example, a bootstrapping technique might also incorporate recommendations and attention management. The following discusses each category.

Recommend Information. Providing the user with tailored content at the appropriate times reduces information overload associated with learning. An adaptive system can make suggestions about where to find appropriate information or can modify the instructional content presented to the user. It could, for example, suggest collaborators with whom the learner could connect with, navigational paths for the learner to follow, or specific actions to perform. A scaffold could even provide step-by-step tutorials for raw novices and remove or tailor them as expertise grows.

Support Bootstrapping. Bootstrapping is brief, direct and specialized pre-task guidance which gets a user started with the task. Bootstrapping adaptations allow novice users to overcome the intimidation and frustration of starting with a blank slate. An example of this type of adaptation is an instructional note identifying possible actions for novice users on how to begin. Interface elements can also be highlighted to achieve this bootstrapping effect.

Clarify Meaning. Visual annotations can be used to elaborate meaning for elements of the methodology, instructional material or the interface. Exaggerated visual elements can implicitly introduce the meaning of a feature and then fade away to a more succinct form once the feature is understood. Fading annotations optimizes the screen space for both novice and expert users. For example, large icons with labels could be used in interfaces for novice users, which gradually transform into an expert interface with smaller, simplified icons.

Manage Attention. Visual signals, such as increasing size, flashing, animating, or highlighting, as well as auditory cues such as beeping, can direct user attention to the appropriate operating control or analysis step. The goal of attention management is to direct user focus. These techniques are disruptive by nature, so they are typically applied when it is certain that the user should be attending to something specific or is doing something incorrect. A scaffold can guide the user through a task by focusing the user's attention on relevant interface elements, thus prompting the user to perform, and learn, the required actions. Attention management is a natural solution for guiding novice users to useful functions.

Support Subtask. A task model can help an adaptive system determine the current and future steps of a user's process. Using a task model, adaptations can be made that specifically support the user's current objectives. An adaptive scaffold can limit the interaction area based on sub-task, provide guidance on a methodology being followed, and enable or disable specific links. For example, the nAble scaffold described in the next section uses a Hidden Markov task model to determine a user's current subtask and provides task-specific instructions on the methodology and tool.

Scale Automation. A novice user's tasks can often be automated, or partially completed with the help of the adaptive scaffold. The balance of control can shift from the system to the user as expertise is developed. A scaffold initially assumes more task responsibility, and then fades this support. This will allow novices to immediately be productive and to gradually perform more tasks independently with flexibility. Adapting the extent of machine automation can improve efficiency and

task performance. Automation may be disruptive to some users and beneficial to others, so tuning it to user properties can help achieve a correct balance.

Adjust Error Recognition. The actions or degree of inaccuracy for an error to be signaled in the system can vary by user. This allows for more appropriate error recognition since knowledge about the user can indicate what to register as an error. An adaptive scaffold can be more tolerant of errors from novice users and expect more precision and accuracy as users grow in proficiency with the system. This technique was used by the human tutors in the initial scaffolding experiments described in [9], who tolerated less thoroughness on evidence examination initially. As expertise developed, tutors increased expectations and consequently showed less tolerance before identifying an analysis error. Errors can be in methodology as in the experiment, or in 'buttonology' when related to tool operation.

Change Feedback. The type and method of feedback can be adapted to suit user characteristics such as learning style or expertise level. Novices might be provided with validation and encouragement for accomplishing basic tasks, while experts might only receive feedback about potential errors. The human tutors in the initial scaffolding experiments continuously validated the actions of complete novices, and this validation was reduced as the subjects gained competence. This type of adaptation is used in intelligent tutoring systems where the feedback given after each step is adapted to the learner [1],[4]. Another form of feedback can be at the interaction level. This can include providing marking menus to help novice users with learning gestures and removing the feedback as their expertise increases [12].

Scale Complexity. An interface's visual and control complexity can be adapted to match the user's skill level. For example, the interface could hide or gray out elements that are not currently useful, and then introduce more functionality when appropriate. Layering the complexity of an application can help reduce cognitive load. This is especially important for complex tools or during early learning when the user is overloaded. Alternatively, the workflow could be scaled to provide a simpler, streamlined methodology for novice users to follow, and progress to more demanding ones as competency is gained.

This taxonomy for adaptive interfaces can be used to guide the selection of techniques for implementing adaptive scaffolds. A system designer may systematically go through each category and decide which adaptations to implement based on what is possible given the original non-adaptive user interface for the system, what would be most useful for the task in question, and the models available to provide input to the reasoning system.

nABLE IN nSPACE2

nSpace2 is a rich, web-based integrated cognitive workspace used in information analysis and is comprised of TRIST [8] for information triage and the Sandbox [11] for evidence marshalling and visual sense making. The Sandbox is a flexible, visual thinking environment that supports both formal and informal analytic methods. The goal of nAble is to provide functionality and training, when needed and just as it is needed, using an adaptive scaffold. The nAble scaffolding system consists of sensors of user activity, Hidden Markov task models, Bayes reasoning engine for expertise and learning style, adaptive scaffolding techniques and dashboards for delivery of operational or methodology instruction. Results from the initial experiment using the nAble scaffold suggest a significant improvement over traditional help and are described in [9].

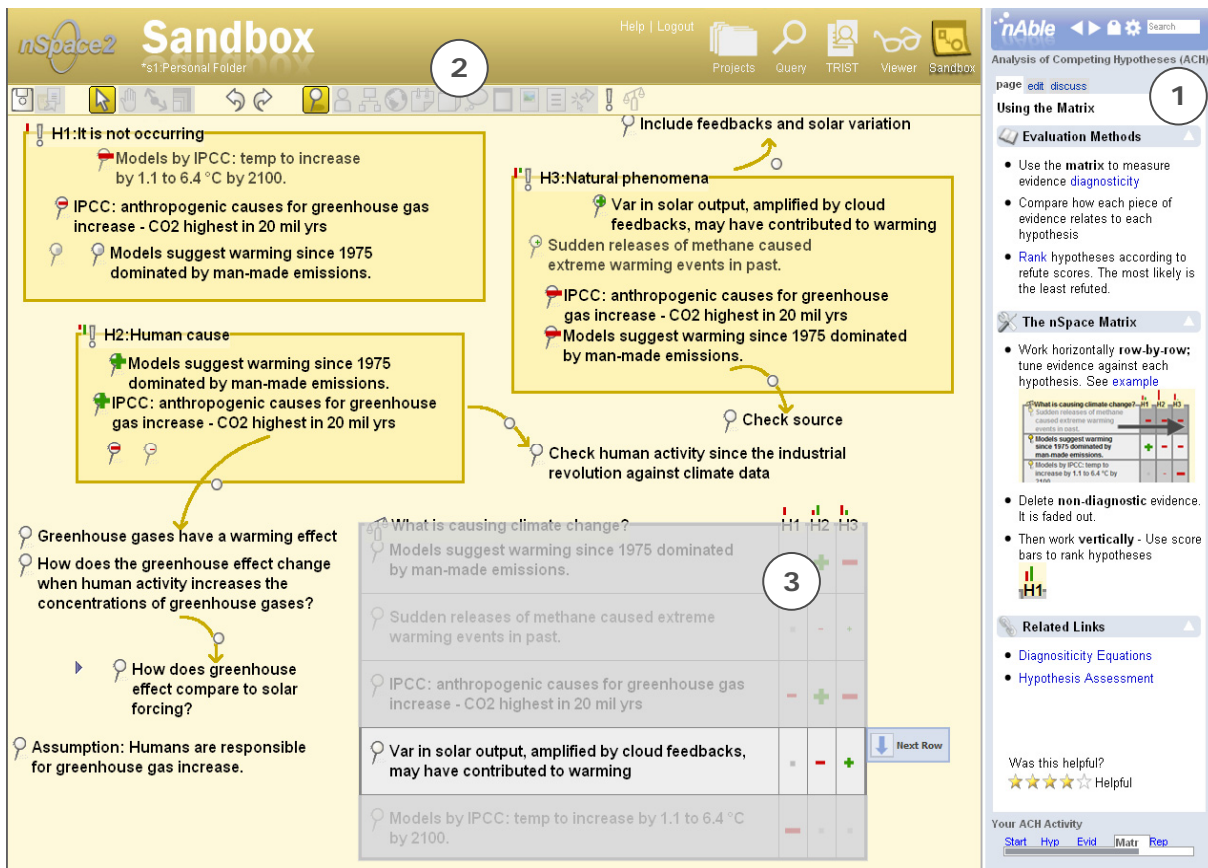


Figure 1. Examples of nAble’s web-based adaptive techniques during the ACH subtask assigning evidence. Wiki-style sidebar (1) for recommendation of information and subtask support with automated scaffold progression. Manage attention and scale complexity of (2) interface and (3) task with masking techniques.

The nAble scaffold described in this paper focuses on one formal analysis tool which implements the Analysis of Competing Hypothesis (ACH) methodology within the Sandbox. ACH is a well documented, formalized approach to weighing alternative explanations in order to minimize the likelihood of analytic errors and bias [5]. The nAble scaffold supports users in performing ACH subtasks, including *naming the issue, identifying hypotheses, gathering evidence, assigning evidence, assessing diagnosticity, reviewing the analysis and writing the report.*

nAble for nSpace2 is a set of network services providing the decision engine, task models and rich internet application for the adaptive scaffolding interface. The scaffold runs in the browser as a layer above the host application and queries the decision engine and task model using Service Oriented Architecture (SOA). User behavior is sensed through the browser, not the host application. Similarly adaptive techniques are sent directly to the browser. This gives the nAble scaffold autonomy from the host application, but consequently also requires a creative approach to implementing some of the adaptive techniques. Figure 1 shows the nAble scaffold for nSpace2 and highlights some of the techniques used in its delivery.

Scaling complexity required a new approach from the previous desktop nAble scaffold [9] since we could no longer directly modify the host’s interface to remove advanced features. We solved this problem using a masking technique to limit user interaction with advanced components while still providing visual information regarding the component’s purpose. The semi-transparent widgets provide the user information about which features were being scaled, which prepares the user for

interaction with that feature. In Figure 1, the system is masking items in the toolbar to reduce the interface complexity for the user. This is also an attention management technique, which guides user attention away from unneeded toolbar items in order to reduce cognitive load.

Recommendations and subtask support are presented to the user with nAble’s Wiki-style sidebar. The sidebar is the focal point for new help content, links to sites of interest and suggestions for expert collaborators. Content is segmented into information chunks, which are organized according to user and task models. Thus, recommended content is adapted to expertise, learning style, and personality traits as detected by the user model. Users can read, rate, tag and even contribute to recommended content. User tags and ratings are used in future decisions about content chunk selection for adaptations. Figure 1 displays the Wiki-style sidebar populated by the information chunks that are relevant to novice users performing the ACH subtask of assessing diagnosticity.

Feedback is given to users about task progression through an interactive progress widget located at the bottom of the Wiki sidebar, (3) in Figure 1. This widget provides users with an overview of the overall objective, feedback on current steps, and a sense of progress as steps are achieved. The steps and progression also adjust according to user expertise, as the steps for a novice user are simplified and initial progress is more visible.

Additionally, instructional pop-up dialogues are used to support bootstrapping, recommend information, manage attention and support sub-task. They actively push relevant information to

the user by appearing in contextually determined locations directly in the user's workspace, allowing the system to effectively "point" things out to the user.

The problem of providing scaled automation as an adaptive technique was also solved using masking techniques. As shown by (3) in Figure 1, when the user's subtask is assessing diagnosticity, the scaffold masks the ACH matrix to force the user through a row-by-row examination of each item of evidence, which is an important step in ACH methodology [5]. This automation is scaled back when the user achieves sufficient competence.

nAble uses a network service exposing a Hidden Markov task model of the ACH methodology. The scaffold can send information to the network service in order to determine which subtask the user is working on. Most of the adaptive scaffolding techniques coordinate with the task model service to determine the timing of each adaptation.

CONCLUSIONS

Adaptive scaffolding is a powerful instructional approach which allows novice users to perform at higher levels while they are learning new tasks and techniques. This paper has presented a taxonomy of adaptive scaffolding techniques that can be used by designers of adaptive interfaces to scaffold learners in complex software environments. A prototype scaffold for the ACH analytic methodology was also presented to demonstrate how key categories could be developed for a web-based analytic system. The nAble scaffold prototype leverages service-oriented architecture to create a scaffold that is autonomous from the host application.

Several avenues of future research are possible. While the generalizability of the taxonomy was tested through additional user interface design exercises and literature reviews, additional experimentation with implemented methods would confirm and refine the taxonomy. Further prototyping and experimenting is required to develop a broader portfolio of adaptive techniques and to evaluate the effectiveness of each. Although adaptive techniques from the taxonomy were implemented in the web environment to illustrate the particular challenges of scaffolding in web-based applications, the taxonomy can be applied across various technologies, including stand-alone systems and synthetic worlds. These adaptations will conceptually work across many domains; however variation in implementation will certainly be required to carry out these techniques.

This presented taxonomy and prototype are a part of a larger project for understanding and improving analytic workflow. We aim to develop a knowledge formalization for information analysis which allows sharing of knowledge between analysts and machine learning algorithms.

Next steps for the nAble scaffold are to build on the task recognition and training system to allow more sophisticated interface adaptations to analytic methodologies. More specifically, we are exploring machine-learning techniques to model various structured analytic methodologies, recognize when they are being used by an analyst, recommend additional

methodologies, and perform aspects of the task automatically where it would be beneficial. These adaptations would offer strong benefits to information analysis communities as new techniques could be captured and propagated as they are developed and used. We aim to develop a reasoning system which assists all users, novice and experts alike, in structured analytic techniques and informal critical thinking methods.

REFERENCES

- [1] Aleven, V. and Koedinger, K. 2002. An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26, 147–179.
- [2] Conati, C. and VanLehn, K. 2001. Providing adaptive support to the understanding of instructional material. In: *Proceedings of the 6th International Conference on Intelligent User Interfaces*. Santa Fe, NM, January 14–17, 2001. ACM Press, New York, NY. Pp. 41–47.
- [3] Fischer, G. 2001. User Modeling in Human-Computer Interaction. *User Modeling and User-Adapted Interaction*, 11(1–2), Kluwer Academic Publishers, Hingham, MA, 65–86.
- [4] Graesser, A., Chipman, P., Haynes, B. and Olney, A. 2005. AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *Education, IEEE Transactions on*, 48, 4, 612–618.
- [5] Heuer, J. 1999. *Psychology of intelligence analysis*. Center for the Study of Intelligence Central Intelligence Agency.
- [6] Jackson, S. L., Krajcik, J. and Soloway, E. 1998. The design of guided learner-adaptable scaffolding in interactive learning environments. In: *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*. Los Angeles, CA, April 18–23, 1998. ACM Press/Addison-Wesley Publishing Co. Pp. 187–194.
- [7] Jameson, A. 2002. Adaptive interfaces and agents. In: *Human-computer interaction handbook*. A. Sears (Ed.). Lawrence Erlbaum Associates Inc., Hillsdale, NJ. Pp. 305–330.
- [8] Jonker, D., Wright, W., Schroh, D., Proulx, P. and Cort, B. 2005. Information Triage with TRIST. In: *International Conference on Intelligence Analysis*, May 2–4, 2005, McLean, VA.
- [9] MacInnes, J., Santosa, S., Kronenfeld, N., McCuaig, J. and Wright, W. 2008. nAble Adaptive Scaffolding Agent – Intelligent Support for Novices. In: *Web Intelligence and Intelligent Agent Technology*, Sydney, Australia, December 9–12, 2008. WI-IAT '08. 3, IEEE Computer Society, Washington, DC. Pp. 365–368.
- [10] Wood, D., Bruner, J. S. and Ross, G. 1976. The role of tutoring in problem solving. *J Child Psychol Psychiatry*, 17(2), 89–100.

Session 14: Understanding Social Dimensions of Activity

Understanding Social Entities, Activity Coupling and Social Computing Affordances in Industrial Work Environments

Heljä Franssila
 Department of Information Studies and
 Interactive Media
 University of Tampere
 33014 University of Tampere, Finland
 helja.franssila@uta.fi

Petri Mannonen
 Department of Computer Science and
 Engineering
 Helsinki University of Technology
 P.O. Box 9210,
 02015 TKK, Finland
 petri.mannonen@hut.fi

ABSTRACT

In this paper we develop a conceptual framework and apply the framework in an case study about exploring intersections of work activity coupling, social entities, social capital and social computing tools in industrial work settings, a process industrial multi unit plant as an empirical example. The conceptual framework can be utilised in the analysis and understanding of interaction and communication practices in the work settings. It can be utilised as an evaluation framework when selecting social computing tools and understanding their potential functional affordances to support communication and coordination in distributed process control work settings.

Keywords

work activity coupling, work system analysis, social interaction in work, social entities, social capital, process industry, evaluation framework

ACM Classification Keywords

H.1.2 [User/Machine Systems] Human factor; H.5.3 [Group and Organization Interfaces] Computer-supported cooperative work, Evaluation/methodology, Web-based interaction.

INTRODUCTION

The power of social ties, connections and different kinds of social conglomerations as a source of valuable capabilities and resources in knowledge intensive work environments has recently gathered considerable interest. Concepts like community, network and crowd appear often in the discussions, especially when different kinds of activity patterns in the online social interaction are touched. The classical debate about the differences and similarities between offline and online communities and networks illustrates the quest for better understanding of these social entities as enabling and grounding intelligent human activity, be it online or offline.

Respectively, it is challenging for industrial work organisations to evaluate which of the social computing tools could

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

appropriately serve their work environments and the work activities their employees are carrying out. Informal interest-based interaction entities like communities of practice and networks of interest supplement and challenge the more formal and traditional forms of work activity coupling like teams and formal work groups. Industrial work organisations have started to seek possibilities to exploit the potential of communities, networks and crowds (e.g.[13, 6]) ICT-companies providing information technologies for diverse organisational use are increasingly integrating social media applications into their portfolios and a plethora of dedicated social applications, both free and commercial is available also for business use. Work organisations in diverse industries are already consciously cultivating and supporting the communities and networks existing among their employees and partners [17]. Still the crucial questions for a single company exists: What are the social entities present in their working ecosystems, what are the qualities and features of them, and what are the feasible possibilities to harness the potential of them for purposes of performance development?

An evaluative framework which help to identify the continuum of social entities and the levels of actor-based work activity couplings present in information-intensive industrial work environments is needed (compare [30]). Detailed understanding of the types of communicative interaction patterns which emerge in the intersections of social entities and the different work coupling intensity levels helps to identify which functional affordancies (see [9] about different facets of affordances) of social computing tools provide proper support for these work activities.

In this paper we develop a conceptual framework and apply the framework in an exploratory case study about intersections of work coupling, social entities, social capital and social computing tools in industrial work settings, a process industrial multi unit plant as an empirical example. We present both conceptual grounding and example of practical application of a framework for understanding social entities, activity coupling and social computing affordances in industrial work environments. The framework can be utilised in the evaluation of social computing tools which are aimed to support communication and coordination in distributed process control work settings.

We explore the levels of actor-based work activity coupling with help of following dimensions of coupling: amount, intensity and predictability of work activity coupling. Next we discuss based on the literature a conceptual continuum of social entities involving actor-based interaction identified in industrial work settings. We discuss both the formal and informal social entities present in the industrial work environments. As an additional analytical framework for analysing social entities and activity coupling we utilise the theory of social capital. Social capital is divided into three main dimensions: structural, relational, and cognitive [22]. The conceptual continuum of social entities

reflects the differences of social, motivational and cognitive properties that characterise different kinds of social entities.

We argue that the dimensions of social capital help to differentiate and understand the elemental differences between social entities operating in the work settings, for example between work shifts, work teams, communities of practice and networks of interest in process industrial work environments. These differences have important implications for the possibilities to harness these social entities for business purposes, for example in the rapid troubleshooting, collective problem solving, collecting organisational memory and experience knowledge cultivation.

WORK ACTIVITY COUPLING

To understand social interaction and communication related to the work, it is essential to understand the grounds and motivations for interaction. One profound possibility to understand and model interaction and communication (and its absence) realising in the work domains and among work related social entities is to analyse the nature and reasons of *coupling of work activities* as the driver for interaction and communication.

In general level work coupling can be seen as the linkage of organization's tasks, duties and activities. Activities are goal-directed bunches of related tasks. When the coupling of these elements links also organizational entities, i.e. people, teams and ICT systems, work activity coupling takes place. Work activity coupling is one of the reasons for collaboration and group work and different kinds of couplings require different levels of collaboration and coordination [23].

There exist different definitions and approaches to work coupling. The concept of work coupling has been defined at least in two differing ways in the literature. The first one is applied in the cognitive ergonomics and another one in CSCW literature. The first one is from Rasmussen et al. [26]. The approach is based on the analysis of the goal, resource and task dependencies in the work system, and they derive the definition of work coupling from understanding these functional dependencies and the regularity and predictability which those dependencies obey or not.

Rasmussen et al. [26] generated a systems analytic continuum of work systems which illustrates one interpretation of work coupling concept. Work systems in general contain a set of enabling means-ends elements and relations – goals, functions, activities, constraints and productive resources. These means-ends elements can have varying amounts of many-to-many relations with each other, thereby generating different levels of complexity into the work activity systems and also creating different levels of controllability to the system. Goals are typically pursued by purposeful execution and control activities of humans and machines which interactively transform energy, material, human and information resources into concrete or abstract products and services.

Especially the successfulness of control activities depends on the regularity and predictability of the processes in work system. There are two different sources of regularity – the functional, causal structure of the work system and the intentional structure (of the actor population) of the work system. Functional structure is typically quite stable and most often governed by the physical laws of nature. Intentional structure instead is governed by the human intentions, preferences, values, policies and practices. The source and level of regularity of work system influence the level of work coupling in the system. Combinations and strengths of these structural elements explain the continuum of work coupling in the work systems.

Rasmussen et al. [26] propose, that primary source of regularity and predictability in industrial work systems like chemical

process plants are the laws of nature governing the function of physical processes and equipments. These kinds of highly complex, technical work systems are *tightly-coupled*, and by definition their functioning is predictable and unambiguous. Predictability of the work activities of the operators is primarily governed by the predictability of the physical connections, functioning and processes of the plant equipment. Personal intentions and working habits of the operators are only secondary (but not totally insignificant) compared to the overall production and automatic control goals, which are hard-wired in the physical design of the equipment and process. It is of critical importance, that communication and information applications in this kind of work system environments support the task of recognising, understanding and communicating the many-to-many connections in the physical process and equipment and the potential disturbances in the plant functions.

Another end of the work coupling continuum is the *loosely coupled* work systems where the goals, rules of conduct, intentions and preferences of the management and workers create the level of regularity of the work activities, and the physical processes are only secondary to the human intentions and just materialise the human intentions, which may vary a lot. Physical equipments and processes allow flexibility to materialise human intentions, they can be used for multiple purposes. According Rasmussen et al. [26], these kinds of work systems one can find for example from offices, hospitals, manual manufacturing shops etc. In this kind of work systems, physical environment and equipment do not govern, only serve as a context and resource for the activities of the workers, who make situated decisions what to do and when based on shared and personal strategies, objectives, rules and preferences. The workers have got freedom to decide how to realise the institutional objectives by utilising the resources available. The work activity system is more complex and dynamic than in the tightly-coupled systems. Of critical importance in these kind of work activities is the understanding, communication and awareness of the workers about each others objectives, strategies, schedules and information resources.

The second approach to understand work coupling is exercised especially in the CSCW, and it concentrates into information sharing and communication in the work. Coupling in work refers to the demands of information sharing [23] and the level and type of communication needed in the work between individual workers [23, 24]. The amount of individual work that can be done before one has to interact and communicate with another also reflects the level of work coupling in that work [23].

Work activity coupling can be categorized both based on the level and the type of it. The level of coupling refers to the intensity of information sharing and required communication [23]. Different types of coupling include for example information and command coupling [28]. In addition the different communication types or media (face to face, telephone, video conference, chat, e-mail etc.) and complexity of tasks (routine vs. ambiguous) can be seen as types of coupling [24]. Thus work activity coupling is in general level seen in this approach as a continuum from lightweight or loose coupling to tight or closely coupled work (e.g. [24, 23]. Neale et al. [23] name the levels of continuum as: lightweight interaction, information sharing, coordination, collaboration, and cooperation.

Interestingly, Olson and Olson [24] propose that loosely coupled work is typically routine, unambiguous and workers tasks have fewer dependencies. Tightly coupled work is nonroutine, ambiguous and components of the work are highly interdependent. Olson and Olson [24] do not name or speculate grounds for the ambiguousness or nonroutiness of certain work as Rasmussen et al. [26] do, but explain that the ambiguousness,

nonroutiness and multiple dependencies are the defining characteristics of tightly coupled work. This is in opposition to the Rasmussen's definition, where the grounds for the intensity level of work coupling are in the *predictability of dependencies* of the work system, not in the amount of dependencies per se. An actor's ability and freedom to plan and execute her own work activity in relation to others' activities depends on the knowledge of sources and nature of the regularity/irregularity in the dependency relations of the work system [26].

As an elaboration of Rasmussen et al. [26] ideas of continuum of work coupling, there can be work systems where dependencies deriving from either functional or intentional structure or from both explain the level of work coupling. The mutual intensity of influence of these two structural drivers varies in different work systems. The dependency structures in the work system create the internal constraints of the work system, and mutual awareness, knowledge and motivation to take into account these dependencies shape the success of the operation of the work system. When it requires extra effort to maintain this awareness of dependencies in the work system, the work is loosely coupled.

Not any work system and workers' activities within that system are all the time tightly/closely coupled or loosely/lightly coupled. In every work domain there are moments and episodes when the level of irregularity and unpredictability rises and episodes when the system is stable and predictable (cp. [24]). For example in chemical plants in their distributed control room environments, the work in normal, stable situations is tightly-coupled, but there are disturbance situations which require special attention and adaptive strategising from the operators. Also in the office work or symbol analytic knowledge work there are periods of routine work, while most of the time the work can be characterised as loosely coupled, nonroutine work.

The difference between the work coupling definitions of cognitive ergonomics and CSCW seems to lay in the focus and level of analysis. Rasmussen et al. [26] analyse the profound dependencies and their sources in the work as the main reason or explanation for the work coupling – the *cause of coupling* – and Olson and Olson [24] and Neale et al. [23] analyse *the effect of coupling* – the need for communication and information sharing in the work – as the essence of coupling concept. These both stances are of empirical importance, but in this article we elaborate further these interpretations.

Nor Rasmussen et al. [26] neither Olson and Olson [24] explicate the profound types or qualities of dependencies which the both functional and intentional structure can create – how the dependency can be experienced by a single actor or social entity in the work system, and what kind of dependency relations organise the interplay of individual and shared work activity? Neale et al. [23] touch briefly this issue when they refer to the amount of individual work that can be done without needing interact and communicate with others as the reflection of the level of work coupling. They also list elements of work which the actors share when work coupling intensifies: goals, plans, common ground and concurrent activities. However, these elements characterise more or less all kinds of goal-oriented shared work executed by a social entity, and they do not differentiate tightly-coupled work from loosely coupled work. Still the question how exactly one activity is dependent on another and how this interdependency can be observed and measured is left open. What kind of dependency elements are present when one actor cannot work independently and without other's contributions as a constraint for one's activity?

DEPENDENCIES BETWEEN WORK ACTIVITIES

To pass the terminological incompatibility of cognitive ergonomic and CSCW use and definition of the work coupling, it may be fruitful to deepen and concretise the understanding of the nature of the dependencies and their regularity which creates work activity coupling phenomenon. Malone and Crowston [21] provide useful framework for classification of dependencies requiring coordination and communication in co-operative work. They name four dependency sources: shared resources, producer-consumer relationships, simultaneity constraints and task-subtask-composition. Dependency of activities realises when activities share certain resource, be it time (people or equipment time), money, materials, information possessed only a certain actor or any other resource needed in the accomplishment of the activity goal. Another general category of dependency is producer-consumer relationship, where one activity produces something that is used by another activity. When certain activities must happen in the same time or they cannot occur in the same time, the dependency relationship is called simultaneity constraint. Finally, a very common dependency relation is a task-subtask relationship where a group of interrelated activities are subgoals of a certain overall goal. Later Malone [20] simplified the dependency types into three classes: *flow dependency*, where one activity produces a resource that is used by another activity; *sharing dependency* where multiple activities all use the same resource; and *fit dependency* where multiple activities must fit together to jointly produce a single resource. Interestingly, central in the Malone's analysis is the concept of *resource*. All of the dependency types are somehow related to the management of resources.

It can be speculated, that any amount of ambiguousness, nonroutiness and unpredictability in the everyday management of these dependencies will create greater need for communication and shared awareness maintenance between actors. Social entities and the conventions of their interaction and communication practices in the work can be interpreted as functional mechanisms and strategies to manage these versatile dependencies.

For the purposes of the evaluation framework we state based on the above discussion our definition of work activity coupling as following: the qualitative nature of work activity coupling is based on the *amount and type of known dependencies, and predictability of dependency effects on work activities of different actors pursuing both individual and shared goals*. The work activity coupling is *intense* when there is *multiple known dependency types* present in the shared activity. The work coupling is *predictable* when the number and type of dependencies between activities remain relatively stable and their effects are predictable. There are work systems where the both the amount and type of dependencies differ for example between projects, and where the predictability of effects of the dependencies vary from work case to case.

SOCIAL CAPITAL AND WORK ACTIVITY COUPLING

Intense and unpredictable work coupling generates need for effective and appropriate information sharing and communication [23, 24]. Creating a work system where information sharing and communication conventions and tools are effective and they fit the needs of both individual and co-operative tasks and user features is far from straight-forward. Certain information sharing and communication conventions and protocols do not fit every work system, and there are lots of examples of failed

information sharing and communication solution implementations in work organisations. Typical reasons for the failure are that the needs, drivers and motivations for information sharing and communication among actors of the work system are not well understood, they are misunderstood or at least the influence of new ICT tools for the information sharing conventions is predicted wrongly [7, 8]. One explanation for the failures can be incomplete understanding of the nature of work activity coupling as profound driver of information sharing and communication practices.

Recently, when the amount and types of technologies and tools to support information and sharing in organisations has expanded tremendously especially due to the new, easy-and-inexpensive-to-deploy web technologies, the application of social capital concept as an analytic device for understanding conditions for information sharing has evolved [10, 11, 17]. When mainly technology-driven and top-down knowledge sharing and online community development initiatives and projects in work organisations have flopped, and at same time online communities and collaborative content creation flourish in the free time virtual environments, questions have arisen if something crucial about the drivers of organisational information and knowledge sharing and communication is poorly understood.

Nahapiet and Ghoshal [22] have created a conceptual framework which describes how multidimensional social capital facilitates the development of the conditions and ability of an organisation to utilise intellectual capital. At the core of this ability are the processes of access, anticipation, motivation and capability to combine and exchange information resources within the social communities of the organisation. Understanding the existence or absence of social capital in the relationships of social entities operating in work context potentially helps to understand also the conditions for the successful application of the new social computing capabilities.

Social capital has several definitions in the literature, but Nahapiet and Ghoshal [22] define social capital as “the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit.” This definition correlates nicely with Malone’s analysis of general dependence types, which all relate somehow to the coordination of resources of the activity.

Social capital inheres in the relationships between persons and among persons, and it is owned by the parties involved in a relationship. It is very difficult to trade or buy, and its value lay in its use to facilitate activities between the parties. Social capital is often crucial resource for the achievement of certain co-operative ends. In other words, if it does not exist between the parties, the co-operative goal will not be reached [22].

Social capital consists of three dimensions: structural, cognitive and relational dimension. Each of these dimensions separate into facets. Structural dimension relates to general pattern of connections between actors and to the actor’s possibilities to reach and contact each other as a resource for action. This connection pattern which an individual or social unit is part of can be characterized by network ties, network configuration and appropriable organization of networks, which means that the networks created for one purpose, can be used for another purpose. Relational dimension refers to assets which are related to trust, shared norms and obligations between actors and to their identification with the social unit they are member of. Cognitive dimension refers to the enabling cognitive resources and common ground which the actors and social units master with their co-actors like shared language, narratives, representation styles, codes and systems of meanings.

The most profound influence social capital has on the interaction of individuals and social units is related to the efficiency, adaptability and creativity of interrelated activities. It can have also drawbacks, for example when mutual norms and identification inhibit creative and flexible action.

If social capital is central for development of the information exchange and combination capabilities of a social community, a question arises how social capital operates as an interaction and communication driver in work domains with differing levels of intensity and predictability of work activity coupling? It is hypothesized in this paper that the nature and intensity of work activity coupling is connected to the existence and extent of social capital among the actors of the coupled work, and further, that the work activity coupling level and existence and extent of social capital jointly explain the diversity of social entities present in certain work settings.

SOCIAL ENTITIES IN WORK SETTINGS

So far we have discussed the grounds for the work related interaction and communication and hypothesized that work coupling can be the driver and social capital the fuel for the social interaction and communication in work settings. What kind of social entities exist in work systems and how different social entities might coexist with the level of work coupling in different kinds of work systems? Or do the different kinds of social entities respond for the different modes of the work coupling within the same work system? Next we discuss based on the literature the continuum of social entities present in the work settings.

The understanding of different kinds of social entities potentially present in work settings has developed enormously within last decades. Teams, groups, group processes in the work, computer supported cooperative work, and design of organizations have been long-term interest in the social, behavioural and management sciences (e.g. [29, 12, 14, 18, 2]). The shift from the quite stable, local and bureaucratic form of task and work design of industrial age towards more flexible ways to organise work reflected the globalisation and pressures work organisations experienced when trying to adapt to the changing operative environments and markets.

Especially in the nineties practical work organisation design was influenced by the strong team approach, and the ideas of self-directed and multi-skilled teams applied also into the industrial work environments. Again, the team movement was quite direct response to the demands of the new business environment. Competition was tough, and work organisations and enterprises needed to focus into their core competencies and outsource functions which were secondary to the core competence, shorten delivery and throughput times, deliver products and services just-in-time, and minimize inventories. All these developments at the same time multiplied the number of co-operating actors, new interdependencies between functions, activities and actors, and intensified the need for explicit coordination, information sharing and communication.

In the nineties, the interest shifted into the more informal social entities and networks, which also crossed the borders of one workplace and surpassed the co-presence limitations. At the same time, information, knowledge and competence were seen as the most important assets of work organisations and enterprises, but also the tremendous difficulty to commoditize, transfer and share these assets was seriously faced and understood. The demands of responsiveness, flexibility and speed were even more intensive. The importance of social interaction as the medium and mode of efficient knowledge transfer and utilisation was recognised in a large scale. The concepts of community of practice [16, 3, 32] and learning organisation [27] emphasised the social interaction, participation and everyday

adaptive and creative practices in the work as important drivers of work performance and development, and also as a target of managerial recognition. The concepts of trust especially in the examinations of networking within and between organisations become central theme [15].

After the introduction of community of practice concept several other information and knowledge centred community concepts like informal communities, strategic communities, informal networks and interest groups, have evolved [1]. The breakthrough of community concept into the interests of social and behavioural scientists of work and technology and information scientists in this decade [31] reflects the wide-spread adoption of internet technologies, online communities [25] and social computing applications [19], which are accessible for nearly everyone. In the accounts of diverse communities as social entities in work settings the importance of shared language, concepts, meanings and narratives, common ground, trust, reputation and mutual accountability is expressed frequently.

The evolution of recognition and understanding of diverse and overlapping social entities in work settings can be seen as reflection of the change in the operation and co-operation environment of work organisations in the last decades. It is one of the hypotheses of this paper that the number and diversity of work systems where work activity coupling has intensified has expanded.

What then are the qualities of social entities operating in the work settings where the actors must manage the effects of multiple and changing dependencies? We argue, that the level of all the dimension of social capital must be rather high within the relationships on the social entities operating in work settings which is characterised with intense work activity coupling. We also hypothesize, that a work activity where there are multiple but still unpredictable and changing dependencies requires greater

amount of social capital to be managed well, and the pursuit to fulfil this requirement finds its realisation in the activeness and multitude of work-related social entities where the workers actively participate. Contrary, when there are lots of dependencies but their logic and effects are predictable and stable, social capital and diversity of social entities might not be that central for the success of the performance of the work system..

EVALUATION FRAMEWORK FOR THE INTERPLAY OF WORK ACTIVITY COUPLING, SOCIAL CAPITAL AND SOCIAL ENTITIES IN WORK SETTINGS

The discussion and hypothesis presented in the above sections can be collected into conceptual framework which collects the dimensions and continuums of work activity coupling, social capital and of social entities in work settings (see Table 1.). These dimensions separate into subcategories and sub-dimensions according their conceptual structure.

The framework can be utilised as conceptual grounding and starting point for the work system analysis via observations, interviews and surveys. In the next section we provide a brief example of the utilisation of framework in the qualitative analysis of work activity coupling and detection of social entities in multisite chemical process plant with the goal to find appropriate social computing application to support the interaction and communication within the social entities.

Most straightforward and practical way to use the framework is to qualitatively score based on the interview or survey data first the level of work activity coupling, next to examine the potential social entities present in then work system and next to analyse the level of social capital dimension within the social entities found.

Table 1. The conceptual framework for the evaluation of work activity coupling, social capital and social entities in work settings.

Evaluation concept	Operationalisation of the concept as dimensions or continuum
Work activity coupling	Amount of dependencies among activities of the actors (hug – low) Types of dependencies between activities (scarce – many) <ul style="list-style-type: none"> - flow of resources - sharing of resources - fitting activities to create to jointly produce a resource [20] Predictability level of each dependency [26] (high – low)
Social capital	Structural dimension of the actor-actor relationships (high – low) <ul style="list-style-type: none"> - network ties - network configuration - appropriable organization Relational dimension of the actor-actor relationships (high – low) <ul style="list-style-type: none"> - trust - norms - obligations - identification Cognitive dimension of the actor-actor relationships (high –low) <ul style="list-style-type: none"> - shared codes and language - shared narratives, meanings and representations [22]
Social entities	A continuum from more formal to less formal social entity: (present – not present) <ul style="list-style-type: none"> work shifts functional work groups self-directed teams communities of practice [32] knowledge communities [1] communities of interest [5] networks of interest

APPLICATION OF THE FRAMEWORK TO EMPIRICAL ANALYSIS AND UNDERSTANDING SOCIAL COMPUTING AFFORDANCIES

As an empirical data to test the conceptual framework we collected exploratory and follow-up semi-structured interviews and a minor qualitative survey data from the employees of five distributed production units of a chemical plant. The interview protocol operationalised the concepts of the evaluation framework presented above. The goal of the observations and interviews was to understand the nature of social interaction, communication and activity dependencies between actors within and between production units of the site. We also explored to communication and information sharing tools used at the moment for different interaction and communication purposes.

The interview respondents were operators, experts and maintenance personnel of the five interrelated production units. We interviewed 2–5 operators and members of maintenance personnel and all of the 3–5 experts in production units in their conventional work environment in the units during their work shift.

First based on documentary material and interview with the expert and production personnel we constructed a configuration map about the material and informational resource flows between the actors within one unit and between the units to understand the flow, sharing and fitting dependencies in the process technical system level. The results of the mapping showed that in overall there were only couple of material and resource flow interdependencies between the production processes of the units, but the dependencies were critical in the sense that disruption of the delivery of a certain resource would disturb and stop quite soon the operation of the production unit chain. Even though the basic process technical dependencies were well known, for example variation in the quality of the raw material which one unit produced to be consumed by another unit was creating unpredictability into the interdependencies.

Also different kinds of shortages of certain materials and especially energy created ambiguous and nonroutine dependencies between the operation activities executed in the units. When these kinds of disturbances occur, the work activity coupling is more intensive than in the normal operation periods.

Next we depicted the typical interaction and communication episodes within and between units with the help of exploratory survey. We applied the three dimensions of the concept of social capital (structural, relational, and cognitive) as a conceptual framework to explain the nature of communicative interactions emerging within and between work shifts in one unit and between work shifts of the separate units.

We also analysed what kinds of active social entities there was present within single production units and between the units in addition to the shifts and functional professional works groups like production operators and maintenance personnel. As the result of the analysis we found that while there are multiple material and informational dependencies between the production units and their personnel, they are most of the time very predictable and the knowledge of the dependencies is widely shared between actors. The level of structural and cognitive dimension of social capital in the relationships within and between the production unit work shifts was quite high. The relational dimension was high in the relationships within work shifts of the units, but only moderate between work shifts of separate units.

The more informal social entities reflected the profession division, so that production staff, maintenance staff and expert staff seemed to constitute also separate communities of practice based on the profession, but a single multi-professional community of interest within on unit. The community of practice of experts also expanded over the unit borders, but there were not any active communities of interest between the production staff of the units.

Based on the empirical observations concerning the social interaction and communication episodes, social entities and the level of social capital dimensions we formulated a preliminary proposal about promising social computing applications to support the information sharing and communication (Table 2.)

Table 2. Promising social computing affordances supporting social interaction and communication needs within social entities present in process industrial work settings. Modified from Franssila and Mannonen 2008.

Social interaction and communication need experienced in the social entities	Information supporting the need	Promising social computing applications supporting the need
maintaining awareness of the personnel doing manual work at the present moment in the field in the plant area	information about the presence and identity of the co-workers, what are their goals, activities and artifacts they are working with; information about common ground	blogs, micro-blog, activity streams, object hyperlinking
maintaining awareness about the present coordination needs and coordination activities between units	information about goals, activities and artifacts co-workers are working with, information about common ground	chat, activity streams, micro-location tracking, mashups
maintaining awareness about the operational situation and activities in the same unit between different time moments	information about co-workers activities, information about common ground	blogs, micro-blogs, discussion forums
maintaining awareness about the operational situation and activities in other units between different time moments	information about co-workers activities, information about common ground	chat, blogs, micro-blogs, discussion forums
maintaining awareness about the past disturbance situation solutions and their applicability into a new process situation and context in the same unit	information supporting the creation of common ground, CoPs, social capital	mashups, video blogs, discussion forums

Couple of these social computing tools (chat, discussion forums and blogs) and the suitability of their functional affordances were piloted in the chemical production site. The pilots and their reporting are in progress.

DISCUSSION

Understanding of the grounds, motivations and practices of everyday social interaction and communication in work settings can be enhanced with the help of conceptual evaluation framework developed in this paper. This understanding can be utilised in the assessment and comparison of different social computing tools and their affordances.

A weakness of the evaluation framework presented in this paper is that there does not exist yet established variable sets that would operationalise the key characteristics of different social entities for evaluation purposes and which would reflect the level of formality in their nature and operation. However, there are couple of operationalisation published which are used to analyse the potential existence of community of practice [33] and sense of community [4].

The evaluation framework should be further elaborated and operationalised into a more comprehensive survey instrument to enable more detailed assessment of its feasibility and usability as a tool for detecting important social entities in work settings.

REFERENCES

- [1] Andriessen, J. H. E. 2006. Archetypes of Knowledge Communities. In: van den Besselaar, P., De Michelis G., Preece, J. and Simone, C. (Eds.). 2005. *Communities and technologies 2005: Proceedings of the second Communities and Technologies Conference, Milano, 2005*, Springer: Dordrecht. Pp. 191–213.
- [2] Baecker, R. M. (Ed.). 1993. *Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*. Morgan Kaufmann Publishers: San Mateo.
- [3] Brown, J. S. and Duguid, P. 1991. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science*, 2(1), 40–57.
- [4] Chibuer, H. M. and Pretty, G. M. H. 2000. A review of the sense of community index: Current uses, factor structure, reliability, and further development. *Journal of Community Psychology*, 27(6), 643–658.
- [5] Fischer, G. 2001. Communities of Interest: learning through the Interaction of Multiple Knowledge Systems. In: 24th Annual Information Systems Research Seminar In Scandinavia (IRIS'24), Ulvik. Pp. 1–14.
- [6] Franssila, H. and Mannonen, P. 2008. Affordances of Web 2.0 in supporting shared situation awareness in distributed process control communities. Paper presented in Internet Research 9.0: Rethinking Community, Rethinking Place -conference 15–18 October 2008, Copenhagen, Denmark.
- [7] Grudin, J. 1988. Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In: *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work*. 26–28 Sept 1988, Portland, Oregon, United States. Pp. 85–93.
- [8] Grudin, J. 1994. Groupware and the Social Dynamics: Eight challenges for Developers. *Communications of the ACM*, 37(1), 92–105.
- [9] Hartson, R. H. 2003. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology*, September–October 2003, 22(5), 315–338.
- [10] Huysman, M. and Wulf, V. (Eds.). 2004. *Social Capital and Information Technology*. MIT Press: Cambridge, Massachusetts.
- [11] Huysman, M. and Wulf, V. 2006. IT to support knowledge sharing in communities, towards a social capital analysis. *Journal of Information Technology*, 21, 40–51.
- [12] Katzenbach, J. R. and Smith, D. K. 1993. *The Wisdom of teams: Creating the High-Performance Organization*. Harvard Business School Press: Boston.
- [13] Koskinen, T. 2006. Social software for industrial interaction. In: *Proceedings of the 18th Australian Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. Pp. 381–384.
- [14] Kozlowski, S. W. J. and Bell, B. S. 2003. Work groups and teams in organizations. In: Borman, W. C., Ilgen, D. R., Klimoski, R. J. (Eds.). *Handbook of psychology: Industrial and organizational psychology*, Vol. 12. Hoboken, NJ, US: John Wiley & Sons Inc. Pp. 333–375.
- [15] Kramer, R. M. and Tyler, T. R. 1996. *Trust in organizations: frontiers of theory and research*. Sage: Thousand Oaks.
- [16] Lave, J. and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University press: Cambridge.
- [17] Lesser, E. L. and Storck, J. 2001. Communities of practice and organizational performance. *IBM Systems Journal*, 40(4), 831–841.
- [18] Levi, D. 2007. *Group Dynamics for Teams*. 2nd Edition. SAGE: Los Angeles.
- [19] Lietsala, K. and Sirkkunen, E. 2008. *Social media: Introduction to the tools and processes of participatory economy*. Hypermedia Laboratory Net Series 17. University of Tampere: Tampere.
- [20] Malone, T. W. 2004. *Future of Work*. Harvard Business School Press: Boston.
- [21] Malone, T. W. and Crowston, K. 2001. The Interdisciplinary Study of Coordination. In: Olson, G. M. et al. (Eds.). *Coordination Theory and Collaboration Technology*. Lawrence Erlbaum Associates: Mahwah, New Jersey, 7–50.
- [22] Nahapiet, J. and Ghoshal, S. 1998. Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, 23(2), 242–266.
- [23] Neale, D. C., Carroll, J. M. and Rosson, M. B. 2004. Evaluating Computer-Supported Cooperative Work: Models and Frameworks. In: *Proceedings of the CSCW'4*, November 6–10, 2004, Chicago, Illinois, USA. Pp. 112–121.
- [24] Olson, G. M. and Olson, J. S. 2000. Distance Matters. *Human-Computer Interaction*, 15, 139–178.
- [25] Preece, J. 2000. *Online Communities: Designing Usability, Supporting Sociability*. John Wiley & Sons: Chichester.

- [26] Rasmussen, J., Mark Pejtersen, A. and Goodstein, L. P. 1994. *Cognitive Systems Engineering*. John Wiley & Sons: New York.
- [27] Senge, P. M. 1990. *The Fifth Discipline: the Art and Practice of the Learning Organization*. Doubleday Currency: New York.
- [28] Sutcliffe, A. G. and Minocha, S. 1999. Linking Business Modelling to Socio-technical System Design. *Lecture notes in computer science*, 1626, 73–87.
- [29] Sundstrom, E., McIntyre, M., Halfhill, T. and Richards, H. 2000. Work Groups: From Hawthorne Studies to Work teams of the 1990s and Beyond. *Group Dynamics: Theory, research, and Practice*, 4(1), 44–67.
- [30] Weiseth, P. E., Munkvold, B. E., Tvedte, B. and Larsen, S. 2006. Wheel of Collaboration Tools: A Typology for Analysis within a Holistic Framework. In: *Proceedings of the CSCW'06*, November 4–8, Banff, Alberta, Canada. Pp. 239–248.
- [31] Wellman, B., Boase, J. and Chen, W. 2002. The Networked Nature of Community: Online and Offline. *IT&Society*, 1(1), 151–165.
- [32] Wenger, E. 1998. *Communities of Practice. Learning, Meaning, and Identity*. Cambridge University Press: New York.
- [33] Verburg, R. M. and Andriessen, J. H. E. 2006. The Assessment of Communities of Practice. *Knowledge and Process Management*, 13(1), 13–25.

Investigating Accountability Relations with Narrative Networks

Daniel Boos, Gudela Grote, Mikko Lehtonen
 Department Management, Technology and Economics
 ETH Zurich, Kreuzplatz 5
 8032 Zurich, Switzerland
 {dboos, ggrote, mikkol}@ethz.ch

ABSTRACT

In this paper, we describe a novel approach to investigate how accountability relations change alongside the introduction of a new Internet of Things application. The approach is based on the narrative network approach which was extended to study accountability relations. We show how this approach is useful to prospectively investigate critical accountability relations already when a new application is being developed.

Keywords

designing for accountability, accountability relations, narrative network, method, internet of things

ACM Classification Keywords

H.5.3 [Group and Organization Interfaces (Evaluation/Methodology)].

INTRODUCTION

Calls to increase accountability, or more specifically, to extend audit practices, are regarded as a part of a larger societal trend [17] and Information and Communication Technologies (ICT) are regarded as useful to answer this call [3,21,22]. Assumption that technology can improve accountability is an argument for funding, researching, developing and introducing new technologies like the Internet of Things [10].

Research on the use of information systems shows the importance to study the link between accountability and new ICT technologies. In the early stage of the development of a new technology designers and involved participants inscribe their tastes, competences and prejudices about a technology [14]. How a new system achieves accountability and who might be held accountable by whom are partially inscribed and prescribed during the process of system development [12]. A new technology may change or even compromise existing accountability relations in an unexpected way leading to conflicts, circumventions and resistance [19, 21, 22]. For instance, in one well-reported case the introduction of a new information system for radiological reports in a hospital created conflicting accountabilities, leading to severe tensions among the medical personnel. The problem was that the same report

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

was supposed to serve multiple audiences with different needs; while clinicians needed information about the patients' medical condition, radiologists intended to account for the results of the radiological study [22].

Many studies only look retrospectively at accountability changes in already deployed or partly deployed ICT systems and provide a critique or recommendations on how to change an already running system to improve problems deriving from misaligned or conflicting accountability relations. However, because of path-dependencies, it might be difficult and costly to resolve encountered accountability issues in a later stage.

We argue that a more promising approach is to look at accountability prospectively in the early design stages. Such an approach allows to more clearly envision the impact of a new technology on newly created and already existing accountability relations. Findings may form the basis for selecting a technical design solution [22], clarify accountability relations to prevent misalignments in advance and foster discussions about how accountability can be achieved and who should become accountable. Such an approach allows to satisfy calls to increase the accountability of designers for their products already during the design stage [1, 9].

In this paper, we present such a new approach to investigate *accountability relations* – relations between actors where one is accountable towards the other – of a new Internet of Things application using *narrative networks* [15]. We extended the narrative network approach to investigate the early stage of technology development focusing on accountability relations.

DESIGNING FOR ACCOUNTABILITY

Several proposals on how to deal with accountability in the design of information systems already exist. Ethnomethodological studies recommend that social actions and interactions with systems should be designed so that they are accountable, that is observable and reportable [2, 4]. Other approaches focus more on organizational aspects of accountability, namely how a new system could support accountability and audit practices [3, 21, 22]. How to improve the accountability of designers for the construction of a new socio-technical system is a core element of Suchman's proposal [18] on located accountabilities of designers in technology production and use.

Designing for the Internet of Things

The term "Internet of Things" describes a number of technologies which enable the Internet to reach out into the real world of physical objects [6]. Internet of Things technologies enrich objects with communication, sensing and computing capabilities (e.g. Radio Frequency Identification (RFID) chips), enabling new services, new business solutions and allowing to redesign business processes. Proponents claim that Internet of Things

applications will make the flow of products along the value chain more accurate, transparent and visible in real time, because information is automatically updated, media breaks eliminated and manual errors reduced. Transparency and an automatic update of information are appealing because they allow knowing the exact location and quantities of products without costly physical counts [7].

The increase in transparency and visibility also has substantial consequences for the users of an Internet of Things application. Users' actions might become accountable in new or changed ways because they become visible to different actors, such as a distant supervisor, or retrospectively through collected records.

Internet of Things applications are distributed applications, connecting different actors and IT systems into a new socio-technical system. Their interconnectedness with a myriad of devices and actors leads to an increased interdependence and complexity, which might reduce a user's freedom in enacting a technology [14]. A rigid prescription of the work process might conflict with his ability to cope with local contingencies at his workplace and to respond to local accountability relations [19].

Being still mostly a vision, today Internet of Things applications mainly exist in research laboratories and thus their real-world use cannot be directly studied and observed. Instead, researchers need to focus on prototypes or trials and use them as a "kind of experimental probe and tool for discovery" [20].

A method for a prospective analysis of the impact of an Internet of Things application on accountability relations needs to take these particularities of Internet of Things applications into account. The method needs to be usable for technologies that are not yet deployed. It should allow investigating the involved actors with an emphasis on a sequence of action and provide guidance for analyzing various pre-existing, newly created and multiple accountability relations.

NARRATIVE NETWORKS AND ACCOUNTABILITY RELATIONS

Narrative Networks

Pentland and Feldman [15] proposed narrative networks as a method for representing patterns of technology in use by visualizing and representing patterns of actions. A narrative network is a representation of an organizational routine, which is defined as "*a repetitive, recognizable pattern of interdependent actions, carried out by multiple actors*" [5]. Narrative networks not only describe existing routines, but they can also be used for designing or discussing how to design new organizational routines. Narrations about future organizational routines allow one to get a more complete description of a future socio-technical system than for example use cases, including all involved actors and accountability relations.

Narrative networks provide a conceptual vocabulary to describe and theorize the interconnectedness of actors, tools and tasks. A narrative network is a collection of functional events performed by actors or artifacts related by their sequential occurrence [16]. Functional events can include human actors and technological artifacts which can substitute each other. For example, in the case of an Internet of Things application, updating the data about the current status of an object can be done manually by a person writing the information into a database or automatically by a reader that interrogates an RFID tag from a distance.

The focus on a sequence of actions is useful because the success of a new ICT system depends on the establishment of a new organizational routine and not on the performance of a

single interaction. The sequence also allows identifying actors along the performance of a future organizational routine, which later can be used to analyze accountability relations. By using the narrative network approach we are able to collect stories of different participants about envisioned organizational routines, which include involved actors, technological systems and their relations.

When the stories include views of different groups such as future users, managers and system developers, the differences in points of view can be used to foster the understanding about possible difficulties.

Accountability Relations

The narrative network approach builds on organizational routines, structuration theory and actor-network theory, which all address accountability. In actor network theory, accountability is part of reporting procedures and giving a report towards others [11]. In structuration theory, being accountable for activities is to be able to give an explanation for them and supply normative grounds to justify them [8]. In the theory of organizational routines, accountability is understood in relation to an abstract, generalized understanding of an organizational routine. Accounting is done to explain performances and to legitimize them by retrospectively referring to the abstract and generalized understandings of routines [5]. A common element of most accountability concepts is that it is about a relation between at least two actors and about who is accountable towards whom. We refer to this as *accountability relation*. How an actor is held accountable and what for, however, differs between the different approaches (cf. [13] for an overview). To answer this question we distinguish between the following three different aspects of accountability. The first is *answerability* which is the ability to provide an account for performing an activity. This is mostly similar to the ethnomethodological understanding of being accountable and performing activities so they are intelligible for other actors. The second is *responsibility* focusing on what an actor is actually responsible for and how he fulfills his obligations. His responsibility might stem from formal or informal rules, compliance to standards or professional norms. The third is *liability* which is about facing consequences mainly towards external actors for not fulfilling formal or legal requirements. For these aspects both individual and organizational levels are considered; for instance, answerability is not only about an individual being able to give an account of his performance towards a supervisor, but also an organization being accountable towards a regulatory body. Accountability relations are enacted during the performance of an organizational routine. Therefore one needs to analyze the whole sequence of actions.

Owing to the networked and distributed nature of Internet of Things technologies and their integration into existing working environments, emphasis needs to be put on multiple and conflicting accountability relations. This is needed because one action or its record might be simultaneously used for different accountability purposes (e.g. accounting for a local and distant actor on different issues at the same time) and new accountability relations might interfere with existing ones.

INVESTIGATING ACCOUNTABILITY RELATIONS

We propose five steps to investigate accountability relations:

- 1) Select future organizational routine and points of view
- 2) Collect narrations to identify actors

- 3) Collect accountability relations
- 4) Identify critical accountability relations
- 5) Discuss critical accountability relations.

We describe and illustrate each step with a case study. We studied a publicly funded research project that developed a new Internet of Things technology for secure authentication of products in a supply chain. In four project trials, branded leather goods, luxury watches, airplane parts and pharmaceutical products were enriched with RFID tags or 2D barcodes. Supported by a Product Verification Infrastructure (PVI) that includes a local client and an online back-end system, test users could immediately identify if a product was genuine or non genuine by scanning the product. For a more detailed account and illustration of our method, we will focus on the pharmaceutical trial only.

Select Future Organizational Routine and Points of View

To be able to start collecting narrations and understand issues arising from accountability relations, one needs to focus on one specific future organizational routine. Our case study focused on two usage scenarios of the pharmaceutical trial: incoming goods scenario and the point of sale scenario. The incoming goods scenario consists out of the following functional events:

- Pharmaceutical employee scans the RFID tag or 2D barcode of a newly delivered package,
- System checks, whether the product is genuine or non genuine based on predefined rules,
- System displays result of the check, and
- Pharmaceutical employee adds genuine goods to the stock or moves non genuine in quarantine.

The point of sale scenario simply describes the sale to customers in the pharmacy. While the incoming goods scenario was performed by trial participants, the point of sale scenario was only discussed with them.

Different perspectives allow deriving a more holistic view and identify differences among groups of actors. The points of view of particular interest are those of: designers, describing the future systems; industry partners, describing the intended use inside the organization; employees, potentially using the system in the future. We collected points of view from two developers, two industry partners and ten trial participants working in the pharmacies. The level of detail varied among these groups of actors; while industry partners had a more distant view and talked more about high level actors, such as governmental agencies, trial participants focused more on those persons using the system.

Collect Narrations to Identify Actors

A real benefit of the narrative networks approach is that once an organizational routine is selected and the points of view defined, it is possible to combine different data collection methods, such as interviews, observation or documents, for gathering narrations. Documents containing a process description can be used as a starting point, followed by interviews revealing a more detailed description. Isolated interviews of one group only give a partial view because some connections between actors and performed activities of the studied organizational routine are opaque to a person narrating the sequence. The networked nature of the technology and the access of distant actors, which might even only be retrospectively involved, make this even more difficult.

In our case the narration of a trial participant only included the verification of a product with an RFID or 2D barcode reader without her being aware of the details of the check and the involved actors and artifacts (e.g. PVI, health authority, manufacturer etc.). We try to overcome the problem by collecting narrations from different persons with different knowledge about the organizational routine and the technology used. To investigate actors, we focus on the sequential order of the organizational routine asking what happens next and who is involved in each step.

We collected narrations through interviews. Developers and industry partners were asked to describe the envisioned process step by step and the actors involved. They were asked to draw the process steps, note down the involved actors and position them in relation to the process and each other. Trial participants were asked to describe their current work processes, their view on the fit of the tested prototype to the work environment and accountability focusing on responsibility. Table 1 shows the different elicited actors:

Table 1. Actors involved in the future organizational routine.

	Directly involved in use	Indirectly involved (setting of standards, rules and systems)
Individual	employee pharmacist customer / patient	developers
Organizational	manufacturer wholesaler delivery company up stream supply chain customs law enforcement federal drug testing lab "Counterfeiter"	standardization bodies industrial groups system producer gov. health agency gov. regulator pharmacists' association

We can use the list of elicited actors to identify additional persons to interview or study and iteratively reach a more complete view about the future organizational routine.

Collect Accountability Relations

The list of actors is used to ask about accountability relations between them. First, participants are asked about their understanding of accountability. Second, they are asked about the different aspects of accountability, such as answerability, responsibility and liability. This is done with each actor for each relation to another actor.

One can also include additional information about envisioned accountability relations by looking at project documents, such as trial description, trial presentation, use cases or written deliverables.

Identify Critical Accountability Relations

Once all investigated accountability relations are collected they can be analyzed. Two strategies can identify critical accountability relations: collecting of issues directly mentioned by the actors during the interview and comparison of the points of view of different actors on accountability. It is useful to proceed one actor at a time and collect all envisioned accountability relations containing this actor.

In our case study, two critical points were the use of the system at the point of sale and the responsibility for checking the

goods for genuineness. Pharmacists are responsible to hand out the right drug to the right customer. In accountability terms they are making their activities answerable by documenting all packages that they bought. This documenting activity might be controlled or used for investigation in the case of an issue by a governmental agency. The information, however, is only available on request and not intended to be shared. Beside very obvious issues (e.g. damaged packages), pharmacist do not regard themselves as responsible for the content of the product. They lack the capability to properly check the identification of products. Checks are made mainly for incoming goods and in case of an issue sent back to the wholesaler (e.g. wrong order, expiry date too close) or to a federal drug testing laboratory (e.g. obviously damaged or suspicious package). If pharmacists receive a package from the proper distribution channel which does not have an obvious issue, and give it to the right customer, their responsibility is fulfilled and they are not liable for the content of the product. The new system adds a verification check. The question is whether pharmacists would become accountable by using such a system. The pharmacist stated that they would probably not use it on a voluntary basis, especially if not everybody else is using it. The association of pharmacists would have to tell pharmacies to use it and pharmacies would then be responsible for using the system and performing checks. The industry partners' view was that pharmacists would have to take over the responsibility for verifying products with the system. Therefore industry partners assume a certain reluctance of pharmacists to use the system. According to the industry partners, the process is meant to become standard and pharmacists not using the system would be responsible for products that are not verified with the new system. The actual liability, however, would still need to be clarified by legislation.

Another issue was the use of the system in front of the customer. From a security point of view, all agreed that it would make sense to check the good at the point of sale. However, open and visible use of the system at the point of sale was regarded as problematic by pharmacists and pharmacy assistants because a check that is visible to the customer might interfere with the client relationship and invoke a trust issue for the pharmacy. Therefore they would rather prefer to use such a system for incoming goods or at least hide the use in front of the customer.

Discuss Results and Decide Further Steps with Project Participants

After having identified critical accountability relations the relations should be presented and discussed with the project participants to decide about further steps. Here the question is whether the elicited accountability relations are intended and if not, how conflicting accountability relations can be resolved. Solutions might not only impact the technological design but also the decisions of where to use a technology, what for, and towards whom users are accountable.

Preliminary findings on critical accountability results were presented to the project consortium and discussed during a project workshop. The discussion focused on the where the verification of goods should be done, namely at the point of sale or at incoming goods. The same issue was brought up also during the luxury goods trial in a retail shop. The findings were used to craft application guidelines and distributed to interested groups.

DISCUSSION AND CONCLUSION

Our novel approach allows a thorough and structured prospective analysis of possible conflicting accountability relations. Accountability relations are an important aspect of why the introduction of a new technical system might cause severe problems and therefore should be taken into account already during the early stages of development. The focus on the sequence of action to investigate involved actors is useful for networked, distributed and business process oriented ICT technologies because all important actors are made explicit and their roles can be analyzed. Our method elicits critical accountability relations and makes them visible so that they can be properly addressed. Besides having the critical accountability relations and the actors elicited, however, it is also important to use the findings to actually derive recommendations for the technology development project, implement them and iteratively reassess them for improvements. This might be difficult if a project has only one trial and accountability relations are not regarded as important enough.

We conclude that our novel approach combining narrative network and accountability relations is applicable to investigate possible issues on accountability relations and can already be applied in the early stage of the development of a new Internet of Things application.

ACKNOWLEDGMENTS

This work is funded by the an internal research grant of the ETH (TH-31 06-1) and by the European Commission within the Sixth Framework Program (2002-2006) project SToP (Stop Tampering of Products), Nr. IST-FP6-034144. We would like to thank Johann Weichbrodt for comments on an earlier version of the paper. We also thank the two pharmacies for their support.

REFERENCES

- [1] Baram, M. 2007. Liability and its influence on designing for product and process safety. *Safety Science*, 45(1–2), 11–30.
- [2] Crabtree, A. 2004. Taking technomethodology seriously: hybrid change in the ethnomethodology-design relationship. *European Journal of Information Systems*, 13(3), 195–209.
- [3] Dillard, J. 2008. Responding to expanding accountability regimes by re-presenting organizational context*. *International Journal of Accounting Information Systems*, 9(1), 21–42.
- [4] Erikson, S. 2002. Designing for accountability. In: *NordiCHI '02: Proceedings of the second Nordic Conference on Human-computer Interaction*. New York, NY, USA. ACM Press. Pp. 177–186.
- [5] Feldman, M. S. and Pentland, B. T. 2003. Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48(1), 94–118.
- [6] Fleisch, E. and Mattern, F. 2005. *Das Internet der Dinge: Ubiquitous Computing und RFID in der Praxis. Visionen, Technologien, Anwendungen, Handlungsanleitungen*. Springer-Berlin.
- [7] Fleisch, E. and Tellkamp, C. 2006. The business value of ubiquitous computing technologies. In: G. Roussos (Ed). *Ubiquitous and Pervasive Computing New Frontiers for Electronic Business*. Springer.

- [8] Giddens, A. 1986. *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press.
- [9] Grote, G. 2009. *Management of Uncertainty Theory and Application in the Design of Systems and Organizations*. Springer.
- [10] Kinder, K. 2009. *Ubiquitous Computing in Industrial Workplaces: Cultural Logics and Theming in Use Contexts*. PHD Thesis. Lancaster University. Lancaster.
- [11] Latour, B. 2005. *Reassembling the Social: An Introduction to Actor-network-theory* (Clarendon Lectures in Management Studies). Oxford University Press.
- [12] Lilley, S. 1996. Refining accountabilities: opening the black box of management systems success. In: Munro, R. and Mouritsen, J. (Ed). *Accountability: Power, Ethos & Technologies of Managing*. London Intl Thomson Business Press. Pp. 118–143.
- [13] Moes, J. et al. 2008. *Accountability-Beziehungen im Wandel – ein sozialwissenschaftlicher Zugang*. Discussion Paper SP III 2008-104. WZB. Berlin.
- [14] Orlikowski, W. J. 2000. Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, 11(4), 404–428.
- [15] Pentland, B. T. and Feldman, M. S. 2007. Narrative networks: Patterns of technology and organization. *Organization Science*, 18(5), 781–795.
- [16] Pentland, B. and Feldman, M. 2008. Designing routines: On the folly of designing artifacts, while hoping for patterns of action. *Information and Organization*, 18(4), 235–250.
- [17] Strathern, M. 2000. New accountabilities, Anthropological studies in audit, ethics and the academy. In: Strathern M. (Ed.). *Audit cultures. Anthropological studies in accountability, ethics and the academy*. London, Routledge. Pp. 1–18.
- [18] Suchman, L. 2002. Located accountabilities in technology production. *Scandinavian Journal of Information Systems*, 14(2), 91–105.
- [19] Suchman, L. 2006. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press.
- [20] Woods, D. and Dekker, S. 2000. Anticipating the effects of technological change: a new era of dynamics for human factors. *Theoretical Issues in Ergonomics Science*, 1(3), 272–282.
- [21] Winthereik, B. R., van der Ploeg, I. and Berg, M. 2007. The electronic patient record as a meaningful audit tool: accountability and autonomy in general practitioner work. *Science Technology Human Values*, 32(1), 6–25.
- [22] Yakei, E. 2001. The social construction of accountability: Radiologists and their recordkeeping. *The Information Society*, 17(4), 233–245.

What Is Real Work? The Case of Process Control

Alexandra Bidet
CNRS, Centre Maurice Halbwachs
ENS, 48 boulevard Jourdan
75014 Paris, France
alexandra.bidet@ens.fr

ABSTRACT

In this paper, we introduce the concept of *real work* in order to investigate what matters for workers. Sociological accounts usually focus on status, wages, and product, and equate *real work* with a good job. *Real work* has more to do with the way work is inhabited and evaluations are produced when coping with troubled or uncertain situations. Following the pragmatist John Dewey's notion of transaction, we join an ecological approach that stresses the human-environment system. The major issue is to understand, better than actor-network theory or activity theory do, the way in which operators find their way at work – and thereby, work as a meaningful practice.

Keywords

activity, real work, evaluation, transaction, process control

ACM Classification Keywords

J.4 [Social and Behavioral Sciences].

INTRODUCTION

“The workers patiently explained to him not the process but their process, their machine, their world. (...) What he constructs is the way in which they cognitively comprehend their work situation. (...) How do they quickly make their way, find the priority paths, in the environment that they recreate (...) It is the dynamic of the relationships that the worker maintains with his work environment which interest him, and their possible disruptions” [7].

Compared to sociology, ergonomics stressed the active and creative relationship of the worker to his environment very early on. Sociologists have misread the operative relationship to the world as much as ergonomists have explored it, as we see in the excerpt above from an homage to Jean-Michel Faverge, pioneer in the study of activities of supervision-control. Until the end of the 1980s, the operative relationship to the world was of little interest to sociologists, with the exception of Pierre Naville: “the study of industrial work conditions was first tied to effects (...) These harmful effects were recognized even before carefully studying technology and the human being's adaptation to these new mechanical beings

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ECCE 2009, September 30 – October 2, 2009, Helsinki, Finland.

Copyright 2009 ACM 978-1-60558-596-3/09/0009...\$5.00.

and to the workshops where they were implemented. It is not the *operation* which first drew attention, but the exhaustion which was a correlate thereof” [20]. Sociologists only studied involvement in work in terms of “social relationships”: resistance (hierarchical relationship), sociability (collegiality) and civility (work relationship). If we wish to find a description of, for example, the grasp of things and themselves that female workers construct in talking, we need to turn to the ergonomists. The same is true of the distinction between static and dynamic environment; the ergonomists were the first to examine – beyond the technical framework and objects *in* the action, the automatism: these systems liberated from the movement of human activity, the widespread use of which produces complex informational ecologies [19].

This excellent understanding of the work environment has, however, no equivalent in examining people. Ergonomics, a discipline with a practical purpose, has focused on forms of expertise and control, symbolic and intellectual functions: “understanding as workers understand” [18]. Jean-Michel Hoc deplored the fact that the “sub-symbolic activities (psychology of motivity, for example)” have been unexplored for so long [10]. He himself understood dynamic environments where human operators are no long the only ones to act in terms of “problem resolution”, “representation”, and “operative image.” The analysis depends on postulates of instrumental rationality, even in its “limited” version: the teleological character of the action, the autonomous individuality of the acting subject, his total mastery of his body [12]. In the field of supervision-control activities, Pierre Naville has, however, opened up lines of inquiry by noting the “worrying psychology” and the “need for meaning” of the operator faced with a task that he has ceased doing himself. In examining the obliteration of “creative curiosity about things” or, on the contrary, the appearance of a “relationship of information and communication of a new sort” between man and equipment, he suggested studying the way in which people deal with the question of the meaning of their activity. This assumes that people are no long separated from their environment, but that we should study the way in which they get their bearings in producing evaluations.

With the concept of *real work* (vrai boulot) [4], this paper proposes to further our understanding of people's involvement in work. The pragmatist inspiration, stressing the relationship of reciprocal constitution between the organism and his environment, allows us to confront the limitations of existing approaches. In actor-network theory, if the questioning of the “subject to inwardness” as the origin of the action allows us to bring people back into a “multitude which makes them act”, it is at the risk of losing them there. With “activity theories,” the reintegration of the creative and affective dimension of the action occurred only through the notion of subject, and thus remains dependent on the classic semantic of action: intentions, goal/means, etc. Finally, these limits have already been noted in HCI and CSCW. Exploring “*the variety of experiences that people engage in*” would also be to reintegrate the person: “Making the person – and particularly the emotional-volitional

character of the person that we recognize in desire, longing and joy – central, would radically challenge the rationalist assumptions of studies of people and technology in ways that HCI and CSCW may not be ready to do” [16].

The idea of *real work* can contribute to this. By *real work*, we mean the part of our activity which we are attached to and that we aspire to find in another job. The evaluations which are associated with it allow us to study the way in which people get their bearings in complex and uncertain environments [21], peculiar to the expanded ecology of contemporary work [3]. These evaluations mark the normative genesis by which each person tries to inhabit his or her work environment and to make it a meaningful practice.

CONTRIBUTIONS OF PRAGMATISM

Whereas the subjectivist phase of European philosophy has depicted an isolated and disengaged individual, the pragmatist perspective, on the contrary, asserts the primacy of the linkage of the organism and the environment. Considering that the organism exists as such only in active connections with its environment, is to maintain that man does not live only *in* an environment but *by* an environment [12]. Attention given to the “*state of affairs*” of the environment then surpasses the attribution of an agency to a distinct and constituted entity to consider the way in which we constantly work to make an environment in constant flux *our own*, in a constitutive relationship: “Agency is not an attribute but the ongoing reconfiguring of the world” [1]; “it resides neither in us nor in our artifacts but in our intra-actions” [24].

If we consider that intentionality emerges within the unit in movement formed by man and his environment, we then undermine the very principle of aporetic questions by the “subject” and “the environment” in the monitoring of the action [14], from which the proponents of “network-actor” thought they had already freed us: “a properly ecological approach is one that treats *the organism-in-its-environment* not as the compound of internal and external factors but *as one indivisible totality*. That totality is, in effect, a developmental system, and ecology deals with the dynamics of such systems” [11]. Here the pragmatist perspective intersects with French technical anthropology or classical sociology. By integrating the world of technical objects into the culture, this perspective comes back to the ontological primacy accorded to the constituted individual, and frees itself from a substantial conception of the individual – “*homo clausus*” to use Norbert Elias’s term – who sees action only through the “tragedy of choice”, as the simple communication of preexisting substances. To give priority to individualization over the constituted individual means also “to look for a sense of values other than in the limited inwardness of the individual who becomes withdrawn and denies the desires, tendencies or instincts which invite him to express himself or to *act outside of his limits*” [23].

The environment here is what the individual can react to – what his sensitivity notes and selects – by eliciting his “responses” [17] or a “*reflective conversation*” [22] which are authentically social. As Georges Herbert Mead has well illustrated, the “form of social coordination between agents and objects of the environment” consists of a “conversation of attitudes” in which we adopt the point of view of “physical things”, by attributing to them an inwardness and response abilities, in order to be able to interact with them, to control and define our own responses. Trade with “physical things” then contributes, just as trade with other “selves”, to ensuring that we become an object for ourselves, i.e. a “self”, capable of

adapting to an exterior point of view on itself and its acting [17, 25].

The ecological perspective, in the midst of a revival, has a rich history in sociology where it has made the pragmatist heritage flourish [9, 13, 5]. By granting a sociological dignity equal to that of the coordination between people to *our ways of entering into a relationship with our environment*, this perspective allows us to study the active and creative relationship to the environment. Refusing to make the “subject” the benchmark of the event or of its meaning is the equivalent of authorizing ourselves to describe the way in which we work to make it *our own*. The affirmation of a power to act is therefore inseparable from a “reticulation” of space and time in “select key points”: “Man finds himself linked to a world experienced as an environment (...) A reticulation of space and time is instituted which underlines places and select moments, as if all man’s power to act and all the world’s ability to influence man were concentrated in these places and in these moments. (...) These points and these moments localize and focus the attitude of the living being vis-à-vis his environment” [23].

In this framework, *interest taken in action* – “creative curiosity” towards things” –replaces the *satisfaction* of a personality, which only aspires to express itself or contemplate itself in the mastery of an object. Pierre Naville proceeds, as John Dewey, with a reassessment of the idea of interest, by prioritizing the point of view of “development of the person,” individualization over the individual. More specifically, John Dewey dismantles the “false concept of the relationship of interest and the self,” which considers the latter as “something that is fixed prior to the action,” “a fixed quantity and thus isolated.” The “development of active lines of interest” involves a correlative transformation of people and the world, associating a process of organization of attention to a “course of events in which we are *involved* and the result of which will affect us” [8, p. 168]. We come back to the etymology of the notion of interest: “what is *between*, what unites two things otherwise distant one from the other” [p. 160]. The idea designates less a state than a “career”, i.e. an effort at transformation in which we relate things to “a situation in continuous development”: “to be interested means to be absorbed, enthusiastic, carried away by an object. To take interest means to be on the alert, vigilant, attentive. We say of an interested person that he or she is both *lost in* and *in* a matter. The two terms express the absorption of the self in an object” [p. 159].

With *real work*, we are not interested in the global assessments – closely studied by sociology – that operators make of their work, but rather in the differentiated relationship that they maintain with the various items in their “bundle of tasks”. How can the evaluation of *real work*, as internal assessment of the activity, become a reference point, i.e. a source of memory, of commitment and professional requirements?

REAL WORK IN PROCESS CONTROL

Two Vocabularies for One Activity

In the context of a three-month observation *in situ*, we conducted in-depth interviews with all of the agents in the telephone traffic control center for Ile de France [Paris region] (n = 21). Organized around the account of the professional career and the concrete exercise of the control activity, the interviews were conducted at the p.c., in front of consoles and screens. How do these technicians describe their activity? Confronted with heterogeneous accounts, the lexical analysis allows us to introduce a systematic comparison, just as in that of their designs, conducted in addition [3]. A count was

conducted on a group of 38 notions or families of notions, assigning priority to those most used by the agents and/or significant with respect to the first analyses.

On the one hand, we have the agents whom we will call “explorers”: they narrate the technical adventures which associate them with automatisms and IT abstractions, and thus concentrate more than 75% of cases of “filte..”, “analysis”, “polluter”, “incident”, “satur...”, “transport...”, “efficiency”, “bundle”, “call”, “to ask”, “traffic”. The most characteristic notions of the accounts of the “explorers” are not those which are peculiar to traffic control, referring to the interface which particularly characterizes this “screen world.” They are related to the very flux of telephone traffic: “filte..”, “analysis”, “polluter”, “incident”, “satur...”, “transport...”, “efficiency”, “bundle” are therefore more distinguishing than “traffic” and “network”. Their accounts speak less of “network” and “traffic” than about the movement of automatisms. On the other hand, we have those whom we will call “troubleshooters” – they wait for breakdowns and keep the control activity at arm’s length, readily criticizing it for its abstraction. With more than two-thirds of the cases of “breakdown”, “organ”, “machine”, “cycle”, “job”, they invoke more often a world populated by technical objects with straight materiality, directly facing the troubleshooter or the programmer, and calling on his intervention. On the one hand, we have the *material* universe of care of machines; on the other, the *electronic* world of IT interfaces and artifacts.

Each of the two vocabularies illustrates a way to identify *real work*. For one group, the explicit reference to *real work* initially indicates a distanced and readily criticized relationship to the control activity which is too distant from “real technique”. Nevertheless, they also encounter *real work* in the present, when they reconfigure this “troubleshooting” activity locally: a part of the activity is in fact assigned value. The job is colored, although too rarely for their taste, by the delights of nostalgia, which we see in our interviews with mentions of the time when: “we were all in our machine, maybe it’s stupid, but we had... When you’re in your machine, it’s like in your car, you pay more attention because it’s yours; we knew the whole history of the machine, the site’s particularities....” The recurrence of the possessive indicates the comfort peculiar to the familiarity acquired over time with an environment which one has come to “inhabit” [6]. The nostalgic comments are thus not secondary, after-effects of a negative experience or rationalization of a difficulty. They first express the emotion peculiar to the familiar world. Therefore, it is an example of the value the “troubleshooters” assign to a part of the activity which, before being made rare by the transformation of switching techniques, was characteristic of their work.

The study of two vocabularies, through which they express their relationship to the activity, thus underlines a professional career effect. The two vocabularies are made up of words coined in distinct technical contexts, one corresponding to a prior period of telephonic switching; the other to current digital technologies. In both cases, the evaluation of certain aspects of work gives proof of our category of *real work*. The commitment to a part of the activity is seen by a series of evaluations dealing with the subtle detail of work in actions. We see the search in the activity itself for an “interesting job”, that is, *real work*. This evaluation leads one group, originally socialized in electrical-mechanical technologies, to keep traffic control at an arm’s length, or to reconfigure it locally, and the other group to be involved in it continuously.

The Real Technique

The “troubleshooters” systematically designate valued activities as “real technique”: *making* and *repairing*, *programming* and *troubleshooting*. What is common to these activities is that they appear in their eyes to require a body of knowledge, likely to guarantee the priority of the technician over the technical object: “That’s more technical, whereas here, the person who has no knowledge will be able to work anyway”; “My brother, he works in a GEC, they really do technical stuff”. The “things which are really worth it” remain for them more certainly “behind” the abstractions of the control operation.

The apparent ease of teleactions is deprecated as a form of passivity: “seeing everything” is “seeing nothing from A to Z” and when “the applications tell you almost everything”, “you have almost nothing left to do”. *Real work* includes more certainly the classical figure of the technical failure, “straight” and “tangible”, and of interventions which are then “repairs”. A “real alarm” is obvious in itself, with the univocity and urgency of the mechanical failure. An agent confirms this unambiguously: “I would rather repair the mistake than look for it for”; that would even be the antitheses: a breakdown that has to be “looked for”, a defect to “find with a scalpel.” The model is more like that of the tree which has fallen across the road: “You’ll be able to react quickly because you have a tree which falls over on the road, ok, so no problem, go ahead, you gotta move it and then it’s all done! But ok, see if there’s wind, lots of wind, if there’s a tree ...try to see if there isn’t a tree which is starting to get a little bit weak. There’ve already been three or four gusts of wind, you can see that it’s leaning a bit more ...pff, that’s not my shtick... it’s not tangible, if you like, or it’s not enough... it’s not my shtick”.

“You get rid of the tree and it’s *done!*”: this way of illustrating “real technique” excludes anything which would come under an active and continuous exploration of the network – road or telephonic. The statements “what’s worth it”, “what matters to me”, “my shtick”, “what I see”, etc., relate human intervention to a *useful mechanical effect*. The “troubleshooters” thus attribute value to a classical technical activity, i.e. subordinated to human work: a “sort of first-aid” based on a model where the artifact like its dysfunction is a product of an effective *cause*, entirely accessible to the technician.

This composition of the activity enables them to find *real work* in the control, by reconfiguring it into “troubleshooting”. Eschewing an indefinite, costly, and futile investigation, they thus display motor values related to the movement of the activity: “you filter the calls, then *done*, it works like new!”; “it takes you thirty seconds to filter a number, it’s *done*, and then, to my mind, the action was of some use”. The evaluation of “corrections” results from the same logic: “for me, what you have to do to give the job more value is *corrections*, remediations, find the trouble, find that there’s a problem which isn’t clear, and then make a *correction*, *that’s good.*”

The Real Work of Teleactions

The category of *real work* also puts a finger here on the evaluation of aspects of work involving movement – which may come as a surprise in such a mediated activity, where the body seems to have given way to technical automatisms. And yet the evaluation of teleactions – “nice”, “comfort”, “speed”, “interest”, “delight” – indeed occurs in reference to motivity: it points to an ease of movement. The activity is highly valued in *motor* terms, as an amplification “powers” of action: the exclamations “it’s powerful”, “it’s efficient” come from all sides, in praise of “power”. The action with the abstractions seems to be of an exceptional lightness, as seen in the “power”

of the spacing of calls: “it’s super powerful, super fast.” They value almost-instant interventions. Thus, when we appreciate being able “to go *faster* with this PCD”: “there won’t be any disconnection, that’s the important thing”; “the *power* of that thing is that *you start up 21 at once*, whereas even with PEC [switch management application], you have to open a window for each bundle, and there you start up everything at once, so *in one minute* you do all your spacings”.

The interviews with these agents thus show an *active involvement*, which also corresponds to continuous research. Traffic should be constantly *researched*: to “find what’s interesting in it”, “what’s happening of interest,” continuous work is necessary. Thus, the “explorers” are not familiar with the porous temporality of the “troubleshooters,” who look a bit condescendingly at these “enthusiasts, who are really into it and will find lots of things to do”: “he *likes to work in there*, he enjoys it, ok, good for him, but as far as I’m concerned it’s good because *he enjoys it*, but it’s something that’s *not really useful or effective*”; “I know some people, who go to look...I know some who go to look, in my opinion, who are going to get worked up for nothing, they *like that, huh...*”. The accounts of “explorers” readily delve into the dynamic of traffic flow, the cumulative effects of which are here the very site of *real work*. And so, with an enthusiasm tinged with concern, they unfurl the “apocalyptic” image of an automatism, no longer speaking to anyone but himself, or sinking into the “madness” of “snowball” effects.

CONCLUSION

Studying what is constructed in the activity – how a professional relates to his environment, assumes a situational and trans-situational focus. The requirements related to *real work* are formed in the course of careers and their moments of “happiness”, when we invent, or reinvent an agreement with our work activity.

When our investigations locate *real work* – “you have a *real goal*”, “you *really* have something to do” – the “truth” test, which makes these moments “true moments”, stresses the perception of interesting opportunities for action. No description here of salaried employees *carried away* by the frenzy of the activity: interest, active involvement in the job, does not emerge from an intense rhythm, but from an uncertain situation. Inversely, in *real work* it is not just a matter of a situation that we hope to *master*. Up against the image of the self as “something fixed prior to the action”, or of an author at the helm, John Dewey noted that “we refer to an interested person by saying that he or she is both *lost in* and *in* a matter. The two terms express the absorption of the self in an object” [8, p. 159]. If memory is attached particularly to often rare moments of *real work*, when action experiences its effects in an acute way– happiness or failure, comfort or complexity, performance or counter-performance – it is not only that we are leaving ourselves out of it, but also that we are creators of ourselves in real work.

REFERENCES

- [1] Barad, K. 2003. Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. *Signs: Journal of Women in Culture and Society* 28.
- [2] Bidet, A. 2007. La mise en récit du trafic téléphonique. Prendre prise sur un travail ‘immatériel’. *Raison Présente*, n°162, Nouvelles formes de travail et d’emploi.
- [3] Bidet, A. 2008. L’homme et l’automate. L’écologie élargie du travail contemporain. *Sociologie du travail*, 50(3), Travail et cognition II. (A. Borzeix, F. Cochoy (Eds.)).
- [4] Bidet, A. Qu’est-ce que le vrai boulot? Le cas d’un groupe de techniciens. *Sociétés contemporaines* (forthcoming).
- [5] Boutet, M. 2008. S’orienter dans les espaces sociaux en ligne. L’exemple d’un jeu. *Sociologie du travail*, 50(4).
- [6] Breviglieri, M. 2004. Habiter l’espace de travail. Perspectives sur la routine, *Histoire et sociétés* 9.
- [7] De Keyser, V. 1982. Les activités mentales dans le processus de production fortement informatisé. *Le Travail Humain* 2.
- [8] Dewey, J. 1975. Intérêt et discipline. In: J. Dewey, *Démocratie et éducation. Introduction à la philosophie de l’éducation*. Paris: Armand Colin (1919).
- [9] Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. New York: Houghton Mifflin.
- [10] Hoc, J.-M. 1996. Supervision et contrôle de processus. La cognition en situation dynamique. Grenoble: PUG.
- [11] Ingold, T. 2001. *The Perception of Environment: Essays in Livelihood, Dwelling and Skill*. London: Routledge.
- [12] Joas, H. 1996. *The Creativity of Action*. Polity Press.
- [13] Joseph, I. 2007. L’athlète moral et l’enquêteur modeste (textes rassemblés par D. Cefai). Paris: Economica.
- [14] Kaptelinin, V. and Nardi, B. A. 2006. *Acting with Technology. Activity Theory and Interaction Design*. Cambridge: The MIT Press.
- [15] Knorr-Cetina, K. 1997. Sociality with Objects: Social Relations in Post-Social Knowledge Societies. *Theory, Culture & Society*, 14(4).
- [16] McCarthy, J. and Wright, P. 2004. *Technology as Experience*. Cambridge/London: The MIT Press.
- [17] Mead, G. H. 1934. *Mind, Self and Society*. Chicago UPress.
- [18] Montmollin (de), M. 1986. *L’intelligence de la tâche. Eléments d’ergonomie cognitive*. Berne: Peter Lang.
- [19] Nardi, B. A. and O’Day V. L. 1999. *Information Ecologies. Using Technology with Heart*. Cambridge: MIT Press.
- [20] Naville, P. 1963. Vers l’automatisme social? Problèmes du travail et de l’automation. Paris: Gallimard.
- [21] Norros, L. 2004. *Acting Under Uncertainty. The Core-Task Analysis in Ecological Study of Work*. Espoo: VTT.
- [22] Schön, D. A. 1983. *The Reflective Practitioner. How Professionals Think in Action*. New York: Basic Books.
- [23] Simondon, G. 1989. *L’individuation psychique et collective*. Paris: Aubier.
- [24] Suchman, L. 2007. *Human-Machine Reconfigurations*. Cambridge: Cambridge University Press.
- [25] Turkle, S. 1984. *The Second Self: Computers and the Human Spirit*. New York: Simon & Schuster.

Doctoral Consortium

Designing Sustainable Tools for Living with the Aid of Mobile Technology in the Emerging Economy Contexts

Karthikeya Acharya

Doctoral Researcher

School of Design / University of Art and Design Helsinki

Hämeentie 135 C / FIN-00560 Helsinki

karthikeya.acharya@taik.fi

+358466367296

Ph.D supervisor: Ilpo Koskinen

ABSTRACT

This research thesis deals with understanding sustainable social practices of specific Indian communities which can inform the design and testing of an integrated mobile system. The system intends to connect user experience and events with user consumption patterns, within the emerging economy context of middle income Indian families. This case based design research thesis proposes to address the idea of creating a conscious society, which would feel accountable for the resources it consumes with the aid of mobile technology. The research process if carried out as planned would also provide rich input to the field of social science research and design.

Author Keywords

consumption patterns, sustainability, Indian middle class consumers, ambient intelligent environments, persuasive mobile technology, mobile user interface, user culture

ACM Classification Keywords

H.1.1 Systems and Information Theory, General systems theory; H.5.2 User Interfaces, User-centered design; H.5.3 Group and Organization Interfaces, Theory and models; J.5 Arts and Humanities, Architecture.

PROBLEM STATEMENT

The globalizing world pumped with the communication revolution, is not just evening the gap of economic disparity between developed, developing and underdeveloped nations but is also conveying an even model of consumption for users all over the world [1]. After creating and producing machines for living which utilize the earth's resources, we humans are now seeing the need for accountability for our consumption [2]. This has created scope for a self monitoring everyday tool which can measure and keep track of usage patterns wherein a user would know her pattern of consumption of resources over a period of time. Mobile communication technologies, with their ubiquitous presence and the deep reach within today's

society has the potential to harbor this tool. This idea of creating a conscious society is the need of the hour now, in the highly populous, resource absorbing developing world. Such a tool would ensure accountability, of use and consumption from the base of usage itself. Ubiquitous computing with persuasive mobile communication technology platforms [3] could be a relevant modality to address this issue. This study and design research is about devising one such tool with mobile technology as a platform, for the culturally diverse Indian middle class consumers, who are fast reaching intense consumption levels. This might prove to be an amicable way towards a sustainable way of living in the emerging economy contexts. India makes up for about 17% of the world's population. Of the 1.13 billion people of this developing nation, about 50 million is the current estimate of middle class consumers. If the current economic reforms continue, a predicted economic growth rate of 7.3% is estimated. This figure of growth would enable a 10 fold surge in the middle class populace to about 583 million in the next 20 years [4]. With such dramatic shifts in spending patterns the straining demand on infrastructure and natural resources is but imperative. Such growth would but increase industrial demand internally again putting pressure on global natural resources.

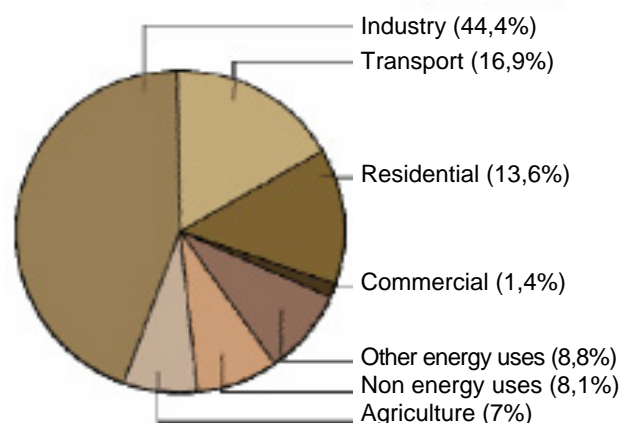


Figure 1. Final energy consumption in India, by sector, in 2005/06.

As shown in Figure 1, the 3 major energy intensive sectors are the industry, transport and the residential sectors [5]. We can assume that the demands put on industry and transport is largely coming from the internal populace which is housed in the residential sector. When viewed as an integrated whole these sectors are responsible for about 3/4th of the entire energy requirement of the nation. This study intends to look

closely into consumption habits of units of the society, i.e. the family. Hence the focus would be the domestic and the transport realm. Households use energy for many purposes: cooking, cooling and heating their homes, heating water and for operating many appliances such as refrigerators, stoves and televisions. At the national level, kerosene and electricity constitute the primary fuel for lighting in 99% of the households. As mentioned in the National Energy Map for India: Technology Vision 2030, the micro-perspective of each consumer is the driving force behind the sector's use of energy, and opportunities for change in the demand and supply patterns. The grounding for this study is based on this presumption, that modifying the micro perspective can change the demand and supply pattern, based on the interactive perspective [6]. The attempt here primarily is to reduce the consumption by designing appropriate design tools for daily living.

RESEARCH QUESTIONS

- Does being conscious of one's consumption pattern on a periodic basis effect or change the decision making process of the user's future consumption? If so, how?
- How cultural background and upbringing effect consumption within the culturally diverse landscape of middle class India?
- How are user consumption and personal identity related?
- User's culture, consumption and space, what can mobile technology offer for tomorrow's world?

APPROACH

While moving ahead wanting to brandish poverty and illiteracy, India seems to be donning many a identities, that of a back office servicing developed nations, an emerging market, future super power and so on. India's identity for its internal populace as a knowledge society with strong cultural roots of the past is an understanding which is reserved only for the cognoscenti, while the large mass struggles for daily living. Such a paradox is common place as disparity lurks around every corner. Another irony is such that while the attempt is on to bridge the disparity, this would lead to further rapid consumption by the masses which nobody can promise to fulfill. The so called Indian way of life, in the form of Hinduism which supposedly is meant to integrate spirituality, commerce, sex and even nature into daily living, maybe did not take into account globalization and western formats of consumption. The texts of such a belief system the Vedanta, containing the Vedas has no single author, but is an amalgamation of concepts over a long period of time. It is known that Hinduism was not created by a single person but seems to have evolved over time. This evolution of such a way of life or the belief system was never really proselytizing unlike other faiths, but had its own mechanism to shield itself. This mechanism can also be viewed as the growth engine for the belief system, and is well known as the caste system. This socio-hierarchical concept which catered then to providing identity to individuals based on their professions also resulted in creating discrimination which can be speculated to the rise in disparity in such a culture. In spite of such shortcomings the traditional way of life still persists and people follow them, religiously, literally. The belief system has evolved and grown and somewhere it does seem to have proven successful, that it has lasted. Where can we spot this success? The Mathas of South Canara which are nodes of social administration run by the upper caste have had large land holdings for the past 800

years. The Matha housed a deity and hence enjoyed immense clout amongst the local populace. It insisted on a ritual, that fresh milk be poured onto the deity everyday, which is called the 'abhisheka'. The milk had to be sourced from cows only, as cows are considered sacred creatures. Hence the Matha now owned cows which gave milk for the ritual. The cows need to graze and hence hay was essential. With ample land holdings and suitable climate, paddy cultivation would work well. That meant involving more people for farming as labor, who got paid in a dual format of rice and money. Such was the socio economic system, that one ritual started a local sustainable community. Are there lessons to be learnt? What esoteric knowledge is augmented in daily Indian living? What systems hold promise that we can integrate with mobile technology which may solve our current problems? The attempt here is to unearth local knowledge which point towards sustainable systems from the existing culture. The intention is to critically analyze the logic of such beliefs as cultural analysis and deconstruct the rituals and social practices to a theoretical framework of sustainable practices of a culture. The study doesn't intend to be theosophical at all. To understand how communities relate to their surroundings and what practices and social activities does it perform to sustain and maintain a balance with the environment is the intention. The systems have to be viewed in a paradigm, of time and context maintaining a strict focus on practices of sustainability, by Indian communities.

The following communities have been identified as possible field case studies for this research:

- 1) The Coffee growing labour communities of Coorg
- 2) The Mattur Sanskrit village
- 3) The Ashta Mathas of South Canara
- 4) The Kudubi tribes of South Canara
- 5) The Boat makers of Koditaleygrama.

This is a tentative list and the exact context identification needs further refinement.

The framework to study such systems is described in the coming section of the hypothetical framework.

HYPOTHESIS

The study seeks mainly to understand user consumption on a day to day basis. The approach intends to utilize the design science framework as a model [7]. This gives rise to a need where the research would involve a combination of social and empirical research. Hence the research is two fold, involving ethnographic research of sustainable practices of existing cultures set in the Indian context and then looking at methods and modalities, empirical in nature for tracking the consumption patterns of middle class users on a periodic basis. One of the main tangents the research seeks to look into is the aspect of consumption pattern and user identity thus trying to understand how a user's culture would determine her consumption pattern. The current hypothesis is that the following parameters would play a crucial part in forming the elements of user consumption.

- 1) The user's physical environment – Ambient temperature with humidity and light conditions.[8]
- 2) User's energy consumption – Direct and Indirect energy consumed by the user.

- 3) Water consumption – User’s water consumption – potable, for hygiene and for user infrastructure maintenance.
- 4) Commodity – The day to day food, clothing and shelter seen as artifacts which are consumed by the user. [9]
- 5) Mobility and movement – Tracking user distances and destinations as patterns of everyday movement. [10], [11]
- 6) ICT usage – Maps of the user’s daily communication and information patterns.

Physical parameters	Social practices/rituals
Environment	
Energy Consumption	
Water Consumption	
Commodity	
Mobility	
ICT Usage	

Phase 1 Theory and Methodology – January 2009 – September 2009, at the University

The initial period of the phase plans to determine methods to study the practices of sustainable living of a community as mentioned in the approach. Then the hypothetical framework of the above mentioned physical parameters will be scientifically analyzed to determine the design of the research tools. The parameters are termed as physical for they entail a very ‘spatial, temporal, infrastructural and environmental’ centric approach towards the user and the context. If this data termed as physical parameters, could be meaningfully put together one maybe could arrive at spatial requirements for any user, hence this data also holds potential in generating a user controlled environment [12]. Thus this part of the study deals with designing the research with a theoretical framework to study cultural practices of sustainability and also how to capture meaningful user data from the above stated 6 hypothetical physical parameters.

Phase 2 Initial Field Studies – October 2009 – April 2010, on field

The next part of the research is about understanding and approaching a community with a culture and cognitive lens paradigm, to determine the relation they establish with their surrounding and environment. To observe and understand what daily practices they perform and what aspects can be extrapolated as sustainable practices. What relation do these practices have with the 6 hypothetical parameters? This would be the first part of the actual field study, where an Indian community/communities would be studied in detail for its daily practices. The framework intends to be viewed with the lens of sustainable practices as mentioned earlier. This would also help in understanding consumption with a cultural framework as a paradigm. The above mentioned hypothetical parameters would be tested through observation and systematic recordings. To evolve a method, such that the research notes and records the contextual information surrounding the social units (which may effect and determine their way of living) over that period of time will be a research task which would have to be addressed in the initial method phase.

Expected outcome – First level of understanding consumption in Indian middle class contexts, case studies of sustainable living practices of X culturally varied social units in the middle class Indian context.

Phase 3 Design prototype – April 2010 – September 2010, at the Lab/University

The next phase would involve analysis of data gathered from phase 2, testing the hypothesis of the 6 physical parameters and arriving at design guidelines, to come up with an interface/device/system which would record and collate the essential physical parameters and generate a user experience of understanding ones personal consumption. The main task during this phase would be the design and production of such a system. The focused task would involve in coming up with a refined prototype for such a system.

Expected outcome – *Production of an Artifact which maps user consumption on a periodic basis.*

Phase 4 Prototype field test and use – October 2010 – April 2011, on field

This phase would involve firstly identifying 3 sets of families and recording their consumption according to the refined set of physical parameters over a fixed period of time. Then the users and their environments from the 3 families would be introduced to the design system which we are assuming would record necessary data as the physical parameters. Notable changes in the actual user context would be recorded, which would need to be taken into consideration during the analysis. The consumption patterns based on the physical parameters of the user, after introducing the prototype of the mobile interface/device/system, would be recorded for the same period of the time as during the initial phase of this field study.

Expected outcome – *Artifact testing in the actual context and deeper understanding of consumption pattern with the aid of the artifact in Indian middle class society.*

Phase 5 Data Analysis and publication – April 2011 – January 2012, at the University

The data gathered from phase 2 and 4 would be analyzed and collated, and a comparative research paper would be published stating the findings of the research. The paper mainly wishes to address three main issues, the first being; what are the social practices of specific communities which lead to sustainable living? The second is, if being conscious of ones consumption pattern, would it effect or change the decision making process of a user’s future consumption, if so then how? The third issue would be to compare how cultural background and consumption patterns are connected within the context of middle class Indian families.

Expected outcome – *Design Artifact and Thesis.*

RELATED WORK

Since the research is in the preliminary proposal stage it is currently looking for a strong theoretical framework to base the research. Looking at vernacular Indian ways of living which had mechanisms of controlled sustainability embedded into daily life with an anthropological perspective is being considered as an approach. The study aims at combining current modalities which technology affords, and on the other hand closely follow Indian ways of daily sustainable living. Prior work and research on the technology front have influenced this study. The following mentioned researchers and their works have influenced this study and this proposal:

- 1) B. J. Fogg's work on Persuasive Technologies [13] and the argument that computers can elicit constructive user behaviour is seen as an important grounding for the hypothesis of this study. This is so because the aim is to make the user understand about daily consumption with the aid of mobile technology and influence future user consumption.
- 2) T. Koskela and K. Väänänen-Vainio-Mattila's [14] work on User interfaces for smart homes from the Tampere University of Technology: The empirical approach towards understanding user interfaces and especially the methodology of research has had an impact on this study. The approach of using the mobile phone for instant control in the mentioned study is seen as an asserting point. But in this proposal on the contrary it is suggested to use the mobile phone as a primary feedback tool.
- 3) Energy use in eight rural communities in India [15] by B. Bowonder, N. Prakash Rao, B. Dasgupta and S. S. R. Prasad which elicits how energy consumption in rural India depends on socioeconomic and agro climatic factors is another study which is seen as an inspiration for this proposal. This empirical communities study however was carried out more than 20 years ago and hence might not be valid in today's context.

CONCLUSIONS AND NEXT STEPS

The study as mentioned is still stands as a proposal and hence is just a seed which needs to be nurtured. Hence it is looking forward to further support from the academia and the industry. The need for technical assistance from an industry leader like Nokia with a sound theoretical guidance from the academia like TaiK is seen as the next level to this study. At this stage the edges are rough but with expert guidance and support the research might actually cater to the much required ways and means of creating a sustainable culture for life.

REFERENCES

- [1] Lahiri, S., Babiker, M. and Eckaus, R. S. The Effects of Changing Consumption Patterns on the Costs of Emission Restrictions, MIT Joint Program on the Science and Policy of Global Change, 2000.

- [2] Wiedmann, T. and Minx, J. Definition of carbon footprint. ISAUK Research Report, 2007.
- [3] Aleahmad, T., Fussell, S. R., Balakrishnan, A. D., Kiesler, S. and Wong, J. Fishing for Sustainability: The Effects of Indirect and Direct Persuasion, CHI 2008.
- [4] The McKinsey quarterly, Tracking the growth of India's middle class 2032.
- [5] TERI, Indian Energy Sector: an overview, Energy Data Directory and Yearbook 2007.
- [6] History and Theory in Anthropology by Alan Barnard.
- [7] Hevner, A. R., March, S. T., Park, J. and Ram, S. Design Science in Information Systems Research, MIS Quarterly.
- [8] Sathish, S. Using Declarative Models in Multi-device Smart Space Environments, Nokia, 2005.
- [9] Gangopadhyay, S. and Wadhwa, W. Changing Pattern of Household consumption Expenditure, (SERFA) August, 2004.
- [10] Ivanov, Y., Sorokin, A., Wren, C. and Kaur, I. Tracking People in Mixed Modality Systems, MERL, TR2007-011 February 2007.
- [11] Rekimoto, J., Miyaki, T. and Ishizawa, T. LifeTag:WiFi-based Continuous Location Logging for Life Pattern Analysis.
- [12] Haque, U. Arquitetura & Urbanismo, AU 149 Brazil, August 2006.
- [13] Fogg, B. J. and Nass, C. How Users Reciprocate to Computers: An experiment that demonstrates behavior change, CHI 97.
- [14] Koskela, T. and Väänänen-Vainio-Mattila, K. Evolution towards smart home environments: empirical evaluation of three user interfaces, Pers Ubiquit Comput. 2004.
- [15] Bowonder, B., Rao, N. P., Dasgupta, B. and Prasad, S. S. R. Energy Use in Eight Rural Communities in India, Administrative Staff College of India, Hyderabad, World Development, 13(12).

Culture, Notation and Formalisation

Laura Benvenuti
 Open Universiteit Nederland
 Valkenburgerweg 177
 6419 AT Heerlen
 The Netherlands
 laura.benvenuti@ou.nl

Gerrit C. van der Veer
 Open Universiteit Nederland
 Valkenburgerweg 177
 6419 AT Heerlen
 The Netherlands
 gvv@ou.nl

ABSTRACT

What we perceive is largely determined by our mind-set. Our personal history colors our perception; this process is called “interpretation”. Interpretation does not just occur at the individual level, but also in collectivity. Our cultural background provides us with a framework for interpretation. In this paper we question to which extent the way we write programs, formalize or model reality is determined by culture.

Keywords

culture, formal languages, notation

CULTURE AND COMMUNICATION

A definition of “Culture” is: *all the knowledge and values shared by a society* (Wordreference, 2008). The culture we were raised in determines the way we think, feel and act. But culture is not the only factor. Hofstede and Hofstede (2004) discern three levels of uniqueness in what they call our mental programming: human nature, culture and personality. Human nature is the universal level, the level we share with other human beings because we are human beings. Human nature is innate. Personality is the level in which we distinguish ourselves from other individuals. Our personality is influenced by our set of genes, which is almost unique, by our education and by our experiences. Personality is partially innate (the set of genes), partially learned (our education) and partially accidental. Culture is the intermediate level, the level where education takes place. Culture is learned. This is the level we share with a group or category: *the collective programming of the mind that distinguishes the members of one group or category of people from others.* (Hofstede and Hofstede, 2004)

Hofstede and Hofstede explain different levels of communication problems in organizations by explicating cultural differences.

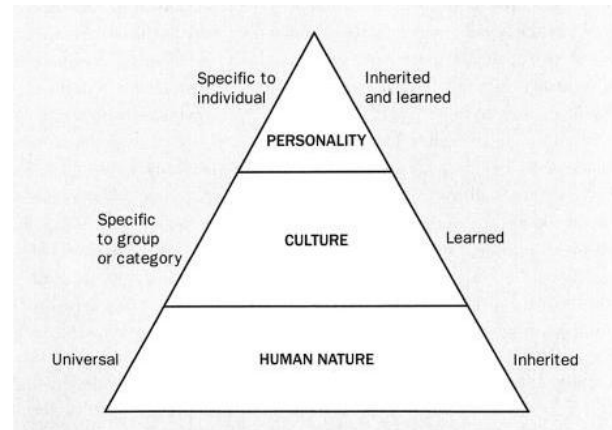


Figure 1. Three levels of uniqueness in mental programming.

FORMAL SYSTEMS

Computers do exactly what programmers tell them to do, nothing more, nothing less. This is not so straightforward: computer programmers are human, and human beings are not used to always communicate explicitly. We are used to fill “gaps” in communication. In the early days of software engineering, ambiguous specification of software led to serious problems. That’s why computer science has developed tools and methods to avoid ambiguity: ambiguity in the communication between software engineers but also between programmers and the providers of their software tools.

Unambiguity is obtained by defining formal settings in which communication takes place. Examples of formal settings are relational algebra, which supports the communication between database administrators and suppliers of relational database systems; and the Unified Modeling Language, which supports the communication between programmers of Object Oriented software.

Software Engineers use formalization as an expedient to control communication. During the engineering process they can minimize the risks of misinterpretation because they agree on the semantics of their statements. When the system is operational, it can communicate with other systems through formal statements, whose meaning is given, immutable and shared by a community of programmers.

Those formal systems are not interchangeable. Each one has his field of application. The choice of the formal setting determines the strengths and the weaknesses of the resulting system. If the software engineer chooses a relational database to store data, the system will gain flexibility in the retrieval of data but will have difficulties with the manipulation of

retrieved data; if he chooses an Object Oriented solution the advantages and the disadvantages will be reversed. Computer scientists mostly do not mention the drawbacks of their choice for a language or formalism. Sometimes they do not even consider those drawbacks.

Occasionally, we can trace the origin of differences between formal systems. Relational Algebra and the Object Oriented paradigm implement different concepts of “meaning”. Database engineers postulate that objects are identified by their properties; Object Oriented programmers identify objects by reference. Those groups disagree on basic philosophical issues. Database engineers follow Frege’s (1948) and Russell’s (1905) positions in the debate on the meaning of proper names; Object Oriented Programmers agree with Kripke (1972) in rejecting a descriptive theory of proper names. In Object Oriented Programming, the content of an object is established by an initial baptism. Engineers are not necessarily aware of the philosophical backgrounds of the formalism they use. The philosophical position is implicit; it is a consequence of the choice for a software environment.

Sharing backgrounds facilitates communication between engineers who work with similar software, but communication with other engineering cultures suffers.

MATHEMATICS

An obvious question to ask is whether we can find formal systems that are universal, i.e., free of cultural values.

It is often assumed that mathematics reasoning or formalisms are free of limits imposed by culture: all arithmetics or geometries are equivalent, etc. To question this assumption, we are investigating moments in history where disruptions in (the use of) scientific notations occurred; we both looked for disruptions in cultural values and in disruptions in the development of notations and mathematical theories. Our aim was to find coincidences: significant changes in one field that preceded (and caused?) significant changes in the other field.

So far we focused on the development of arithmetic in the 12th-13th century. In the Middle Ages, Christian Europe used Roman numerals. Arab mathematicians developed arithmetic and calculus for merchants and used the positional system that included a symbol to represent the zero (Maracchia, 2005). Their notation made calculus easier than it was in the Christian world. In those days, every city state had her own sizes, weights and monetary system, Italian merchants needed a formalism or a notation who would support them in their unceasing calculations.

Leonardo Fibonacci introduced Arab numerals in Christian Europe with his *Liber Abaci* in 1202. In 1241 the city of Pisa employed him as “consultant in arithmetic”. In the following centuries other municipalities in Tuscany opened school where this new calculus was taught (*Scuola d’Abaco*); Arab calculus spread first in Tuscany and subsequently in Europe (Giusti, 2002).

The introduction of the *Liber Abaci* is considered a turning point in the development of European arithmetic. This is a clear disruption in the history of mathematics in Europe. This disruption, in turn, preceded a booming era of Italian commerce and the start of the Renaissance.

Cultural revolutions can be triggered by technical developments. This is happening now with the introduction of ICT in developing countries. The other implication, a disruption in the development of arithmetic or formalisms preceded by a turn in cultural values is less obvious, maybe because cultural forces often defeat the publication and spread of scientific opinions that are inconsistent with ruling values. Look at the position of Darwinism in the present: there are orthodox religious cultures who do not appreciate this theory to be mentioned.

One of the turns in moral values we can easily trace is the attitude of the Catholic Church towards the financial system. “Financial system” is the translation of the Medieval terms “bank loans” and “usury”. In the Middle Ages, it was considered a sin to make money without working. Usury was banned by the 3rd Council of the Lateran in 1179 (Usury); this interdict remained in force until Pope Leo X’s bull “*Inter Multiplices*” in 1515. Leo X wrote that it was not forbidden to take collaterals on loans if it was necessary to cover the costs, although he confirmed that taking excessive interests was sinful (Capone, 2002). He was the first pope to admit banking practices (pawn shops or lombards).

Even though Fibonacci had solved a problem raised by interest in the *Liber Abaci*, a general treatment of this problem requires more advanced algebra (Giusti, 2002, pp. 103–106). Medieval maestri d’abaco were aware of the shortcomings of their calculus. But arithmetic did not develop in Europe until the 16th century, when Cardano wrote the *Ars Magna*, the first treatise on Algebra in Europe. Algebra developed in Europe since then. (Maracchia, 2005)(School..., 2008).

We summarize this development relations in Figure 2.

In the upper line of Figure 2, we represent the history of algebra in the Arab world. The middle line contains landmarks of the history of algebra in Christian Europe. The lowest line reflects disruptions in moral and cultural values in the European Christian world.

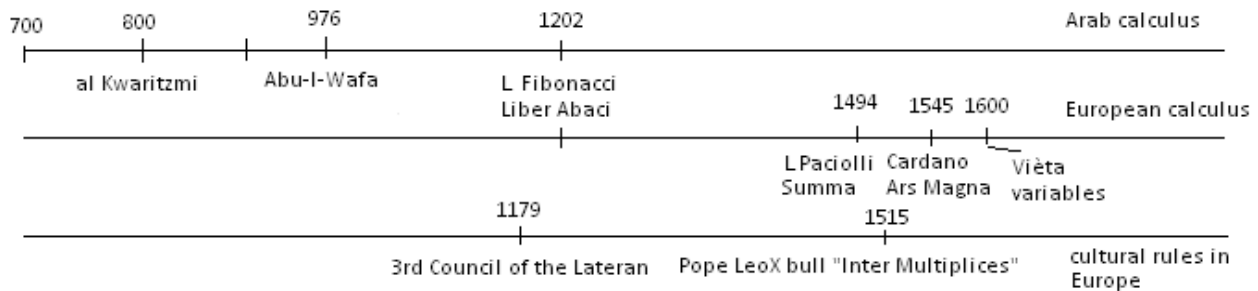


Figure 2. Development disruptions and relations. Axis A represents the development of Algebra in Arab Muslim culture, axis B shows the development of arithmetic and algebra in Europe, axis C indicates the official cultural rules developed in catholic Europe.

after Fibonacci only began three centuries after the Liber Abaci, almost synchronous with a Pope's new verdict, and that epoch making results were obtained in 1545.

We can not say that the changes in moral values have caused the revolution in mathematics, but we can state that we have found an interesting coincidence between the development of culture and the development of a formalism who is supposed to be independent of culture.

Our research project aims at identifying similar instances of co-development between culture and science (in the sense of systematic developments of notation, formal languages, and reasoning). Currently we are considering several episodes: the relation between European and Chinese mathematics and physics between the early 17th and the 18th/ century; South American pre-Columbian development of mercantile rope-knot notation in relation to inter-tribal or "international" trade

Our ultimate goal is to acquire indications to establish to what extent the process of formalisation is dependent of culture. This will help us understand the current development of languages people use in working with computers, both for hard-core programming, for using operating systems, and for using embedded IT applications. We observe the co-development and wars of cultures and notations in this field (Timmerer, 2009): the rise and fall of popular operating systems (MS-dos vs the Mac- ref the "anti-mac" pannel at CHI about 10 years ago), the fierce debates between Linux believers and WIMPS adapts. But also Edsger Dijkstra's strong opinions about what is good programming and good education (visualisation leads to curriculum infantilisation, Dijkstra, 1988) versus Ben Shneiderman's opposite view (Shneiderman, 1983).

REFERENCES

Capone, D. 2002. Profili dell'usura e della polemica antiebraica nel Rinascimento, tesi di laurea
<http://www.tesionline.it/default/tesi.asp?id=8533>, retrieved 24.1.2009.

Dijkstra, E. W. 1988. On the cruelty of really teaching computing science, transcription by Javier Smaldone, retrieved 16.4.2009 from <http://www.cs.utexas.edu/users/EWD/transcriptions/EWD10xx/EWD1036.html>.

Giusti, E. (Ed.). 2002. Un ponte sul mediterraneo. Leonardo Pisano, la scienza araba e la rinascita della metematica in Occidente, edizioni Polistampa, Firenze.

Frege, G. 1948. On sense and Reference, translated by M. Black, first appeared in *The Philosophical review*, 57, 207–230.

Kripke, S. A. 1972. Lecture II of Naming and Necessity, From *Readings in the Philosophy of Science*, P. Ludlow (Ed.). MIT press 1997.

Hofstede, G. and Hofstede G. J. 2005. *Culture and Organisations: software of the mind*: McGraw-Hill.

Maracchia, S. 2005. *Storia dell'algebra*, Liguori editore, Napoli.

Russell, B. *Descriptions*, first appeared in *Introduction in Mathematical Philosophy*, George Allen and Unwin, London 1919.

School of mathematics and Statistics, University of St Andrews 2008. Quadratic, cubic and quartic equations, <http://www-history.mcs.st-andrews.ac.uk/Indexes/Algebra.html>. Retrieved 27.1.2009.

Shneiderman, B. 1983. *Direct Manipulation: A Step Beyond Programming Languages*, in: *Computer*, august 1983.

Timmerer, Christian (interviewer) 15.4.2009. A video interview with Nikolaus Wirth part 2, in *Computing Now*, retrieved 16.4.2009 from <http://www2.computer.org/portal/web/computingnow/videos/niklauswirth>.

Usury, retrieved 27.1.2009. From <http://www.catholic.org/encyclopedia/>, retrieved 5.4.2009 from <http://en.wikipedia.org/wiki/Usury>.

Wordreference.com 2008. Retrieved 5/4/2009. From <http://www.wordreference.com/definition/culture>.

Friendship Online: Users' Experiences of Relationships in Social Networks

Rob Comber

People and Technology Research Group
Department of Applied Psychology
University College Cork, Ireland
r.comber@ucc.ie

ABSTRACT

This research examines user's experiences of one-to-one relationships set within the boundaries of social media websites, such as online communities and social network sites (SNSs). Using ethnographic methods a rich description of life in two websites (Deviant Art and Bebo) is created. This description serves as a building block for an analysis of the roles of technology in mediating relationships and also the roles of those relationships in technology adoption and appropriation.

This research looks at how user experience unfolds over time particularly with respect to other people and the technology that mediates relationships with them. The changes that were observed in the two communities relate to three specific connecting experiences and an unfolding of the apparent 'function' of the sites and their position in the user's everyday life. The unfolding phases also change the user's experiences of control and their emotional response to the sites and other users.

Unraveling the contexts of interaction highlights the extent to which varying social media services, whether interest-based communities, or social network sites, are embedded in our everyday lives and the extent to which they can and do exist as useful communication spaces.

Keywords

social networks, community, friendship, online ethnography

ACM Classification Keywords

H.5.3 [Information Interfaces and Presentation (e.g., HCI)] Group and Organization Interfaces – *Web-based interaction*.

INTRODUCTION

In 1996, Parks and Floyd [1] set out to "examine the relational world actually being created through Internet discussion groups." In the intervening 13 years, momentous changes have occurred in the geography and culture of the Internet, social media and technology. It is a change characterized by the popularization and proliferation of Social Network Sites/Services (SNSs), that has seen mediated communication penetrate further into the everyday, personal lives of millions of Internet users.

This research begins with an interest in the changes people experience as they connect with one another through these new

and increasingly prevalent media environments. Social Media websites, such as YouTube, Bebo, Facebook, DeviantArt and Flickr have been one of the key areas of Internet growth over the past 5 years. Through these websites users connect with one another and share both social resources, such as companionship and advice, and digital resources, such as images, code, and music.

The connections people form within these sites are typically labelled 'friendships'. However, the meaning and value of individual relationships, and the socio-mental ties that sustain them, are unclear in the design and analysis of these technologies. The value of relationships within social media websites is dictated by a network metaphor. This model of interaction has been in use since the 1950s, but has gained notoriety through Social Network Services (SNSs) such as Bebo, Facebook and MySpace. It has also recently seen a wide range of application, from in-house organisational networks to educational settings. When social networks are to be used in such a variety of settings, understanding how users interact within social networks is essential.

The Social Network

In the past 5 years a network model of interaction, or Social Network, has exploded into popular culture in the form of the Social Network Site/Service (SNS). Social network technologies have also been added to many other websites, such as interest based communities. Popular SNSs allow users to create an expression of and expand their networks of social ties. Users begin by creating a personal profile page, or online identity. This page represents a node within the network that can be connected to each other node, or user page, through the act of "Friending".

In SNSs, users are faced with new ways to express, maintain and develop relationships. According to the social network model, all connections, relationships or "friendships" within the network are treated equally. They are non-directional, public, and univariate [2, 3]. The explicit nature of connectedness in SNSs alters how we can define our close relationships, particularly when a 'friendship' label is applied to each and every relationship within our networks [4, 5]. For many users the friend label becomes ambiguous. However, Fono and Rayne-Goldie [5] argue that this conflict between label and value often goes unexpressed where users of SNSs do not engage in public interaction.

This research examines user's experiences of close tie relationships set within the boundaries of both an online community where public interaction occurs (DeviantArt.com) and a social network site where public interaction is less visible (Bebo.com). The sites offer users differing sets of social spaces and technologies in which users interact publicly. This research asks how users' experiences of socializing, socialization and connectedness in

relationships unfold within these culturally and technologically different spaces.

Contribution

The research seeks to create an understanding of the experiences of users as they navigate online social spaces to create meaningful relationships online. Not only is the immediate experience of connecting important but also the decisions we make about who to befriend and why. Friendships are important social relationships that provide belonging, identification, emotional and physical support. They have been shown to have noticeable effects on our physical, emotional and psychological well-being [6, 7]. Understanding the emotional, ethical and political considerations of making explicit connections with other people in mediated environments and the impact that navigating within a defined network have on our experiences of relating to others will serve to illuminate the developing cultural practices of social media websites. These cultural practices can inform design in social media and social technologies.

METHODOLOGY

Computer Mediated Communication is generally understood as offering limited bandwidth for the transfer of information, and lacks non-verbal cues present in face-to-face communication [8]. Walther's Social Information Processing theory [9], suggests that rather than prohibiting relationship development, these limitations only imply that relationship may take longer to develop. Virtual Ethnography [10], unlike surveys, interviews or network analyses, can be appropriately receptive to changes over time, avoiding mere snapshot representations. This is of particular importance for understanding ongoing social practices such as friendship.

As a result, virtual ethnography is seen as the most suitable tool for the engagement with and understanding of online communities and social networks as it responds to the ongoing development of cultural practices and norms and the affective and emotional responses of users to online interaction.

Furthermore, virtual ethnography recognizes that online interaction cannot command the same depth of immersion in a given culture that traditional ethnography involves. It allows that interaction fits into and around other activities, reflecting the embedded computing practices of participants. Where communities and social networks traverse physical or digital boundaries the virtual ethnographer is positioned to follow and take into account technology use as it is embedded in daily life.

Further still, online interaction is conceived as textual, and ethnography often presents cultures as amalgams of texts. A virtual ethnography, in order to be more than a simple mirror of user-presented text, highlights the personal, meaningful interaction of the researcher as important data. This participatory experience helps the researcher "read between the lines".

Added to this, the multi-method data collection employed in ethnographic studies reflects the multi-media nature of online interaction. When users interact online they share media rich data and personal texts. The ethnographer can respond to the multiple media and genres of data through interviews, photo-ethnography and participant observation. Many of the 'things' of the internet may not be as well represented in traditional data collection methods such as surveys.

The current research is carried out through participant observation as a virtual ethnographic [10] study of two social media services. The ethnography took place over a three year

period. Online and face-to-face interviews were also carried out alongside the participant observation. The participant observation occurs within two social media websites. One site is the art-based online community, DeviantArt.com, and the other is the out-and-out social network site Bebo.com.

DeviantArt.com is the leading digital art community on the Internet, with over 10 million registered accounts and 75 million submitted artworks (for comparison the Tate Collection is 66,000 pieces). In 2008 the site attracted approximately 36 million unique visits (the Tate Modern attracted just over 5 million visitors in 2007). The site is built around galleries where users can display their work and "channels" that display work in various categories, including, for example, "most recently submitted" and "most popular" works. In addition the site contains many public social media, including forums, synchronous chat, comments, and a social networking system for creating personal profiles associated with member galleries and connecting with other users.

Bebo.com is an out-and-out social network site [11]. It is primarily concerned with replicating users' offline social networks in an online space. The site launched in 2005 and within 3 years it had over 40 million registered users worldwide. In March 2008 the site sold to AOL for \$850million. During its swift rise in popularity Bebo captured the Irish social network market and in January 2007 it reached over 1 million registered Irish users, over 25% of the population and approximately 4 users per domestic broadband subscription in the country.

Individual profile pages form the basis for the site. Each profile represents a user within the social network and they can connect to any other node in the network. Users do this by creating publicly viewable friend lists, by 'friending' other people.

RESULTS

DeviantArt.com

The unfolding of the user experience on DeviantArt follows three phases of interaction that roughly reflect the user's transition from lurker to newbie to regular. During each phase users experience changing relationship processes and are involved in a growing set of activities. In addition, a change occurs in how users experience connecting to other users. The connecting experiences can be usefully categorized as "found", "sudden", or "built".

Relationships Found

Found connections occur most often when the user assumes a peripheral position within the website's membership as a lurker. At the edge of the community and as a newcomer the user must rely on ego-centric navigation for the discovery of relationships. At this point all viable connections to other members are possible and valuable. Without prior knowledge of cultural norms surrounding the "friending" process, users begin to create extensive friend lists, adding any user they see as interesting. Many of these users experience frustration when promising connections do not result in meaningful relationships. The opportunity provided by the technology to connect with any other user is at odds with the users' ability to respond to each and every connection.

For this reason, found relationships tend to be one-sided and informational, rather than dialogical and meaningful. As such, 'found' relationships do not provide for users' engagement with a community. Instead, they are most closely associated with a lurker perspective. For many users the frustration they

experience at this stage of their involvement with the website is enough to drive them away from the site.

Sudden Connections

“Sudden” connections occur when a user is added by another without any prior interaction. If users proceed beyond this lurker stage, they begin to engage in more interactions around the websites galleries. The website will begin to seem more like a community of users, centered on the gallery practices. A user will experience “sudden” connections increasingly as they post artworks themselves and begin to comment on other users work and activities around the website. In essence, sudden connections begin to occur at this stage because other users begin to recognize the user as a community member.

Sudden connections can come from new and regular users, and are a source of frustration for many users when unaccompanied by contextual information. In most other social situations, unsolicited and decontextualized connections are often treated as intrusions or spam. On the other hand, for artists on DeviantArt these connections can be admirers of their work.

The experience of dealing with sudden connections changes how the user sees the site. These connections add a sense of responsibility to others to the experience of being an artist and a participant on the site. Sudden connections create unease for some users, particularly artists who have a large following. With limited time available to them, these users feel that they cannot adequately respond to each user. As such, no norm of reciprocity of connection or interaction ever fully develops on the site.

Sudden and found relationships carry with them a number of ethical and political concerns. For instance, when users add each other on DeviantArt, they may do so in order to receive updates on an artist’s work or to receive personal updates, such as journal entries. The artist has no easy way to differentiate between the two connections. When faced with the sudden connection the artist can adopt one of two strategies; answer them all or ignore them all. Unless the user initiating the sudden connection provides further contextualizing information about the connection, the artist cannot appropriate the connection into their active personal community. Even if the user provides the contextualizing information, the artist does not have the technological tools available to express the distinction between a valuable or meaningless relationship.

Building Relationships

Where sudden and found connections apply mostly to situations where users are engaging with unknown others, “built” connections follow a period where users construct a meaningful relationship before creating an explicit connection. At this stage of their interaction a user will start to become a full member of the website’s community and see the community on its own merit. As this occurs the user becomes less interested in the website as an art gallery and more as a community.

On DeviantArt, “built” relationships are most visible in the forums. In order to begin to build relationships users must first learn to act in the way expected by other users. For this, users are expected to temporarily return to the “lurker” role of their participation, as regular users advise newcomers to “lurk more” in order to learn how to fit in. This is because, for most users, participation in the forums is an entirely different set of cultural practices than those experienced in the rest of the site. Relationships are no longer a binary connection, but a lived, dynamic and meaningful interaction.

Bebo

The “friending” experiences of Bebo users are extremely different than those of DeviantArt users. To begin, Bebo users are involved in the expression of existing, rather than developing, networks. Most Bebo users’ social networks are their offline friends and family, that is, for the most part the relationships that are expressed in Bebo could be considered built connections though little of the interaction occurs within Bebo itself.

In most cases, because the network already exists offline, the expression of that network can reach a relatively stable state in the very early stages of interaction. However, users are still faced with difficult decisions regarding the connecting experience, such as whether one user is a friend or not. Sudden or found connections become more difficult for Bebo users because those connections are supposedly from people with whom they have a pre-existing or built relationship. Because the social network model only allows for a binary representation of either friend or not, the user must compromise the value of close relationships and overvalue distant relationships in order to grow their social network.

The inherent value in flattening the social area is unclear. In Ireland, at least, the value of Bebo does not appear to reside on the site itself. Instead, as the site became increasingly popular in schools and colleges around the country, membership on the site became an indicator of social openness. However, while most 13–25 years olds have Bebo profiles, many more don’t update regularly. Engagement with the practice at any level appears to include the individual in the community of users.

Participation in the community involves losing a certain amount of control over issues such as privacy. People who use the site open themselves to being “Bebo stalked”, essentially allowing other users to see personal information. This is an unwritten but accepted feature of the site for its users. In Ireland, this “stalking” is an extension of the social and dating scene, which mostly occurs in pubs and nightclubs. The social events of one space are linked to the other. Users plan for and share photographs and stories of nights out online.

So, unlike DeviantArt, the user’s engagement with Bebo may not evolve on the site. However, the use of Bebo as an identity platform, a sort of holding-place for expression of their social selves, does change as the practices intertwine with offline activities.

DISCUSSION

As the user becomes more involved in either website they begin to unfold new functions for the site. These functions are increasingly interpersonal and offer the user changing experiences of participation and interaction. At the final stage of engagement the user is immersed in a community of practice, where the practice may no longer be the originally apparent function of the site, and instead will be a process of relationship and community management. Future social network design must account for this changing use and experience of connectedness. In out-and-out social networks, such as Bebo, the technology must help users to better define their relationships. For interest-based communities, the design must help to contextualize incoming and outgoing connections, and provide for the richer connections between users. Facilitating one-to-one friendships will allow social media services to retain users and foster meaningful engagement with the site.

For users to achieve a sense of community [12] in both cases they must participate in public interaction. On DeviantArt this sense of community is available through interaction in a

multitude of social spaces. Bebo fails to keep this sense of community online because it does not provide shared social spaces for its users. For Bebo the community exists outside the site itself, in the offline lives of its users. Introducing synchronous chat rooms or asynchronous forums would help Bebo users establish stronger ties to the service and each other.

The results also bring into question how we value one-to-one relationships online. By creating uni-directional and public ties, social networks can reduce the user's control over differentiating between users in the networks. Conversely, allowing the user to make explicit their relationships gives the user a sense of agency that is absent in haphazard face-to-face social interactions. As the social network model of connectedness limits friendships to a binary value, users explore new ways to define the grey areas between, for instance, close friend and acquaintance. Future iterations of social network technologies must take this into consideration. Rather than insisting on reducing or removing physical distance, technology must also preserve the user's sense of personal and psychological space.

Limitations and Critical Reflection

Online navigation and interaction can be highly ego-centric. Added to this, ethnography relies on the researcher's experiences as a participant to inform the analysis. While an ethnographer does not solely rely on their own experience as the explanation of a culture, indeed the narratives of the ethnography should belong to the members of the culture, there is a greater difficulty, when researching everyday cultures, differentiating between the everyday experiences of the researcher and the user. This may well be the case for the ethnography presented here.

The findings for Bebo can also be questioned. Although they may represent Irish users' experiences, they may not reflect the broader perspectives of users of SNSs. In particular, the highly localized and parochial nature of Irish social life, coupled with relatively low levels of broadband adoption, may mean that Bebo use cannot and does not fully enter the everyday lives of Irish users. Although SNS use is growing rapidly in Ireland it is still grossly overshadowed by mobile phones and particularly SMS.

FUTURE RESEARCH PLANS

As the research develops the role of individual, one-to-one relationships begin to produce the sense of community that most users, designers and service providers hope for in an online social media service. The research now looks to establish how best to facilitate these relationships through technology design. Space also opens up to examine the role of friendships in technology use trends. Particularly, the shift of users between social network sites, such as has occurred in Ireland, where Bebo use has dropped off while Facebook use has risen. Examining how users experience their connections differently before, during and after such a transition may be

useful for service providers and designs to help limit market share loss.

ACKNOWLEDGMENTS

This paper was supported by funding from the Irish Research Council for the Humanities and Social Sciences (IRCHSS).

REFERENCES

- [1] Parks, M. R. and Floyd, K. 1996. Making friends in Cyberspace. *Journal of Computer-Mediated Communication*, 1(4). Retrieved January 17, 2006, from <http://jcmc.indiana.edu/vol1/issue4/parks.html>
- [2] Donath, J. and Boyd, D. M. 2004. Public Displays of Connection. *BT Technology Journal*, 22(4), 71–82.
- [3] Jones C. and Esnault L. 2004. The metaphor of networks in learning: communities, collaboration and practice. In: *Proceedings of the Fourth International Conference on Networked Learning*, Pp. 317–323. Retrieved from WWW November 2007: <http://www.networkedlearningconference.org.uk/past/nlc2004/proceedings/>.
- [4] Boyd, D. 2006. Friends, Friendsters, and MySpace Top 8: Writing Community into Being on Social Network Sites. *First Monday*, 11(12).
- [5] Fono, D. and Raynes-Goldie, K. 2006. Hyperfriendship and Beyond: Friends and Social Norms on LiveJournal. In: Consalvo, M. and Haythornthwaite, C. (Eds.). *AoIR Internet Research Annual Volume 4*. New York, NY: Peter Lang.
- [6] Duck, S. W. 1994. *Understanding Relationship Processes*. Thousand Oaks, CA: Sage Publications.
- [7] Duck, S. W. 1983. *Friends, for life: the psychology of close relationships*. Brighton: Harvester.
- [8] Chan, D. K. S. and Cheng, G. H. L. 2004. A comparison of offline and online friendship qualities at different stages of relationship development. *Journal of Social and Personal Relationships*, 21(3), 305–320.
- [9] Walther, J. B. 1992. Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research*, 19, 52–90.
- [10] Hine, C. 2000. *Virtual Ethnography*. London: Sage Publications.
- [11] Boyd, D. M. and Ellison, N. B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 11.
- [12] McMillan, D. W. and Chavis, D. M. 1986. Sense of Community: A Definition and Theory. *Journal of Community Psychology*, 14(1), 6–23.

Assessment of Usability Evaluation Methods

Sirpa Riihiaho
 Helsinki University of Technology
 P.O. Box 9210, 02015 TKK
 Finland
 Sirpa.Riihiaho@tkk.fi

ABSTRACT

My research on usability evaluation methods is based on research projects and course assignments made at Helsinki University of Technology during the last 16 years. My personal aim has been to offer usability experts and system developers such a wide and flexible set of user testing methods that a suitable one is always found for each situation. Our goal in the user testing is getting lots of qualitative data about the problems that the users encounter as well as reporting the success stories. Therefore, we have modified the original usability testing methods over these years and learned which ones are the best in certain situations and what attributes affect to the situations. The material gathered during these years offer many perspectives to the evaluation methods, and I shall present some of them in this paper for discussion in the doctoral consortium.

Keywords

usability evaluation methods, user testing, usability test

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)] User Interfaces – ergonomics, evaluation/methodology, user-centered design

INTRODUCTION

We have been doing research and given courses on usability evaluation at Helsinki University of Technology (HUT) since 1993 [9]. During these years, we have evaluated nearly 200 systems in our research projects and course assignments. The evaluated systems have included desktop systems for professional use, smart products such as heart rate monitors and televisions, gaming slot machines and virtual worlds. The goals of the evaluations have varied as well as the systems, so we have needed a wide range of user testing methods to reach our goals.

Our usability research and teaching at HUT is strongly based on iterative user-centered product development process presented, for example, in the ISO 13407 standard [6] (see Figure 1). Our research projects are usually done in cooperation with companies, and the course assignments contribute either to the research projects or are commissions from the companies. Therefore, the methods that we teach and apply in our research projects always try to influence the development of a certain product and are tied to a certain phase of a product development process.

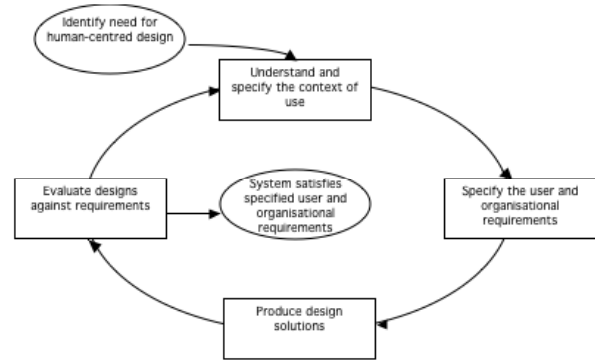


Figure 1. Iterative user-centered design activities [6].

The goals of our evaluations fall into two major categories:

1. finding usability problems in a product under development to make it better before releasing it
2. finding problems in products already in use to get ideas for improvement in the next releases or new products.

We divide the usability evaluation methods into two categories by the extent to which users are involved: those that require users' attendance, i.e., user testing methods, and those that can be used without the users, i.e., usability inspection methods (see Figure 2). On each assignment we use multiple methods: usually one or two inspection methods (typically a heuristic evaluation [11] and a cognitive walkthrough [20]) and one or two user testing methods supplemented with interviews and questionnaires.

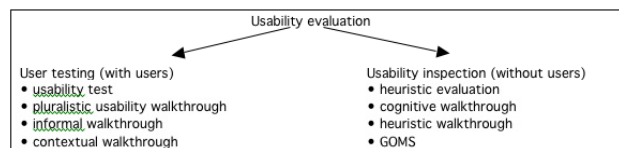


Figure 2. Two categories of usability evaluation methods.

The aim of my research has been to give usability experts and system developers a set of usability evaluation methods, especially user testing methods, wide enough and flexible enough to offer a suitable set of methods for each situation. As for example Nielsen [11] and Gould [4] present various methods for different phases in the product development cycle for the usability practitioners' methodology toolbox, my research has focused on variations of usability testing in laboratory conditions or in the field. The other methods we have applied as they have been presented in the literature.

Most of our user testing methods emphasize the users' cognitive skills and limitations instead of their effectiveness in accomplishing the test tasks. We are more interested in why the

users do something than how quickly they can do it. Therefore, we use the think-aloud method whenever possible. The respect that we have on users is reflected on the setting that we use in our tests: a test moderator is always beside the test user making the thinking aloud more natural and creating a relaxed atmosphere. This way, we get lots of qualitative data from the tests in addition to ideas for improvements. The test sessions might not be comparable to each other, but neither are the users of the real systems or the situations in which they use the systems.

Based on the experiences in these evaluations, I have written my licentiate's thesis "Experiences with usability evaluation methods" [13], some Finnish papers and one reviewed article "The pluralistic usability walk-through method" [14]. The evaluations done offer many perspectives to the material, including comparison of the methods, finding patterns and guidelines for choosing suitable methods in various situations, recognizing trends in usability evaluation and required methods, or presenting various case studies. In this doctoral consortium, I hope to get discussion on which perspectives I should take for my doctoral thesis.

OBJECTIVES OF RESEARCH

The aim of my research work has been to develop usability evaluation methods that help system designers to take into account the users during the whole development process, and to present so many alternatives for varying the basic methods that a suitable modification is always found. I have shared the idea that "something is better than nothing" that Nielsen presented in his book in 1993 [11] claiming that the simple methods have a better chance of actually being used in practical design situations than more careful and thorough methods. Therefore, we have tried to find a justified set of methods applicable with even minimal resources to help in getting the systems more usable.

There is always the risk, that if a novice evaluator makes usability inspections applying too difficult methods or methods requiring strong experience in usability evaluation, false alarms are made and the system may become even worse than before. Greenberg and Buxton also warn on the risk that wrongfully applied evaluations can quash potentially valuable ideas early in the design process and misdirect developers into solving minor problems instead of the major ones [5]. Therefore, the guidelines for selecting suitable methods should take into account many aspects, including the evaluators' expertise level, not just the number of required evaluators. Guidelines for communicating the results to the development team are also needed to make real affect on the system instead of just giving the label "usability tested".

More specified research objectives depend on the perspective chosen for my research. Different alternatives for themes include:

1. Assessment of usability evaluation methods
2. Comparison of usability evaluation methods
3. Usability evaluation patterns
4. Trends in usability evaluation
5. Variations of usability testing
6. Case studies
7. Development of new evaluation methods

The next subsections will elaborate these themes further.

Assessment of Usability Evaluation Methods

The first option is to elaborate the theme of my licentiate's thesis by assessing the evaluation methods more thoroughly in quantitative means, as for example Karat [8] does. The validity and reliability of the methods could be studied, and metrics for assessing and comparing the methods could be set and confirmed. Attributes to be measured include

- quantity and severity of usability problems revealed
- problem types revealed
- resources required to apply the methods, e.g., evaluators' skill and time
- impact of the evaluations to the development process.

The impact of the methods should be measured by studying how many of the recommended changes are really implemented to the design either in the present design or in the next versions. The ability of the methods to change the designers' and developers' attitude more positive to the usability issues could also be assessed as part of the impact ratio.

Separating the effect of various factors to the results is challenging. For example, the effect of the number of evaluators [19] and their level of expertise [11] has been studied, as well as the impact of the evaluations to the development process [17] and even the practices that make the usability reports less usable and thereby having less impact on the development process [12]. Lindgaard and Chatratchart also studied the effect of the task coverage on the test results [10]. This all shows that there are multiple factors influencing the evaluation, and varying just one factor in our experiments is not easy, if even possible. Therefore, getting statistically meaningful and valid quantitative and comparable results on various methods is very challenging and requires more research than I can conduct alone.

Comparison of Usability Evaluation Methods

The second alternative concentrates on comparing the usability evaluation methods. In my licentiate thesis, I presented a plan for comparing the methods on the basis of how well they reveal typical usability problems, the quantity and seriousness of the problems found, the time and expertise it takes to apply the method, and the phase of product development process where the method can be applied. The comparisons presented in literature, usually compare the methods or the effect of the evaluator's expertise by the number of problems they help to reveal (e.g. [7] and [3]). Comparisons like this seem a bit simplified, and do not generally take into account the differences in classifying the problems. For example, an expert evaluator may make only one remark on the order and labeling of buttons that do not follow the general standards, whereas a novice evaluator may list each button and each view having the same problem separately. The comparisons also seem to neglect the idea that the evaluation methods are originally developed for certain use and certain phases in the development process, and they are not very effective in other situations.

To support the generation and validation of metrics, a systematic comparison of various evaluation methods should be conducted. Two software systems and two smart products could be evaluated with the same set of methods. A suitable set of methods would be

- a heuristic evaluation by four usability experts
- a cognitive walkthrough with two designers and two usability experts

- a pluralistic usability walkthrough with two users, one designer and one usability expert
- a usability test with four users.

Since both heuristic evaluation and cognitive walkthrough are inspection methods, it might be better to have different evaluators applying these methods for the same system to avoid bias in the results. Therefore, six experts would be needed in the minimal: four experts could make a heuristic evaluation to the system in the beginning, after that, one of them could be the moderator in a pluralistic usability walkthrough and one in a usability test, and the ones that did not do a heuristic evaluation could participate in a cognitive walkthrough. The evaluators should change roles between the systems to compensate possible differences in their skills and experiences.

Inspection methods such as heuristic evaluation and cognitive walkthrough are developed for situations where users are not available or there are not enough resources for involving them. Therefore, comparing their results to those of user testing methods is not quite “fair”, but nevertheless gives interesting academic results on their effectiveness. The problem of separating the effect of various factors that was discussed in the previous subsection also holds here.

Usability Evaluation Patterns

The third option is to develop guidelines for selecting a suitable set of methods in various situations. This perspective requires that certain attributes affecting the situation must be identified for the base of the guidelines. For example, the level of the prototype, the resources required or available (time, users, usability experts etc.), the purpose of the evaluation and the type of the evaluated system all have an effect on the selection of suitable methods.

An important aspect in the guidelines is to give instructions on how to divide the available resources, since a combination of at least two methods is more likely to reveal more problems than one method with more test users or more evaluators.

This research would continue the work that for example Rosenbaum [16] has presented. He gave guidelines on selecting and combining usability methods depending on:

- Where you are in the product development cycle
- What questions you want to answer
- Which users you want to study
- Which use scenarios are of special interest, and why.

Rosenbaum’s guidelines deal with the whole product development cycle including both user studies and usability evaluations, whereas my research concentrates on the evaluation methods and especially on the user testing methods.

Trends in Usability Evaluation

The fourth theme concentrates on the history of the usability evaluations, i.e., how the needs for usability evaluation have evolved during these years and what the future trends look like. This theme resembles the research that Barkhuus and Rode [1] included in their study on 24 years of usability evaluation in CHI conference papers. One of their research questions was: how has empirical evaluation developed in scope during these 24 years.

We started the evaluations at HUT by applying heuristic evaluations and basic usability tests with thinking aloud. The systems included consumer electronics, such as televisions and

videotape recorders, and professional systems, such as anesthesia monitors and CAD systems. Comparisons to competing systems or comparisons between competing designs were also common at that time, but they soon faded.

Nowadays, web services are the most common systems to be evaluated, and the leisure time applications have also become more general than the professional systems. Therefore, the methods that we apply have also been modified to reflect the informal and varying use of the leisure time systems. One basic user testing method with fixed test tasks is not enough any more, if it ever has been [16].

Variations of Usability Testing

The fifth theme concentrates on the modifications that we have made to the original user testing methods, giving up-to-date examples of their use and the results they give. The full paper in this conference [15] represents this perspective as well as my article on pluralistic usability walkthrough [14]. Articles such as Rosenbaum (2000) [16], Greenberg and Buxton (2008) [5], and Scott (2009) [18] indicate that there is a need for user testing methods that are mobile, modular and contextual, and provide room for the user experience and intuitions. Contextual walkthrough and informal walkthrough methods that are presented in my full paper try to answer these challenges [15].

Case Studies

One perspective on the methods is to present some product development projects describing what methods were applied during the project, how they were applied, what results were gained, how the results were presented to the company representatives and how they greeted the results. The idea in this perspective is to present the whole development process and to analyze how the evaluations affected it.

New Evaluation Methods

The future work can also orient toward developing new evaluation methods and detailed instructions for applying them. The goal of this work would be to combine the benefits of the known methods as much as possible. The assessment and comparisons presented in the previous sections would help in finding the strengths of present evaluation methods. Features of these methods could be combined so that various aspects of usability would be evaluated and effective recommendations for improvements could be made.

The new methods should be tested, and a few iterations would probably be needed to refine the methods and to develop instructions for applying the methods. Variations of usability testing could be a part of this work.

CONTRIBUTION TO THE COGNITIVE ERGONOMICS FIELD

The emphasis on our user testing is to get into the users’ thoughts, to understand what they expect from the system, what they perceive from it and how they analyze and structure the system. In other words, we are not interested in the performance times or in getting comparable results between different users. Instead, we put much effort on getting the users relaxed and feeling comfortable to think aloud and share their thoughts with the test moderator. From this perspective, I feel that the goal in our evaluations is similar to the goals in cognitive ergonomics, i.e. getting information about the users’ thoughts instead of pure performance.

One possible contribution to the field of cognitive ergonomics from our research is, that we may find patterns, how users' perceive and structure various systems: how they proceed in first time use and what sort of models they build from the systems. These results should be reflected on the research on learnability of systems. Even without these patterns, I find that the evaluation methods emphasizing the users' thoughts and their experience on the demands that the use of the system requires are valuable in setting the researchers' and designers' thought from effectiveness to the quality of use.

METHODOLOGY

The research is based on empirical studies carried out in our research projects and as course assignments during the last 16 years. The original methods have been adopted from literature in the early 1990's (e.g. [2], [11] and [20]), and have been modified to meet our needs and values.

FUTURE PLANS

For my doctoral thesis, I have two main options: I may pick one of the perspectives, study it really deep and thorough, and write a monograph dissertation, or I can write a few more articles presenting some of the perspectives, collect them as a bundle and write a summary as an introduction to the selection of papers.

REFERENCES

- [1] Barkhuus, L. and Rode, J. A. 2007. From mice to men – 24 years of evaluation in CHI. In *Extended Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, April 28 – May 3, 2007. CHI '07. ACM Press, New York, NY. <http://www.itu.dk/~barkhuus/barkhuus-altchi.pdf>
- [2] Bias, R. 1994. The pluralistic usability walkthrough: Coordinated empathies. In: *Usability inspection methods*, J. Nielsen and R. L. Mack (Eds.). John Wiley & Sons, Inc. Pp. 63–76.
- [3] Cuomo, D. L. and Bowen, C. D. 1994. Understanding usability issues addressed by three user-system interface evaluation techniques. *Interacting with computers*, Vol. 6, No. 1, 86–108.
- [4] Gould, J. D. 1996. How to design usable systems. In: *Readings in Human Computer Interaction: Towards the Year 2000*, R. Baecker, J. Grudin, W. Buxton and S. Greenberg (Eds.). Morgan-Kaufmann. Pp. 93–121.
- [5] Greenberg, S. and Buxton, B. 2008. Usability evaluation considered harmful (some of the time). In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Florence, Italy, April 5–10, 2008. CHI '08. ACM Press, New York, NY. Pp. 111–120. DOI = <http://doi.acm.org/10.1145/1357054.1357074>.
- [6] ISO 13407. 1999. *Human-centred design processes for interactive systems*.
- [7] Jeffries, R., Miller, J., Wharton, C. and Uyeda, K. 1991. User interface evaluation in the real world: a comparison of four techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New Orleans, Louisiana, USA, April 27 – May 2, 1991. CHI '91. ACM Press, New York, NY. Pp. 119–124. DOI = <http://doi.acm.org/10.1145/108844.108862>.
- [8] Karat, J. 1997. User-centered software evaluation methodologies. In: *Handbook of human-computer interaction*, 2nd ed., M. Helander, T. K. Landauer and P. Prabhu (Eds.). Elsevier Science. Pp. 689–704.
- [9] Koivunen, M., Nieminen, M. and Riihiahho, S. 1995. Launching the usability approach: experience at Helsinki University of Technology. *SIGCHI Bull.* 27, 2, Apr. 1995, 54–60. DOI = <http://doi.acm.org/10.1145/202511.202526>.
- [10] Lindgaard, G. and Chatratchart, J. 2007. Usability testing: What have we overlooked? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, April 28 – May 3, 2007. CHI '07. ACM Press, New York, NY, 1415–1424. DOI = <http://doi.acm.org/10.1145/1240624.1240839>.
- [11] Nielsen, J. 1993. *Usability Engineering*. Academic Press.
- [12] Redish, J., Bias, R. G., Bailey, R., Molich, R., Dumas, J. and Spool, J. M. 2002. Usability in practice: Formative usability evaluations – Evolution and revolution. In *Extended abstracts on Human Factors in Computing Systems*, Minneapolis, Minnesota, USA, April 20–25, 2002. CHI '02. ACM Press, New York, NY, 885–890. DOI = <http://doi.acm.org/10.1145/506443.506647>.
- [13] Riihiahho, S. 2000. Experiences with usability evaluation methods. Licentiate's thesis. Helsinki University of Technology. Available as a pdf file: http://www.soborit.hut.fi/~sri/Riihiahho_thesis.pdf.
- [14] Riihiahho, S. 2002. The pluralistic usability walk-through method. *Ergonomics in Design. Human Factors and Ergonomics Society*, 10(3), 23–27.
- [15] Riihiahho, S. 2009. User testing when test tasks are not appropriate. In: *Proceedings of the European Conference on Cognitive Ergonomics*, Helsinki, Finland, September 30 – October 2, 2009. ECCE 2009. (In press.)
- [16] Rosenbaum, S. 2000. Not Just a Hammer: When and how to employ multiple methods in usability programs. In *Proceedings of the Usability Professionals' Association Conference 2000*. Ashville NC, August 2000. UPA 2000. <http://www.teced.com/PDFs/Upa2000-sr.pdf>.
- [17] Sawyer, P., Flanders, A. and Wixon, D. 1996. Making a difference – The impact of inspections. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, British Columbia, Canada, April 13–18, 1996. CHI '96. ACM Press, New York, NY. Pp. 376–382.
- [18] Scott, K. M. 2009. Is usability obsolete? *Interactions*, 16(3), 6–11. DOI = <http://doi.acm.org/10.1145/1516016.1516018>.
- [19] Virzi, R. A. 1992. Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors, Human Factors and Ergonomics Society*, 34(4), 457–468.
- [20] Wharton, C., Rieman, J., Lewis, C. and Polson, P. 1994. The cognitive walkthrough method: A practitioner's guide. In: *Usability inspection methods*, J. Nielsen and R. L. Mack (Eds.). John Wiley & Sons, Inc. Pp. 105–140.

User Experience of Mobile Systems and Services – A Case of Mobile News Journalism

Heli Väättäjä
Tampere University of Technology
Korkeakoulunkatu 6
33720 Tampere
heli.vaataja@tut.fi

ABSTRACT

Understanding the factors that affect user experience and how it is formed is important when developing new mobile systems and services. In my research I use case study approach to explore user experience and factors affecting it. Research is conducted with field studies and it has main emphasis on the qualitative data. Mixed methods research designs are used when applicable. Earlier theoretical frameworks and models of user experience are used in research designs as well as a basis for elaborating a framework for user experience. The research cases are related to mobile news reporting with mobile multimedia phones by professional. The aim of my work is to contribute to the theory building on user experience in mobile work, resulting in a framework of user experience based on analytical generalization and synthesis of the research results combined with earlier theories. The secondary goal is to contribute to the methods for studying and evaluating user experience as well as to provide implications for design and development of mobile systems and services in the context of the study.

Keywords

user experience, journalism, work, evaluation

ACM Classification Keywords

H.5.2 [User Interfaces]: Theory and methods, Evaluation/methodology. H.5.1 [Multimedia Information Systems]: Evaluation/methodology.

INTRODUCTION

During the last ten years user experience (UX) has been gaining increasing interest in the area of human-computer interaction (HCI). Technology in different forms is already part of people's everyday lives both in their personal life and working life. We are moving from the world of single devices and applications into a world of device and service ecosystems. These ecosystems are complex with multiple manufacturers, service providers, service developers and dynamically changing and evolving elements. Creating for example a satisfactory or engaging user experience is more challenging than before, since one manufacturer, service developer or service provider is not

able to control all of the aspects of the user experience. In addition, on one hand the convergence of devices into one device, such as a mobile multimedia phone and on the other hand the divergence of devices and their interoperability poses various challenges for users. Therefore, to be able to enable and create e.g. a "good", "fun", "pleasurable" or "satisfying" user experience, there is an urgent need to understand the phenomenon and factors affecting user experience of different types of interactive systems and services both by academia and by industry.

Current research on user experience concentrates mainly on consumer products and services. Relatively little research on user experience and its evaluation has been ongoing on work-related systems and services (see e.g. [26]). Related to user experience is the research on hedonism and user satisfaction in using IS systems for work, which has been reported for example in [15]. In addition, theory on technology acceptance [3], [5], [11], [25] and task-technology fit [7] as well as the field of ergonomics [21] have studies related to similar themes. However, since the term user experience is rarely used in work related research and even more rarely defined in the context of work, exploring user experience and factors affecting it in work related use is needed. In addition, by studying the use in real-life context enables a holistic understanding of the phenomenon. Furthermore, it is of interest, is it possible to combine the approaches from different disciplines within this research.

Mobile multimedia phones seem to provide both technical and functional capabilities, which could fulfill the demands for capturing of news content, creating news stories as well as enabling fast publishing from the field. They could be used as an all-around tool for professionals working in news organizations. However, using mobile multimedia phones in news journalism poses questions not only on the suitability of the devices on the functional level, but on a wider set of factors that affect the user experience in professional use. My research aims to gain a holistic view of these factors, taking also into account the work processes and workflows. I aim to generalizing some of the results to other fields of work, where mobile multimedia phones are used as well as looking at possible generalizations to the field of consumer products. In addition, I aim to contribute to the theory building on user experience in the context of mobile work.

RELATED RESEARCH

At the moment we are still lacking a widely accepted definition of user experience. ISO standardization is ongoing to give it a definition, but until approved I rely on definitions found in earlier literature. User experience is often defined as a consequence of the interaction between a user and a product (see e.g. [4], [6]). It is affected by the characteristics of the user, system and contextual factors ([6], [10], [4], [20]). User experience is also characterized as being subjective [27],

having a temporal dimension [19], [27] and being both reflective and recursive [31], [27].

There exists a number of theoretical frameworks for user experience (e.g. [4], [6], [9], [19], [14]) and the characteristics and components of user experience have been studied for example by Roto [20], Hassenzahl [9] and Mahlke [16]. Recently values have gained increasing interest in relation to user experience (see e.g. [2]). Furthermore, user experience of mobile devices and applications has been studied for example by Roto [20] and Swallow et al. [22]. Evaluation and measuring user experience has also been gaining increasing interest due to both needs of academia and practitioners [13].

Although there has been a considerable amount of work on user experience, we are still trying to gain a holistic understanding of user experience as a phenomenon, what contributes to it and what is in fact meant by it [1], [17], [31]. In fact, the term user experience is still often used as a synonym for usability, not making a clear distinction between what is their relation to each other or how do they differ. For example, a recent book in the field of HCI uses these terms interchangeably (see e.g. [24]). It is also of interest what is the relationship of user experience to theories on technology acceptance [3], [5], [11], [25], task-technology fit [7], and user satisfaction [15] or the research in the field of ergonomics [21]. Therefore, in my research the theoretical foundation is based on existing user experience literature, but I reflect the findings with other theories as well and aim for a synthesis from different theories and results of the research.

RESEARCH OBJECTIVES

In short – the aim of the work is to “make sense of user experience”, quoting Wright et al., [[31]] in forms which are usable for both the academia and practitioners working with mobile system and service development especially in a work related context. Main research questions are:

- What factors affect the user experience of mobile systems and services and why?
- How is user experience shaped and what are the resulting consequences?

Also a supporting question is needed in order to be able to carry out this research and to:

- How to study and evaluate user experience?

RESEARCH APPROACH AND METHODS

The research approach is empirical and mainly qualitative. Mixed methods research design is used in some case studies [23]. The case study approach is used in this research, which is exploratory by nature. Either several single case studies or multiple case studies are used to study the phenomenon of user experience and based on the synthesis of results either a framework or a model of user experience is presented. Multiple sources of data are used in each case to gain an insight into various aspects that affect the user experience and how it is formed. Examples of used data collection methods are interviews, questionnaires, observations, and focus groups. In addition, photos, videos, logs, and emails are used as a source of data. Data is analyzed primarily by content analysis [23]. For questionnaire data relevant statistical analysis methods are used.

Table 1 presents four tests which are used to deal with the quality of the research [32]. and the tactics chosen in this research to deal with them. All the research cases will use multiple sources of data, build explanations and address rival explanations, apply using theory and use case study protocol. In addition a case study database is built for each case. Other presented tactics are used in cases, in which they are applicable.

The research in the cases is designed based partly either on selected existing frameworks or models of user experience or gained pre-understanding of the subject of the study to be able to gather relevant data. However, in the analysis of the collected data the aim is not only to see do the findings fit these frameworks, but to “listen to” the data and see what emerges from it, be it completely new themes or support for earlier research. This type of approach ensures, that on one hand earlier knowledge in the field is utilized in the research design, but on the other hand new knowledge and theories can be built based on the empirical data.

Table 1. Tests and tactics of this research for dealing with quality of the research (Source: Yin, 2003).

Tests	Chosen Case Study Tactic	Phase in which chosen tactic occurs
Construct validity	Use of multiple sources of evidence (i.e. data), establish a chain of evidence, key informants review draft case study	Data collection
Internal validity	Explanation building, address rival explanations	Data analysis
External validity	Use theory in single-case studies (analytical generalization)	Research design
Reliability	Use case study protocol, develop a case study database	Data collection

Frameworks from earlier research, which have had most influence so far on my research, have been presented by Hassenzahl [9], Roto [20] and Mäkelä & Fulton Suri [19]. Hassenzahl’s framework for pragmatic and hedonic aspects of user experience [9] has been applied both in the design phase of the case on mobile journalism as well as in reporting of the results as well as in developing a questionnaire Attrak-Work for assessment of users’ perceptions of the pragmatic and hedonic qualities and overall judgment of appeal for mobile systems in the context of news journalism [29].

Due to the exploratory nature of this research, the role of the researcher is to observe and study user experience in as close to real-life like context as possible and not to affect the usage of the systems or services by interfering. However, since in some of the cases research prototypes are evaluated, expert evaluations (e.g. heuristic evaluations) and usability tests may be used before the real-life testing. In addition, field study setups when using research prototypes may not be in the context of real media organizations, but in as close to similar setups as possible.

CASE STUDIES

Case studies are conducted in co-operation with students of journalism and visual journalism at University of Tampere, Nokia Research Center, and a local newspaper as well as

professionals working for newspapers. The number of users in the cases varies from five to about twenty participants.

The system used in the field trials is developed by Nokia Research Center (NRC), and NRC provides the devices, the mobile client and the platform for mobile journalism trials. The mobile client used in the trials has evolved during the study and three different versions are tested as separate cases. Feedback from the field trials is used by NRC to develop the mobile client further. The reason for using students in a number of studies is that trialing with a research prototype with professionals in real-life usage situations may be problematic due to possible problems in the prototype.

Case 1: Field study with students of journalism and visual journalism (field study in spring 2008) – publishing directly from the field to an online web magazine [28], [29], [30].

Case 2: Interviews with early adopter professionals & design implications based on this and previous part of the study (summer 2008) [12].

Case 3: Study with students working for local news paper (fall 2008) – videos with mobile phones.

Case 4: Study with students at Mediapäivät (spring 2009).

Case 5: ? (upcoming): Field trial to be defined (fall 2009).

RESULTS OBTAINED AND A CRITICAL ANALYSIS OF THE RESULTS

The first results of the first two cases have been analyzed and also being published. From the two further cases we have preliminary results, but a further analysis is ongoing. A brief summary of some of the results is given here, more results can be found in [28], [29], [30], [12]. In the first case I utilized a number earlier theories (e.g. [9], [19], [20]) on user experience which I used in the research design for example as interview themes and in questionnaire design. I also used them as some viewpoints in analyzing and reporting the results.

The first case helped to gain a holistic picture into the field of mobile news reporting with mobile phones to an online magazine. Furthermore, it helped to gain a preliminary understanding of the factors affecting user experience in the studied field related to user, system and context. I found for example that earlier negative experiences, which had been extremely frustrating, on using mobile phones for similar tasks, lowered the expectations and the motivation to use mobile phones for given tasks. They had a clear effect on the attitudes towards further use before the trial. During and after the trial the ambition, the goals of the person herself has and tries to accomplish in conducting the assignments, the fit to professional identity and the reflection of the professional status was found to effect user experience.

I also found that in news journalism there exist requirements due to the professional publishing, which pose further criteria for the used systems and the related workflows and work processes, such as reliability of the mobile submission, deadlines and speed of publishing and error-freeness and quality of the material. The students of journalism and visual journalism emphasized different things in the used system as affecting their user experience. The limitations in the multimedia capture were emphasized by the students of visual journalism, whereas students of journalism found the usage of the mobile phone as a multimedia tool supporting them in new type of storytelling, although not favored it as a work tool. In a multiple case study

article written with a colleague, we also identified five dimensions for mobile work context, and exemplified these through the cases. I also used an elaborated version of AttrakDiff questionnaire [8] to evaluate user experience and further study would be needed on its usage in real-life work context.

To summarize the results in relation to previous results in the area, there have been partly similar findings related to for example the mobile system in several studies, also in the field of ergonomics. However, looking at the results from the point of view of journalism reveals new insights into how and why users make their judgments, reveals the situational aspect of the user experience and what the overall user experience in this context is and how it is formed. Also using the clear distinction between user, system and context related factors, offers a way of conceptualizing user experience and discussing the phenomenon in these cases. On the other hand, partly similar issues that arise in the user experience related findings have been discussed in relation to user satisfaction, technology acceptance and task-technology fit, and this needs to be critically discussed in the future papers.

Therefore, establishing what I mean by user experience based on the findings and how it differs from previous theories is important for my further work and publishing. In addition, publishing results from a mixed methods research design has proven to be sometimes challenging, and I need to find the right way of presenting the results. From the qualitative data it is not always relevant to quantify the results as sometimes asked for, if one is presenting emerging themes from the data and the number of participants is low. I need to take this into account in the future research designs as well as more carefully think about how to present the results and in what forums.

ACKNOWLEDGMENTS

This work is supported by the Ministry of Education, Finland (project PALTI, 2007–2009).

REFERENCES

- [1] Blythe, M., Wright, P., McCarthy, J. and Bertelsen, O.W. 2006. Theory and Method for Experience Centered Design. Workshop at CHI 2006. Pp. 1691–1694.
- [2] Cockton, G. 2006. Designing worth is worth designing. In Proc of the 4th Nordic Conference on Human-Computer interaction: Changing Roles. ACM Press, New York, NY. Pp. 165–174.
- [3] Davis, F. D. 1989. Perceived Usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319–339.
- [4] Desmet, P. and Hekkert, P. 2007. Framework of Product Experience. *International Journal of Design*, 1(1), 57–66.
- [5] Dishaw, M. T. and Strong, D. M. 1999. Extending the Technology Acceptance Model with Task-Technology Fit Constructs. *Information & Management*, 36, 9–21.
- [6] Forlizzi, J. and Ford, S. 2000. The Building Blocks of Experience: An Early Framework for Interaction Designers. Proc. DIS'00, Brooklyn, New York. Pp. 419–423.
- [7] Goodhue, D. L. and Thompson, R. L. 1995. Task-Technology Fit and Individual Performance. *MIS Quarterly*, 19(2), 213–236.
- [8] Hassenzahl, M. 2007. AttrakDiff™. Information available at: <http://www.attrakdiff.de>.

- [9] Hassenzahl, M. 2003. The thing and I: Understanding the relationship between users and product. In: *Funology: From Usability to Enjoyment*. Blythe, M. A. et al (Eds.). Kluwer Academic Publishers, The Netherlands.
- [10] Hassenzahl, M. and Tractinsky, N. 2006. User experience – a research agenda. *Behaviour and Information Technology*, 25(2), 91–97.
- [11] Kaasinen, E. 2008. User Acceptance of Mobile Services. In: Lumsden, J. (Ed.) *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, Vol. 1, Information Science Reference. Pp. 102–121,
- [12] Koponen, T. and Väättäjä, H. Early Adopter’s Experiences of Using Mobile Multimedia Phones in Mobile News Journalism. To appear in *Proceedings of European Conference on Cognitive Ergonomics ECCE2009*, Finland.
- [13] Law, E., Bevan, N., Christou, G., Springett, M. and Larusdottir, M. (Eds.). 2008. *Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM 2008)*, Reykjavik, Iceland, June 2008. Available at: <http://www.cost294.org> .
- [14] Law, E. and Vermeeren, A. P. O. S., Hassenzahl, M. and Blythe, M. 2007. *Towards a UX Manifesto*.
- [15] Lee, C.-P., Shim, J.P. 2006. An Empirical Study on User Satisfaction with Mobile business Applications Use and Hedonism. *Journal of Information Technology Theory and Applications*, 8(3), 57–74.
- [16] Mahlke, S. 2007. *User Experience: Usability, Aesthetics and Emotions in Human-Technology Interaction*. In: *Towards a UX Manifesto*. Law, E., Vermeeren, A. P. O. S., Hassenzahl, M. and Blythe, M. (Eds.).
- [17] McCarthy, J. and Wright, P. 2004. *Technology as Experience*, MIT Press, USA. 211 p.
- [18] Miles, M. B. and Huberman, A. M. 1994 *Qualitative data analysis: An expanded sourcebook*. 2nd Ed. Sage Publications, USA .
- [19] Mäkelä, A. and Fulton S. J. 2001. Supporting users’ creativity: design to induce pleasurable experiences. In: *Proc. Affective Human Factors Design, Asian Academic Press*. Pp. 387–394.
- [20] Roto, V. 2006. *Web Browsing on Mobile Phones – Characteristics of User Experience*. Doctoral dissertation, Helsinki University of Technology.
- [21] Salvendy, G. (Ed.). 2006. *Handbook of Human Factors and Ergonomics*. 3rd Ed. John Wiley & Sons, Hoboken, New Jersey.
- [22] Swallow, D., Blythe, M. and Wright, P. 2005. Grounding experience: relating theory and method to evaluate the user experience of smartphones. In: *Proc. of the 2005 Annual Conference on European Association of Cognitive Ergonomics*, ACM Int. Conf. Procc Series, Vol. 32. Pp. 91–98.
- [23] Tashakkori, A. T. and Teddlie, C. (Eds.). 2003. *Handbook of Mixed Methods in Social and Behavioral Research*. Sage Publications, USA.
- [24] Tullis, T. and Albert, B. 2008. *Measuring the User Experience: Collecting, Analyzing and Presenting Usability Metrics*. Morgan Kaufmann, USA.
- [25] Venkatesh, V., Morris, M. G., Davis, G. B., Davis, F. D. 2003. User Acceptance of Information Technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- [26] Vuolle, M., Kallio, T., Kulju, M., Tiainen, M., Vainio, T. and Wigelius, H. 2008. Developing a Questionnaire for Measuring Mobile Business Service Experience. In: *Proceedings of MobileHCI 2008*, ACM, NY. Pp. 53–62.
- [27] Vyas, D. and van der Veer, G. 2006. Rich Evaluations of Entertainment Experience: Bridging the Interpretational Gap. In: *Proc. 13th European Conference on Cognitive Ergonomics: trust and control in Complex Socio-Technical Systems*, ACM. Pp. 137–144.
- [28] Väättäjä, H., Männistö, A., Vainio, T. and Jokela, T. 2009. Understanding User Experience to Support Learning for Mobile Journalist’s Work. In: *The Evolution of Mobile Teaching and Learning*. Guy, R. (Ed.). Informing Science Press.
- [29] Väättäjä, H., Koponen, T. and Roto, V. Developing Practical Tools for User Experience Evaluation – A Case from Mobile News Journalism. To appear in *Proceedings of European Conference on Cognitive Ergonomics ECCE2009*, Finland.
- [30] Wigelius, H. and Väättäjä, H. Dimensions of Context Affecting User Experience in Mobile Work. To appear in the *Proceedings of INTERACT 2009*, Uppsala, Sweden.
- [31] Wright, P. C., McCarthy, J. C. and Meekison, L. 2003. Making Sense of Experience. In: *Funology: From Usability to Enjoyment*. Blythe, M. A., Overbeeke, K., Monk, A. F. and Wright, P.C. (Eds.). Kluwer Academic Publishers, The Netherlands. Pp. 43–53.
- [32] Yin, R. K. 2003. *Case Study Research – Design and Methods*. 3rd ed. Sage Publications, USA.

Is It Possible to Design for Aesthetic Experience without Damaging Agency?

Mórna Ní Chonchúir
People and Technology Research Group
Department of Applied Psychology
University College Cork, Ireland
m.nichonchuir@ucc.ie

ABSTRACT

This research examines evocations of aesthetic technological experience in Interaction Design (IXD) literature. Specifically, in theory and applications devoted to designing such aesthetic technological experiences as, *inter alia*, beauty, ambiguity and calmness. A pragmatist, cultural psychological framework, amplifying personal agency is used as an analytic framework for this work. Drawing on the pragmatist perspectives of Dewey [1] and Bakhtin [2], agency for cultural psychology underscores peoples' situated, embodied, expressive, intellectual and sensual needs and values, particularly in the context of aesthetic experience. Because agency in the cultural tradition is sensitive to the qualities of aesthetic experience to which IXD literature attests, it is well placed to assess and contribute to this line of thought and design practice. Analysis is both conceptual and empirical, the former forming the basis for this paper.

Conceptual analysis of IXD aesthetics literatures, employing the concept of agency as analytic lens, contends that agency is undervalued when aesthetic experience is evoked as (i) situated in-product (ii) deeply reflective (iii) located outside of the everyday. Broadly, the thesis seeks to stress the ethical and social responsibilities of design in its various bids to draw more humanistic concepts and experiences into the design argot.

Keywords

interaction design, agency, aesthetics, user experience

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]
User Interfaces – *user-centered design*.

AIMS AND OBJECTIVES

To provide a conceptual critique of IXD aesthetics literature and its alleged dedication to a more sensual, emotional, situated, creative and therefore agentic user. The aim here, however, is not to pour cold water on the admirable enterprise of embracing a more broad range of human experience in design. What is of value, however, is a consideration of the possibility that certain methods of designing for aesthetic experience might deplete the agentic faculties that they wish to

promote. Cultural psychology has a rich history in conceptualising the agentic textures of experience, and is particularly comfortable with agency in terms of aesthetic experience. Thus, cultural psychology offers a framework that is uniquely placed to critique claims for the designability of aesthetic experience, its implications for personal agency, and the possible pitfalls of misusing aesthetic concepts.

Design for aesthetic experience, in this research, also represents a flashpoint at which the social and ethical responsibilities of design practice and comment move into sharp focus. A broad aim of this work is to illuminate a growing attention to and critical thinking upon the possible impact of design and design discourse – of conceptualising, creating, producing and putting artifacts into the world. The phenomena that we choose to study and the ways in which we study them must be tempered by a consideration of the impact that our enquiries and modes of practice might have on those phenomena. These implications become particularly acute in the conceptualisation of aesthetic concepts in design. As concepts whose value may lie in their emergent, complex quality, an added weight of responsibility falls upon design practice and comment in their efforts to transpose humanistic concepts into design.

INTRODUCTION

Technologies are artifacts through and with which we live, work and play. The once indelible lines between work-related uses of technologies and leisure/domestic uses have become blurred so that such decipherability feels off topic. Thus, the range of experiences to which design must attend has been broadened from the original HCI concerns of usability, functionality and instrumentality to include more personally-meaningful, value-laden, and emotional as well as cognitive technological experiences. In particular, Interaction Design (IXD) is devoted to understanding the influence of modern technologies on people's cognition, emotion and value with a view to informing and critically evaluating design. This constitutes an experiential element traditionally explored by social sciences and particularly ripe for psychological contribution. The last fifteen years has seen various attempts in IXD to conceive devices and systems that engage the user in aesthetic and personal ways as might a work of art or literature, experiences at the heart of cultural psychology. IXD aesthetics evoke an holistic user with capacity and desire to engage creatively with technologies. Deeply personal experiences such as, amongst others, ambiguity, beauty and calmness, are being drawn into the design remit in a bid to flesh out and complexify the user, their capabilities and desires. As design desiderata shift to encompass the holistic user, an interface occurs between the formula of design and the richness of user experience. This is particularly acute in the case of aesthetic technological experiences, being especially difficult to square with clear-cut design heuristics. This alleged dedication to

creative, interpretive experience in a design-driven arena presents an interesting conflict that may have implications for personal agency. Conceptual analysis bears out this conflict in design discourse and reflects on the possibility that people's resources of trust, self-confidence, and imagination may be imperiled by undercooked design ontologies and practices.

METHODOLOGY

Analysis is both conceptual and empirical. The conceptual leg of the research is detailed in this paper.

Conceptual Analysis

Conceptual analysis of the IxD literature was carried out with reference to the following main agency framework foci:

1. Aesthetic experience hinges on the agentic negotiation of meaning between agent and artifact, a meaning that is amenable to change and reformulation in situated, everyday transactions [3, 4].
2. Agency conjures people as desiring openness of meaning in ways that are sensual, felt, creative and emergent in space and time [5, 6].

Technologies are cultural artifacts that structure and are structured by mediating lived action [2, 4].

RESULTS

Conceptual Analysis

Referring to these foci, conceptual analysis has identified the following main ways in which agency is underplayed in IxD:

Situating aesthetics in-product

Aoki and Woodruff [7] locate ambiguous experiences in-product and detached from actual, lived use. Interface options pre-determine the quality of ambiguity, allowing for a set number of ways in which ambiguity might be experienced. Implicit here is an adjudication of ambiguity as one or a number of experiential threads, as opposed to other less 'desirable' elements. Here, a value judgment is placed on certain instantiations of ambiguous encounters, hence relegating the possibility of others and the emergent quality of the experience in the everyday interchange between artifact and agent. Similarly, Hassenzahl [8] splits the experience of technologically-mediated beauty into visceral and reflective elements, seeing intense reflection as the true path to aesthetic experience and reducing beauty to participant-ratings on a scale from 'ugly' to 'beautiful'. This discredits the lived negotiation of what a beautiful experience might mean to the situated person. The wealth of cultural baggage – needs, values and desires bodied forth in and through the encounter – and imaginings of future, possible meanings are dealt short shrift. More than this, this work rests on the assumption that ugliness is an experience that is at odds with that of beauty; that purity of experience is desired by the user; that desire being measurable and not amenable to change. These approaches also assume that 'ambiguous' or 'ugly' are fixed effects, meaning the same to designer and user, hence, again discrediting the possibility of change, the significance of the user's sense-making faculties and particular aesthetic values and the possibility of intersection between pure and impure flavours of these experiences.

Dewey and Benson [1, 3] disputed the idea that aesthetic is locatable within an artifact. In the context of experiences of art (closely allied to the current critique of aesthetics design), artwork and the experience of it are so inseparable that art is experience in its active negotiation of meaning between object and person. Bakhtin [9] calls this agentic negotiation dialogue; people make sense of their experiences with, as well as through, an 'Other' – in this case a technological artifact – finding engagement by acting from unique emotional and volitional perspectives with agency and intention [5]. In this light, thinking of an experience of beauty, for instance, as only bearing significance in its pure form does an injustice to the aesthetic experience as integrative and oscillating between emotional, volitional, agentic and intentional threads. For instance, I might find one of the MP3 player skins in Hassenzahl's [8] study 'ugly' based on how it looks and feels in my hand and report it thus. But my imaginings of how my friends might react to my owning a flashy product when they know me as someone who rejects such trappings, coupled with the banter that I imagine that this might generate, may for me, be beautiful. My perspective on what beauty or ambiguity is and could be for me is emergent or 'becoming' – constantly re-negotiated in interpretive, everyday action. In Dewey's words: "It [experience] is 'double barreled' in that it recognizes in its primary integrity no division between act and material, subject and object, but contains them both in an unanalyzed totality" [10 pp. 10, 11]. If we can agree that strict dichotomies between actor and environment hold little water in our conceptualizations of experience, then talking about aesthetic constructs like ambiguity or beauty as inscribed in-product – invisible in a sense to the experiencing person – then the visibility or tangibility of such constructions moves into focus.

User as deeply reflective

Sengers et al. [11] suggest that designer intentions should translate to users. Reflective design encourages the user to consider the place and impact of technologies and design intentions in our lives and to consider the tacit assumptions of design practice. Though the impulse of this work is admirable – the visibility and tangibility of designer intentions and the place of the user becoming more central – there are problems with this logic. The assumption here is that the user has the time and interest to linger over and imagine the intentions of the designer of a particular artifact, that a reflective looking-under-the-hood of the system is what is desired and valued by the user. It sketches a stymied, in-the-head way of being, thereby depleting the more socially wrought, emotive and ethical 'participatory thinking' that occurs within lived interaction [2]. In this respect it is too easily squared with a cognitively-driven, reflective being. Those fleeting, spontaneous impressions of our experiences – those that occur bodily, are intensely felt and occur in-the-moment – are dealt short shrift. In the same vein, Hallnäs and Redström [12] design for 'slow' experience by suggesting that the user might work out the designer's intended use. This evokes users with time and desire to elucidate design intentions and devalues moment-to-moment interaction in the everyday.

This trend does represent a move away from approaches that try to "script and control user experience as tightly as possible...and position[s] personal experiences as engineerable" [13 pp. 347], being much more closely aligned with a user capable of and desiring agentic and intentional sense-making. However, portraying the person's role in these interactions as invested in disinterring the possible implications of design undercuts the potential for on-the-fly, serendipitous sense-making and, in the end, locates aesthetic meaning back in-artifact. Pragmatist philosopher Shusterman [14] appeals to the significance of integration

between immediate and intellectual elements of experience. The sensual and reflective coalesce in the sort of ‘felt’ experience that is fundamental to aesthetics so that reducing experience to either aspect cuts out the inter-relation between these elements and the embodied action of experience [15]. On the issue of trust in ubiquitous computing environments, Mads Bødker scorns ‘universalising’ approaches to design in which transparency of product or interface – where the ‘materiality of the technology disappears, becomes invisible’ – is mistaken for promoting user agency and control [16 pp. 2]. For Bødker, reflective HCI represents a counter to such approaches in that it presents an opportunity for the user to bear witness to and become aware of the ways in which design ontologies impact on their experiences and lives. They represent ‘a way to explore and articulate the moral and ideological underpinnings of technology’ [16 pp. 2]. Folding back onto personal agency as embodied action, however, the sensual and social characters of experience are underplayed when visibility and materiality of artifact become so prominent. Here, the medium of the aesthetic experience – the mechanics of its inscription – become too focal. The sous-layers of the experience are so prominent that the immediate, visceral and emotional may be eclipsed.

“The nature of experience can be understood only by noting that it includes an active and passive element peculiarly combined. On the active hand, experience is ‘trying’...On the passive side it is ‘undergoing’ [17 pp. 146].

Aesthetic experience as ‘extra-ordinary’

Bill Gaver and collaborators’ design-led approaches to aesthetic experience aim to create challenging, ambiguous and playful artifacts for everyday life. As a contribution to design discourse, the devotion in this work to ‘digital devices [that] help us to reflect, daydream and explore, systems...designed for our private idiosyncrasies as well our public personas, and technologies [that] don’t always tell us what to do and who to be’ is a virtue [18 pp. 2]. Further, the commitment, stitched into this body of work, to opening out a dialogue between designers and users through cultural probes bespeaks a sensitivity to the agentic person. These strands promote a design climate in which the character of aesthetics is treated as culturally-rich and indistinct – to be encountered obliquely and idiosyncratically and treated thus in design discourse – as opposed to tapped directly as an entitative, homogenised design resource. All of this said, the ‘everydayness’ of these devices and systems is disputable; the vast proportion of the pieces being more accurately described as artistic pieces and installations than everyday artifacts. Describing *The History Tablecloth*, Gaver notes that ‘it isn’t for anything, and that’s the point.’ [18 pp. 4]. Peeling back the layers of meaning that have grown around the term ‘technology’ with the digitization of our lives, we find an artifact with particular affordances and limitations that shape and are shaped by goal-directed use in the tangle and messiness of real-life use and abuse [4]. In this agentic, lived negotiation of meaning with and through artifacts we find a distinction between dead tool and culturally-charged artifact.

In this sense, the resistance in these artistic pieces to commonplace utility means that the agentic and playful working through the constraints and affordances of a given artifact in everyday activity – the obstacles and boundaries around which meaning is navigated – are dropped out. For Gaver, Beaver and Benford, ‘Allowing this ambiguity to be reflected in design... allows designers to engage users with issues without constraining how they respond’ [19 pp. 233]. This admirable bid to cede ultimate design control has ended in a place where constraint is framed as restrictive. As a counter to the tight fist of domineering modernist design ontologies that claim for the designability of

experience, the aversion to constraint here seems to go too far by letting the rein slip on constraint or designability in any guise. The claim here is not that design should aim to produce clear, communicable knowledge about what we are designing or who we are designing for, as this reinstates a discourse of ‘fit’ between user and experience and the clear-cut designability of aesthetic experience [as described in 20]. However, Gaver [18] recurses to a discourse of ‘fit’ when aesthetic experience is situated outside of the commonplace and portrayed as more readily available to the aesthetically-inclined, as the article title ‘Curious Things for Curious People’ implies [18]. There is an air of exclusivity to this discourse, as aesthetic experience is placed within a narrative of ‘extra-ordinariness’, thereby diluting the potential for the agentic person to imagine and seek out beauty in the everyday.

CRITICAL ANALYSIS IN RELATION TO PREVIOUS RESEARCH

For Bennett [21], to be entrusted with the capacity to seek out moments of wonder in the everyday is to be empowered to imagine a world in which wonder is always possible. In this way, the agentic person is conjured as a responsive agent, intoning their interactions with unique moral and ethical value in response to the other. Not only can such a climate nourish people’s propensities toward turning up aesthetic encounters, it also flatters a more overarching fullness of imagination and trust in which the *possibility* of turning up wonder in the everyday is energized. It follows, then, that openness to aesthetic experience can foster a more ethical, responsive and agentic being. With respect to this, and potentially the most worrying aspect of various IxD uses of aesthetic concepts, is what Bennett describes as a whittling down of people’s resources to *imagine* the world as a place where aesthetic experience *might* be found.

The idea of potential imaginings here idea is pivotal. For Bennett, such whittling down cuts out the embodied agent and champions atomised, reified knowledge – the stuff of myth in our always-fluid acts of meaning making. What Bennett [21 pp. 3] calls a ‘disposition’ to forging wonder can be depleted and future imaginings of, as well as self-trust in, turning up such experiences in the mundane can be deadened. Agency and the desire to negotiate creatively can be wrung dry by such discourses and moments of possible wonder can be passed by.

‘The depiction of nature and culture as orders no longer capable of *inspiring* deep attachment inflicts the self as a creature of loss and thus discourages discernment of the marvelous’ [21 pp. 7, my emphasis].

This second point is crucial for IxD aesthetics. It speaks to the importance of personal agency in aesthetic encounters not only in terms of its emergent nature. More than this, it suggests that if design can find a way to deal delicately with this emergent, agentic nature, it might imbue the person with a bigger feeling of trust in their own capacities to forge wonder and trust in technologies as possible sources of wonder. The various IxD depictions of aesthetic experience outlined above chime with Bennett’s and Black’s [22] fears about the far reach of discourses that reify aesthetic experience, expunge the embodied person or evoke aesthetic experience as extraordinary.

‘In airbrushing humanity from the world, reification renders it flat, featureless, and without wonder – in a word, reification disenchant the world’ [22 pp. 39].

The deep involvement of technologies in everyday life makes it reasonable to imagine that the theory and practice of aesthetics design carries with it a heavy duty of care. The idea that our

depictions of aesthetic experience have a ‘rhetorical power [that] has real effects’ becomes especially apparent in the context of design, as conceptualizations of aesthetics ooze down into design practice and real-world use [21 pp. 21].

‘All technologies develop within the background of a tacit understanding of human nature and human work. The use of technology in turn leads to fundamental changes in what we do, and ultimately in what it is to be human...we recognize that in designing tools we are designing ways of being’ [23 pp. xi].

FUTURE RESEARCH PLANS

The empirical research phase is now in progress. This qualitative leg offers lived exemplars of aesthetic experiences, analysed with sensitivity to the agency framework foci. This phase serves to lend real-life relevance to conceptual claims, as well as highlight the value of in-depth, empirical enquiry for IxD aesthetics. To date, interview data articulates (i) the emergent quality of aesthetic encounters, (ii) the importance of real-life use and abuse, and (iii) an overarching feeling of trust and possibility that can be fostered by aesthetic experience. The agency framework sits well with this empirical leg, being pragmatist in tradition, and because agency presents a nuanced quality most aptly illuminated by qualitative enquiry. The intention here is to offer critical considerations for IxD as opposed to prescriptive recommendations that might lend themselves to more ‘efficient’ uses of aesthetics concepts in design [as argued in 20]. This close study of lived aesthetic experience also finds root in calls for empirical research as a valuable starting point for rich research practices as opposed to formalized knowledge [15].

ACKNOWLEDGMENTS

This research is funded by the Irish Research Council for the Humanities and Social Sciences (IRCHSS).

REFERENCES

- [1] Dewey, J. 1934. *Art as Experience*. Perigee: London.
- [2] Bakhtin, M.M. 1993. *Toward a Philosophy of the Act*. M. Holquist (Ed.). V. Lianupov (Trans and notes). University of Texas Press: Austin, TX.
- [3] Benson, C. 2001. *The Cultural Psychology of Self: Place, Morality and Art in Human Worlds*. Routledge: London.
- [4] Cole, M. 1996. *Cultural Psychology: A once and future discipline*. Harvard: London.
- [5] Bruner, J.S. 1990. *Acts of Meaning*. Harvard University Press: Cambridge, MA.
- [6] Ratner, C. 2002. *Cultural Psychology: Theory and Method*. Kluwer: New York.
- [7] Aoki, M. and Woodruff, A. 2005. *Making Space for Stories: Ambiguity in the Design of Personal Communication Systems*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. Portland, Oregon, April 02–07, 2005. ACM Press, New York. Pp. 181–190. Doi: <http://doi.acm.org/10.1145/1054972.1054998>.
- [8] Hassenzahl, M. 2004. *The Interplay of Beauty, Goodness, and Usability in Interactive Products*. *Human-Computer Interaction*, 19(4), pp. 319–349.
- [9] Bakhtin, M.M. 1981. *The Dialogical Imagination: Four Essays*. M. Holquist (Ed.). C. Emerson and M. Holquist (Trans.). University of Texas Press: Austin, TX.
- [10] Dewey, J. 1925. *Experience and nature*. Open Court: La Salle, IL.
- [11] Sengers, P., Boehner, K., David, S. and Kaye, J.J. 2005. *Reflective design*. In: *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*. Aarhus, Denmark August 20–24, 2005. ACM Press, New York, 49–58. Doi: <http://doi.acm.org/10.1145/1094562.1094569>.
- [12] Hallnäs, L. and Redström, J. 2001. *Slow Technology – Designing for Reflection*. *Personal and Ubiquitous Computing*, 5: 201-212. Doi: 10.1007/PL00000019.
- [13] Sengers, P., Boehner, K., Mateas, M., and Gay, G. 2008. *The disenchantment of affect*. *Personal and Ubiquitous Computing*, 12: 347–358. Doi: <http://doi.acm.org/10.1007/s00779-007-0161-4>.
- [14] Shusterman, R. 2000. *Pragmatist aesthetics: Living Beauty, rethinking art*, 2nd Ed. Rowman and Littlefield, Boston, MA.
- [15] Dourish, P. 2001. *Where the action is: The foundations of embodied interaction*. MIT Press: Cambridge, MA.
- [16] Bødker, M., 2008. *Dealing with pervasive uncertainty – trust, digital environments and reflective HCI*. Retrieved: <http://cemcom.infosci.cornell.edu/reflectiveHCI/papers/bodker.pdf> on October 21st 2008.
- [17] Dewey, J. 1966. *How we think*. Free Press: New York.
- [18] Gaver, W. In Press. *Curious Things for Curious People*. Retrieved: <http://www.goldsmiths.ac.uk/interaction/pdfs/36.gaver.curiousThings.inPress.pdf> on November 5th 2008.
- [19] Gaver, W., Beaver, J. and Benford, S. 2003. *Ambiguity as a resource for design*. In: *Proceedings of CHI 2003*. Ft. Lauderdale, Florida April 5–20, 2003. ACM Press, New York. Pp. 233–240. Doi: <http://doi.acm.org/10.1145/642611.642653>.
- [20] Redström, J. 2006. *Towards user design? On the shift from object to user as the subject of design*. *Design Studies* 27, pp. 123–139.
- [21] Bennett, J. 2001. *The Enchantment of Modern Life: Attachments, Crossings and Ethics*. Princeton University Press: Princeton.
- [22] Black, D. 2002. *The Politics of Enchantment: Romanticism, Media and Cultural Studies*. Wilfrid Laurier Press: Ontario.
- [23] Winograd, T. and Flores, F. 1991. *Understanding computers and cognition*. Addison-Wesley: Reading, MA.

Author Index

Acharya, Karthikeya.....	423	Grote, Gudela.....	411
Ala-Laurinaho, Arja	289, 301	Gunawan, Lucy T.	85
Alers, Hani	85	Halonen, Minna	119
Andersson, Jonas	107, 111	Hammoud, Lamis	69
Bannon, Liam	29	Hänninen, Riikka	73
Barcenilla, Javier	374	Haradji, Yvon	139, 145
Bastien, J. M. Christian	374	Harr, Rikard	307
Beale, Russell	154	Hasu, Mervi	297
Benvenuti, Laura	427	Hébert, Anne-Marie	355
Bergström-Lehtovirta, Joanna	389	Helfenstein, Sacha	236
Bidet, Alexandra.....	416	Heylen, Dirk	131
Biemans, Margit	147	Hoc, Jean-Michel.....	194
Bligård, Lars-Ola.....	111	Hoggan, Eve	363
Boos, Daniel.....	411	Hutton, Robert	92
Boy, Guy A.	187	Hyll, Nikolaj	202, 267
Brangier, Éric	61, 374	Ihamäki, Pirita.....	177
Brinkman, Willem-Paul.....	69, 85, 169, 323	Jonker, Catholijn M.	323
Broekens, Joost.....	323	Kaasinen, Eija	127
Burkhardt, Jean-Marie	355	Kallio, Katri	119
Burmistrov, Ivan.....	115	Kämäräinen, Anna	393
Cañas, José J.....	45	Kaptelinin, Victor	29, 307
Cao, Yujia.....	96	Kariniemi, Arto.....	289
Charron, Camilo	194	Keinonen, Turkka	19
Clarkson, P. John.....	370	Kerosuo, Hannele	281
Colford, Nick.....	92	Khan, Nawaz.....	381
Comber, Rob	430	Kivisaari, Sirkku	293
Dadashi, Nastaran.....	210	Kodagoda, Neesha	381
Détienne, Françoise	355	Koester, Thomas	202, 267
Dijk, Betsy van	147, 349	Koponen, Tiina	65, 240
Dinet, Jérôme.....	374	Kopp, Steve	61
Dörner, Ralf.....	161	Koskinen, Hanna.....	275
Dugdale, Julie.....	139	Kronenfeld, Nathan.....	397
Edmondson, William.....	154	Kühn, Guido	161
Ericsson, K. Anders.....	333	Kujala, Tuomo	341, 389
Flach, John M.	11	Kuuva, Sari	73
Franssila, Heljä.....	403	Kynsilehto, Minna	123
Fréjus, Myriam	139, 145	Laarni, Jari	275
Galliers, Julia.....	315	Lagerstam, Else.....	177
Grommes, Patrick.....	92	Langdon, Patrick.....	370

Leclercq, Pierre	355	Safin, Stéphane	355
Lehtinen, Ville.....	73	Salembier, Pascal.....	139
Lehtonen, Mikko	411	Salo, Leena	258
Lievonen, Mirja.....	161	Salovaara, Antti	236
Liinasuo, Marja	92	Sandino, Guntur.....	169
Lind, Morten.....	37	Santosa, Stephanie	397
Loomes, Martin	79	Savioja, Paula	92, 258
Love, Steve.....	69	Schaupp, Marika.....	289
MacInnes, Joseph	397	Seppänen, Laura	301
Mancero, Gabriela.....	79	Sharples, Sarah	210
Mannonen, Petri	403	Sobolewska, Emilia	221
Mast, Charles van der.....	169	Stage, Jan.....	202, 267
Mattelmäki, Tuuli.....	297	Suomalainen, Tiia.....	389
McCuaig, Judi	397	Theune, Mariët.....	96
Michel, Gabriel.....	374	Thunberg, Anna.....	251
Mieczkowski, Anna	370	Toiviainen, Hanna.....	281
Milleville-Pennel, Isabelle.....	194	Toivonen, Ville.....	73
Neerinx, Mark.....	85	Turner, Phil.....	53, 221
Nemery, Alexandra.....	61	Turner, Susan.....	53
Ní Chonchúir, Mórna.....	442	Tuuri, Kai.....	363
Niemelä, Marketta	127	Väänänen-Vainio-Mattila, Kaisa	177
Nijholt, Anton.....	131, 96	Väätäjä, Heli	65, 438, 240
Norros, Leena	92, 258, 275	Veer, Gerrit C. van der.....	349, 427, 131
Olsson, Thomas	123, 127, 177	Ventä-Olkkonen, Leena.....	177
Oomes, Augustinus H. J.	85	Vet, Paul E. van der	349
Osvaider, Anna-Lisa.....	107, 251	Vivian, Robin.....	374
Oulasvirta, Antti	73, 236, 333, 389	Vyas, Dhaval.....	131
Perron, Laurence	355	Wahlström, Mikael	236
Pirhonen, Antti	363	Walkowski, Swen	161
Poizat, Germain	145	Wassink, Ingo	349
Pommeranz, Alina.....	323	Wiggers, Pascal.....	323
Rämänen, Jussi	389	Wilson, John R.	210
Randell, Rebecca.....	315	Wilson, Leanne.....	53
Riihiaho, Sirpa.....	228, 434	Wilson, Stephanie	315
Roos, Marco	349	Wong, B. L. William	381
Rosenberg, Duska.....	161	Wong, William	79
Roto, Virpi.....	240	Woodward, Peter	315
Saari, Eveliina	119, 293	Wright, William.....	397
Saariluoma, Pertti	45, 73, 393	Ylirisku, Salu	297

Author(s) Leena Norros, Hanna Koskinen, Leena Salo & Paula Savioja		
Title ECCE 2009 – European Conference on Cognitive Ergonomics Designing beyond the Product – Understanding Activity and User Experience in Ubiquitous Environments		
Abstract ECCE 2009 was held 30 September – 2 October 2009 in Finland. The conference site was Dipoli Conference Center located in the campus of Aalto University in Otaniemi. The event was organised by the European Association of Cognitive Ergonomics and hosted by VTT Technical Research Centre of Finland. The conference dealt with the challenges of designing intelligent human-system interaction in work and everyday environments. ECCE 2009 focused on the topic “Designing beyond the product – Understanding activity and user experience in ubiquitous environments”. The conference provided an opportunity for researchers, practitioners, and designers to exchange new ideas and practical experience e.g. on the following topical areas: <ul style="list-style-type: none"> • Challenges in work and everyday activities: Does the invisible computer change our ways of acting? • Means for coping: How to support sense making in ubiquitous environments? Research and design approaches: What is happening to cognitive systems engineering?		
ISBN 978-951-38-6339-5 (soft back ed.) 978-951-38-6340-1 (URL: http://www.vtt.fi/publications/index.jsp)		
Series title and ISSN VTT Symposium 0357-9387 (soft back ed.) 1455-0873 (URL: http://www.vtt.fi/publications/index.jsp)		Project number 25749
Date September 2009	Language English	Pages 447 p.
Name of project		Commissioned by
Keywords human factors, human-centred design		Publisher VTT Technical Research Centre of Finland P.O. Box 1000, FI-02044 VTT, Finland Phone internat. +358 20 722 4520 Fax +358 20 722 4374

European Conference on Cognitive Ergonomics 2009

DESIGNING BEYOND THE PRODUCT – UNDERSTANDING ACTIVITY AND USER EXPERIENCE IN UBIQUITOUS ENVIRONMENTS

ECCE 2009 was held 30 September – 2 October 2009 in Finland. The conference site was Dipoli Conference Center located in the campus of Aalto University in Otaniemi. The event was organised by the European Association of Cognitive Ergonomics and hosted by VTT Technical Research Centre of Finland. The conference dealt with the challenges of designing intelligent human-system interaction in work and everyday environments. ECCE 2009 focused on the topic “Designing beyond the product – Understanding activity and user experience in ubiquitous environments”. The conference provided an opportunity for researchers, practitioners, and designers to exchange new ideas and practical experience e.g. on the following topical areas:

- Challenges in work and everyday activities: Does the invisible computer change our ways of acting?
- Means for coping: How to support sense making in ubiquitous environments?
- Research and design approaches: What is happening to cognitive systems engineering?

ISBN 978-951-38-6339-5 (soft back ed.)
ISSN 0357-9387 (soft back ed.)

ISBN 978-951-38-6340-1 (URL: <http://www.vtt.fi/publications/index.jsp>)
ISSN 1455-0873 (URL: <http://www.vtt.fi/publications/index.jsp>)

