Dissertation

VTT PUBLICATIONS 765

Juha Pärkkä

Analysis of Personal Health Monitoring Data for Physical Activity Recognition and Assessment of Energy Expenditure, Mental Load and Stress

VTT PUBLICATIONS 765

Analysis of Personal Health Monitoring Data for Physical Activity Recognition and Assessment of Energy Expenditure, Mental Load and Stress

Juha Pärkkä

Thesis for the degree of Doctor of Technology to be presented with due permission for public examination and criticism in Tietotalo building, room TB222 at Tampere University of Technology, on the 21st of June 2011 at 12 o'clock noon.

ISBN 978-951-38-7740-8 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7741-5 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

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

Edita Prima Oy, Helsinki 2011

Juha Pärkkä. Analysis of Personal Health Monitoring Data for Physical Activity Recognition and Assessment of Energy Expenditure, Mental Load and Stress [Henkilökohtaisessa terveydentilan seurannassa syntyvän mittaustiedon analyysiä fyysisten aktiviteettien tunnistamista sekä energian-kulutuksen, henkisen kuormituksen ja stressin arviointia varten]. Espoo 2011. VTT Publications 765. 103 p. + app. 54 p.

Keywords personal health monitoring, biosignal processing and classification, physical activity, activity recognition, energy expenditure, mental load, stress

Abstract

Personal health monitoring refers to the long-term health monitoring that is performed in uncontrolled environments instead of a laboratory, for example, at home or by using wearable sensors. The monitoring is done by individuals alone, usually without guidance from health care professionals. Data produced by personal health monitoring (for example, actigraphy, heart rate, etc.) are currently used more in personal wellness monitoring rather than in clinical decision-making, because of challenges in the interpretation of the long-term and possibly unreliable data. Automatic analysis of long-term personal health monitoring data could be used for the continuous recognition of changes in individual's behavior and health status, and to point out which everyday selections have a negative effect on health and which have a positive effect. This can not be achieved by using sparse measurements in controlled environments.

In this thesis, data analysis was carried out for the recognition of physical and mental load using data from wearable sensors and other self-measurements. Large, annotated data libraries were collected in real-life or realistic laboratory conditions for the purpose of the development of practical algorithms and the identification of the most information-rich sensors and signal interpretation methods. Time and frequency domain features were computed from raw sensor data for the correlation analysis and the automatic classification of the personal health monitoring data. The decision tree, artificial neural network, K-Nearest Neighbor and a hybrid of a decision tree and artificial neural network classifiers were used.

Automatic activity recognition aims at recognizing individual's activities and postures using data from unobtrusive, wearable sensors. Similarly, the unobtrusive, wearable sensors can be used for the assessment of energy expenditure.

The quantities measured in this thesis include acceleration, compass bearings, angular rate, ECG, heart rate, respiratory effort, illumination, temperature, humidity, GPS location, pulse plethysmogram, skin conductance and air pressure. The results indicate that several everyday activities, especially those with regular movements, can be recognized with good accuracy. The energy expenditure estimate obtained using movement sensors was found accurate in activities involving regular movements. The sensors that react to the change of activity type without delay were found the most useful for activity recognition. These include accelerometers, magnetometers, angular rate sensors and GPS location sensors.

Automatic assessment of mental load aims at measuring the level of mental load during everyday activities using data from wearable sensors. The assessment of long-term stress aims at finding measures that reflect the perceived stress level, either directly or as observed through changes in behavior. Data were collected with people suffering from long-term work-related stress and participating in a rehabilitation program. Automatic measurements of recovery, measured with a bed sensor, actigraphy and bedroom illumination sensors were found to correlate best with the self-assessed stress level.

Careful selection of sensor types, sensor locations and input features played a more critical role in successful classification than the selection of a classifier. Computational complexity of the classifier's classification phase has an impact on the power consumption of a hosting mobile terminal. Power consumption is one of the bottlenecks in long-term personal health monitoring solutions today.

Juha Pärkkä. Analysis of Personal Health Monitoring Data for Physical Activity Recognition and Assessment of Energy Expenditure, Mental Load and Stre [Henkilökohtaisessa terveydentilan seurannassa syntyvän mittaustiedon analyysiä fyysisten aktiviteettien tunnistamista sekä energiankulutuksen, henkisen kuormituksen ja stressin arviointia varten]. Espoo 2011. VTT Publications 765. 103 s. + liitt. 54 s.

Avainsanat personal health monitoring, biosignal processing and classification, physical activity, activity recognition, energy expenditure, mental load, stress

Tiivistelmä

Henkilökohtaisen terveydentilan seuranta viittaa pitkäaikaismittauksiin, joita tehdään laboratorion sijaan kontrolloimattomissa oloissa, esimerkiksi kotona tai puettavien antureiden avulla. Mittauksia tekee yksilö itse, yleensä ilman terveydenhuollon ammattilaisten ohjausta. Henkilökohtaisen terveydentilan seurannasta kertyvää dataa, esimerkiksi aktigrafiaa tai sykettä, käytetään tällä hetkellä enemmän henkilökohtaiseen terveyden seurantaan kuin kliiniseen päätöksentekoon. Tämä johtuu paikoin epäluotettavan pitkäaikaisdatan tulkinnan haasteista. Pitäkaikaismittauksilla voidaan kuitenkin jatkuvasti arvioida muutoksia yksilön käyttäytymisessä ja terveydentilassa ja osoittaa, millä valinnoilla on terveyden kannalta positiivisia, millä negatiivisia vaikutuksia. Tähän ei päästä harvoilla yksittäismittauksilla kontrolloiduissa oloissa.

Tässä työssä käytettiin puettavien antureiden ja muiden henkilökohtaisten mittausten avulla kerättyä dataa yksilön fyysisen aktiivisuuden ja henkisen kuorman profilointiin automaattisen data-analyysin avulla. Tutkimuksissa kerättiin laajoja, annotoituja datakirjastoja jokapäiväistä elämää vastaavissa ympäristöissä. Datakirjastojen avulla tunnistettiin parhaita antureita sekä kehitettiin käytännönläheisiä algoritmeja datan automaattista tulkintaa varten. Aika- ja taajustason piirteitä laskettiin antureiden tuottamasta raakadatasta korrelaationanalyysiä ja henkilökohtaisen terveysdatan automaattista luokittelua varten. Työssä käytettiin luokittelijoina binäärisiä päätöspuita, neuraaliverkkoja, k-lähimmän naapurin luokittelijaa (KNN) sekä päätöspuun ja neuraaliverkon yhdistelmäluokittelijaa.

Automaattisen aktiviteettien tunnistuksen tavoitteena on tunnistaa käyttäjän aktiviteetit ja asennot päälle puettavien mutta huomaamattomien ja liikkumista häiritsemättömien antureiden avulla. Samoja antureita voidaan käyttää myös automaattiseen energiankulutuksen tunnistamiseen. Tässä työssä mitattiin kiih-

tyvyyksiä, kompassisuuntaa, kulmanopeutta, EKG:ta, sykettä, hengitysliikkeitä, valoa, lämpötilaa, kosteutta, GPS-paikkaa, pulssipletysmogrammia, ihon johtavuutta sekä ilmanpainetta. Tulosten perusteella useita arkipäiväisiä aktiviteetteja, erityisesti toistuvaa liikettä sisältäviä, voidaan tunnistaa automaattisesti hyvällä tarkkuudella. Liikeantureiden datasta laskettava energiankulutusarvio toimii hyvällä tarkkuudella toistuvaa liikettä sisältävillä aktiviteeteilla. Aktiviteettien tunnistuksen kannalta parhaiksi antureiksi osoittautuivat ne, joiden ulostulosignaali muuttuu välittömästi aktiviteettityypin vaihtuessa. Näitä antureita olivat kiihtyvyysanturit, magnetometrit, kulmanopeusanturit sekä GPS-paikannin.

Henkisen kuormituksen automaattisen arvioinnin tavoitteena on henkisen kuormituksen tason mittaaminen puettavilla antureilla, osana arkipäivän elämää. Pitkäaikaisen stressin automaattisen arvioinnin tavoitteena on löytää mittareita, jotka kuvastavat yksilön kokeman stressin voimakkuutta joko suoraan tai käyttäytymismuutoksia seuraamalla. Työssä kerättiin dataa pitkäaikaisesta työstressistä kärsivien ja kuntoutusohjelmaan osallistuvien henkilöiden avulla. Automaattisesti palautumista mittaavien antureiden, sänkyanturin, aktigrafin sekä makuuhuoneeseen sijoitetun valoanturin, tuottama data korreloi voimakkaimmin itseraportoidun stressin kanssa.

Anturityypin, anturin sijainnin sekä piirteiden valinnalla oli suurempi rooli onnistuneessa luokittelussa kuin luokittelijan valinnalla. Luokittelijan luokitteluvaiheen laskennallinen monimutkaisuus vaikuttaa akkukäyttöisen laitteen tehonkulutukseen. Tehonkulutus on tällä hetkellä yksi pitkäaikaisen, henkilökohtaisen terveydentilan seurannan pahimmista pullonkauloista.

Preface

The research presented in this thesis was carried out in years 2003 to 2010 at VTT Technical Research Centre of Finland, in Tampere. At that time, I had the privilidge of working with a wonderful group of people, who were very enthusiastic researchers and very kind and open human beings at the same time.

I want to express my most sincere gratitude to my instructor, Docent Ilkka Korhonen. During this time he led our team, our cluster and acted as a research professor at VTT. He inspired many thoughts presented in this thesis and gave practical advice and guidance during the research and the process of writing my dissertation.

My supervisor, Adjunct Professor Alpo Värri, had already guided my M.Sc. thesis at Tampere University of Technology. I was lucky to find him again to be my supervisor when the time for this doctoral thesis came. His suggestions were very useful during the writing of this thesis.

I also want to express my gratitude to my colleagues at VTT Tampere. Miikka Ermes was my closest colleague during the research on activity recognition. I want to thank him for the fruitful collaboration and for co-authoring many of the publications. I also want to thank Luc Cluitmans for fruitful collaboration and for the professional implementation of many useful software applications. I also thank Mark van Gils for his advice and comments, especially on statistical issues. I also want to thank Kari Antila for fruitful collaboration and professional project management. I also want to express my gratitude to Juho Merilahti who was my closest colleague during the stress monitoring study. The project would not have been as productive and enjoyable without his collaboration and humour. I also want to express my gratitude to Elina and Jussi Mattila for collaboration on the stress monitoring study and to Pasi Välkkynen for collaboration on the first data collection on activity recognition.

I express my gratitude to colleagues at VTT in Oulu: Panu Korpipää, Jani Mäntyjärvi, Arto Ylisaukko-Oja, Jouko Vilmi and Johannes Peltola for collaboration on activity recognition studies and for Esko Malm for collaboration on the stress monitoring study.

I want to express my gratitude to Pertti Huuskonen, Jari Kangas, Jukka Salminen, Heikki Nieminen and Ole Kirkeby of Nokia. I thank them all for fruitful collaboration and inspiring discussions.

I also want to thank Ari Mänttäri, Ulla Hakala, Kirsi Mansikkamäki and Mikael Fogelholm of the UKK institute for collaboration on the assessment of energy expenditure. I thank them all for sharing their expertise with us.

I also want to express my gratitude to Veikko Koivumaa of Suunto and Akseli Reho of Clothing+ for collaboration on the activity recognition studies. I thank them for their constructive and inspiring comments during the studies.

I also want to express my gratitude to Ari Saarinen of Rokuan Kuntokeskus and Martti Tuomisto of University of Tampere for their efforts and comments during the stress monitoring study.

I also express my gratitude to Antti Särelä of IST International Security Technology for constructive comments during the stress monitoring study.

I also want to express my gratitude to the pre-examiners of my thesis, Professor Tapio Seppänen of University of Oulu and Research Director, Docent Heikki Tikkanen of University of Helsinki and Foundation of Sports and Exercise Medicine for their constructive comments.

I express my gratitude to VTT Technology Manager Eero Punkka for his support during my dissertation process. The financial support from the Signe and Ane Gyllenberg Foundation is gratefully acknowledged.

Last but not least, I want to thank my wife Maarit for all her support during my dissertation process, and my two sons Matias and Mikko for asking many questions about my work and also for giving me other things to think about during the process of writing my thesis.

Tampere, May 2011

Juha Pärkkä

Contents

List of Publications

P1. Pärkkä, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J. & Korhonen, I. **Activity Classification Using Realistic Data from Wearable Sensors**. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 10, No. 1, January 2006, pp. 119–128.

P2. Ermes, M., Pärkkä, J., Mäntyjärvi, J. & Korhonen, I. **Detection of Daily Activities and Sports with Wearable Sensors in Controlled and Uncontrolled Conditions**. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, No. 1, 2008, pp. 20–26.

P3. Pärkkä, J., Ermes, M., Antila, K., van Gils, M., Mänttäri, A. & Nieminen, H. **Estimating Intensity of Physical Activity: A Comparison of Wearable Accelerometer and Gyro Sensors and 3 Sensor Locations**. *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, France, 23–26 August 2007, pp. 1511–1514.

P4. Pärkkä, J., Cluitmans, L. & Ermes, M. **Personalization Algorithm for Real-Time Activity Recognition using PDA, Wireless Motion Bands and Binary Decision Tree**. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 14, No. 5, 2010, pp. 1211–1215.

P5. Pärkkä, J., Ermes, M. & van Gils, M. **Automatic Feature Selection and Classification of Physical and Mental Load using Data from Wearable Sensors**. *Proceedings of the 10th International Conference on Information Technology and Applications on Biomedicine (ITAB 2010)*, Corfu, Greece, 2.–5. November 2010, pp. 1–5.

P6. Pärkkä, J., Merilahti, J., Mattila, E., Malm, E., Antila, K., Tuomisto, M., Saarinen, A., van Gils, M. & Korhonen, I. **Relationship of Psychological and Physiological Variables in Long-term Self-monitored Data during Work Ability Rehabilitation Program**. *IEEE Transactions on Information Technology in Biomedicine*, Vol. 13, No. 2, 2009, pp. 141–151.

Author's Contribution

This thesis is based on 6 publications. Author's contribution in these studies and publications was as follows:

P1. The author had the main responsibility for the study design and running the experiments. He shared the responsibility for feature extraction, feature selection and classification with M. Ermes. He was the main author of the publication.

P2. The author had the main responsibility for updating the study design from Study 1 and running the experiments. He shared the responsibility for feature extraction, feature selection and classification with M. Ermes. He co-authored the publication.

P3. The author shared the responsibility for the study design with the co-authors. He had the main responsibility for feature extraction and data analysis. He was the main author of the publication.

P4. The author had the main responsibility for the study design and running the experiments. He had the main responsibility for feature extraction, feature selection and classification algorithms. He was the main author of the publication. The online implementation of the algorithms on a PDA was written by L. Cluitmans.

P5. This study used the data set collected in Study 2. The author had the main responsibility for the new feature extraction, feature selection and the classification. He was the main author of the publication.

P6. The author shared the responsibility for the study design and running the experiments with the co-authors. He had the main responsibility of feature extraction and data analysis. He was the main author of the publication.

Abbreviations

ANN	Artificial Neural Network
BBI	Bergen Burnout Inventory
BMI	Body Mass Index
DFT	Discrete Fourier Transform
DSP	Derogatis Stress Profile
DSPtss	Derogatis Stress Profile, total stress score
ECG	Electrocardiogram
EDA	Electro-Dermal Activity
EE	Energy Expenditure
EMFit	Electromechanical Film Technology
FFT	Fast Fourier Transform
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability
IQ	Intelligence Quotient
KNN	K-Nearest Neighbor Algorithm
MBI	Maslach Burnout Inventory
MET	Metabolic Energy Turnover (otherwise Metabolic Equivalent)
MLP	Multi-Layer Perceptron
PDA	Personal Digital Assistant
PPG	Pulse Plethysmogram
PSD	Power Spectral Density
RIP	Respiratory Inductive Plethysmogram
SpO_2	Blood Oxygen Saturation (from pulse oximeter)
SFS	Sequential Forward Search feature selection algorithm
WD	Wellness Diary mobile phone application
WHO	World Health Organization

1. Introduction

Long-term personal health monitoring, as part of everyday life, is a central element of care in certain areas of health and disease management. For example, long-term monitoring of weight, blood pressure and blood glucose is used in health and disease management (Lappalainen et al. 2005, Verberk et al. 2007, Martin et al. 2006). However, the overall wellbeing of an individual is influenced by physiological, psychological and social factors. All these three factors interact as determinants of health. In fact, behavioral and social factors explain more than 50% of health outcomes (McGinnis et al. 2002). Thus, there is a need for comprehensive, long-term health monitoring that uses both physiological and behavioral monitoring.

Sedentary lifestyle is a common risk factor in chronic diseases in a modern lifestyle. At least 60% of the world's population fail to achieve the minimum recommendation of 30 minutes of moderate-intensity physical activity daily (Puska et al. 2004). High mental load is a major problem in public health in modern society. In 13 countries of the Organisation for Economic Co-operation and Development (OECD), mental disorder causes one third of disability pensions (OECD 2009). Prolonged stress has been suspected to cause physical illness (Honkonen et al. 2006), and burnout has been found to be related with all-cause mortality in people under 45 years of age (Ahola et al. 2010). Both the volume of physical activity and the volume of mental load can be modified with behavioral changes. Thus, there is a need to provide new tools that help individuals to analyze their current lifestyle and motivate them to a healthier lifestyle.

Today, an ageing population is a burden and cost on health care systems in many countries. This has accelerated the ongoing change from reactive and centralized care to proactive and patient-centric health care, where the patient is given more responsibility for his/her care, with health care professionals supporting him/her in the care process. A central theme is to improve patients' self-efficacy. Self-efficacy refers, for instance, to the patient having the correct information, skills and motivation to do self-care. Thus, the emphasis is moving from improving patients' compliance to improving patients' self-efficacy. The patient-centric care has been shown to be beneficial, for example, in terms of changes in the health status, patients' self-efficacy and health care resource utilization (Lorig et al. 2001).

The miniaturization of sensors, and electronics in general, have made many new sensors truly wearable. Similarly, developments in solid-state memory technology and communications technologies have opened new possibilities for long-term personal health monitoring. The solid-state memories can store persistent data in ambulatory conditions more reliably than the earlier micromechanical drives with moving heads and spinning disks. In addition, the capacity of the solid-state data storage devices has grown and allows longer recordings with sufficient sampling frequencies for many new applications. Developments in wireless communications technologies have enabled wireless data transfer in body area networks as well as over wider area networks. Mobile phones have become one possible platform for wearable systems (Amft & Lukowicz 2009). Personal health records are being developed to be an Internet-based set of tools that people can use to access and manage their life-long health information (Tang et al. 2006). The personal health record is a central element in a whole new ecosystem of services that cooperate with the patient as co-producers of health (Saranummi 2009). The components of an ecosystem are made interoperable through international standardization that is being specified, for example, in the Continua Health Alliance (Carroll et al. 2007).

Different sensors, for instance, stand-alone, embedded and wearable, are available on the market today and a lot of data are being acquired by individuals. However, all the acquired data are not yet utilized efficiently. Utilization of the data comprises an enormous task of data analysis and data interpretation. The long-term data measured in uncontrolled environments by individuals themselves are not something that can be easily interpreted. Thus, there is a need to develop new data analysis methods for personal health monitoring data. The long-term personal health monitoring improves outcomes and reduces health care resource utilization (Cleland et al. 2005, Celler et al. 2003). The acquired data can be useful for both the patient and the medical professional, as they can be used in health and disease management as well as in prevention.

The studies on automatic recognition of physical activities and energy expenditure are commonly based on signals obtained from wearable sensors. The goal of automatic recognition of physical activities and automatic assessment of energy expenditure is to show people current distribution of their daily activities, level of their energy expenditure, and to motivate people to a more active lifestyle. The rough estimates of daily energy expenditure provided by simple activity monitors, such as the pedometers, can be further improved by using carefully selected sensors and by placing them on well-chosen points on the body, where they can be carried unobtrusively, without disturbing the user. Physical fitness has many components, which require different types of exercise: endurance can be enhanced by long-duration activities, balance can be improved, for example, by playing ball games, and muscle strength can be improved by exercising at the gym. In order to estimate a particular user's need for different activities, a more advanced analysis of user's physical activities is needed than what is available from simple activity monitoring devices. Continuous, automatic monitoring of daily physical activities and energy expenditure could provide more advanced information that can be used to promote a more active lifestyle with a wider spectrum of different exercises.

The advanced studies on mental load use data from both self-assessments as well as automatic measurements and self-measurements. No gold standard exists for the objective measuring of mental load. Clinical studies on mental load make use of, for example, salivary samples and blood tests. In everyday, continuous use, these are not applicable. Using a combination of several wireless sensors and self-assessment tools, a new type of data can be collected, which can allow studying new measures for the objective quantification of level of mental load unobtrusively and in the long-term. Supporting the interpretation of the long-term self-measured data might also make the health care professionals more willing to use this information in the future.

This thesis contains publications on the author's research in the years 2003 to 2010 on the analysis of sensor data for personal health management. This covers applications for automatic recognition of physical activities, automatic assessment of energy expenditure, and automatic recognition of mental load and stress.

2. Background and Literature Review

The analysis of personal health monitoring data measured in uncontrolled environments requires several steps in order to guarantee good quality data and the reliable interpretation of data. A human uses his/her senses every day, for example to recognize faces or traffic signs. This is an example of pattern recognition in humans. The senses acquire data from the target objects and the brain performs the classification. An automatic pattern recognition system takes raw measurement data and classifies patterns automatically. The general design cycle of an automatic pattern recognition system is shown in Figure 1. The design of automatic pattern recognition systems for the analysis of personal health monitoring data follows the general design cycle.

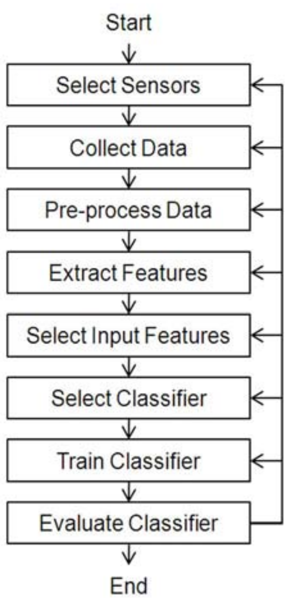


Figure 1. Design cycle of a pattern recognition system ((Duda et al. 2001), modified).

Once a target for measurements has been set, and a physical quantity, the measurand, has been selected, sensors that can measure the target well are selected. Next, the aim is to collect a data set that contains a representative variation of the data in different circumstances. The data are used for training and the evaluation of the system. The data collection is a laborious task, the cost of which can be easily underestimated. The raw data obtained from sensors may contain artefacts that have to be removed at the signal pre-processing stage. Pre-processing may also involve the synchronization of data from different sources, resampling, scaling the sensor outputs and signal filtering.

Next, features are extracted from the pre-processed signals. Sliding window is the most commonly used technique for extracting features from signals. Only the part of the signal which is within the window is processed at a time. Signal characteristics, for example the mean, are computed for the window contents. Then, the window is slid to its next location on the signal. Both block windows, that is, the windows next to each other, and overlapping windows can be used.

Feature selection aims at selecting the most information-rich subset of features to be used as inputs to the classifier. The different computed features of the pattern are components of a feature vector in a feature space. For classification, features that maximize the inter-class distance and minimize the intra-class variability are the best features. Both the prior knowledge and the training data are used for feature selection.

Next, the classifier and its structure are designed. Characteristics of the data influence the design. A more complex classifier is required, when the classes are not easily separable. The classifier is trained to recognize the patterns with the training data set. In supervised learning, the training data contains both the input patterns and the desired outputs for each pattern. Too complex a classifier may overfit to the training data. Such a classifier would not generalize and give satisfactory results with unseen patterns. An evaluation data set is used to evaluate the resulting classifier performance with unseen data. The evaluation may also show needs for the re-designing of some parts of the system.

Once the classifier is designed and ready for use, its components for online use can be described as in Figure 2. The physical measurand is converted to an electric signal using the sensors. The obtained electrical signal is pre-processed and the selected input features are computed and fed to the classifier, which assigns class labels to the patterns. In active use, both the accuracy and the computational complexity of each component become an issue.

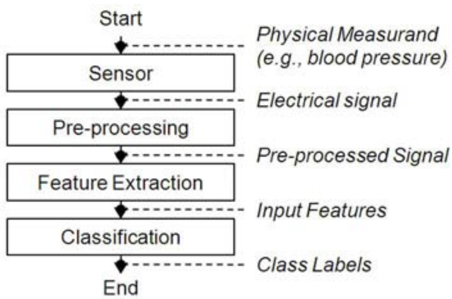


Figure 2. Components of a pattern recognition system during online recognition.

This chapter explains the requirements for sensors, central feature extraction, feature selection and classification methods used in this thesis. It also reviews the literature on the three application domains: 1) automatic activity recognition, 2) automatic assessment of energy expenditure and 3) automatic assessment of mental load and stress.

2.1 Sensors and Data Analysis Methods

2.1.1 Requirements for Sensors

A sensor converts a physical quantity to electric output. For example, a pressure sensor converts pressure to electric output. An optimal sensor responds to the physical energy available via the measurand and excludes other sources of energy (Webster 2010). Many sensors require a calibration signal for maintenance of accurate output. Many sensors also require an external power source for converting the energy from one form to another. The power consumption is critical, especially in the case of long-term monitoring using wearable sensors.

Traditionally, most of the studies on automatic health monitoring in uncontrolled environments (environments other than hospitals) focus on health monitoring at home. These solutions use mostly *embedded sensors* that are available in the home environment, either as separate devices or embedded into structures such as furniture. More recently, with advances in electronics and communication technology, *wearable sensors* have attracted increased attention. It is possible to monitor user's health status and behavior at home and during movement by combining the fixed home sensors and the wearable sensors.

Because the sensors in the home environment are used by non-professionals, certain characteristics are required of sensor systems (Korhonen et al. 2003). They should be *reliable*, *robust and durable*, because the environment of use may vary and a non-professional user may not be aware of limitations of use. The wearable sensors should be *unobtrusive* and have an *attractive design*, because they are used in various situations of everyday life. They should not limit normal activities, but rather have *attractive features that motivate to long-term use*. Preferably, the sensors should *automatically communicate* data to central data storage or have *processing power and storage space* for storing enough data for history and trend visualization. The sensor should be *self-calibrating* and have *long battery life*. The *user interaction with the device should be minimized* and the *achieved benefits should overcome the potential burden to the user*. The sensor should also be *robust against artefacts* such as movement or environmental light.

One of the key aspects of wearable sensor systems is the ability to *use the system while engaged with real-world tasks* (Amft & Lukowicz 2009). Further user needs for wearable sensors and sensor systems are *usability*, *embedded medical decision making*, and *clinical validity* (Lymberis 2004). *Usability* refers to the easy daily use, power autonomy and informative user interface. *Embedded medical decision-making* refers to algorithms that integrate collected data, and analyze trends and changes in data and suggest the best possible medical diagnosis. Today, no such algorithms are in clinical use in homecare or ambulatory devices. *Clinical validity* tells us if a device has clinical use, and whether it can accurately discriminate between normal and abnormal data.

Further requirements for psychometric monitoring are *accuracy*, *validity*, *sensitivity to change* and *incremental clinical validity* (Haynes & Yoshioka 2007). *Accuracy* refers to the degree to which the measure obtained reflects the target variable. If no gold standard measurement exists, the *degree of validity* of a measure can be estimated by estimating the degree of its co-variance with other measures. Validity can sometimes be conditional and unstable, thus, the same validity cannot be obtained in different settings (for example, subjects of different ages, a different severity of conditions, a different measurement context, etc.). *Sensitivity to change* describes the degree to which obtained measures reflect changes in a target variable, thus, whether the changes are delayed or immediate. *Incremental validity* refers to the degree to which the measure strengthens the validity of the clinical decision-making beyond the measures normally used.

With these requirements in mind, the sensor set for each study was selected. With activity recognition and estimation of energy expenditure, the focus is on wearable sensors. For mental load and stress monitoring, both wearable and embedded sensors are used.

2.1.2 Feature Extraction

Feature extraction aims at extracting such characteristics of the input patterns that enable their classification into distinct classes. In the analysis of personal health monitoring data, the features are computed from the content of a sliding window. Feature extraction methods commonly used to study signal characteristics include 1) time-domain features and 2) frequency domain features. *Time-domain features* include the mean, median, variance, skewness, kurtosis, range, means for high frequency data and low frequency data, and the integral of modulus of accelerations for estimation of energy expenditure. *Frequency domain features* include the peak frequency, peak power, spectral power on different frequency bands and spectral entropy. The central features are presented in this chapter in more details. The emphasis is on computationally less complex features that are more useful in embedded and wearable devices and in long-term monitoring rather than computationally more complex features.

2.1.2.1 Sample Mean

Mean $\bar{x}$ represents the sample mean of values within the sliding window. In equation (1), *n* denotes window length and *x[i]* denotes the i^{th} value of the signal. The sample mean of accelerometer data has been used to identify postures. This can be done, because the gravitational acceleration component shows on signal and can be used to distinguish, for example, vertical and horizontal postures.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x[i] \tag{1}$$

2.1.2.2 Sample Variance

Sample variance s^2 represents the signal variance around the sample mean within the window. It is often used to represent signal energy.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x[i] - \bar{x})^2 \tag{2}$$

2.1.2.3 Median

Median $\tilde{x}$ represents the point for which half of the distribution is higher and half is lower. Variable y denotes the sliding window contents sorted into ascending order. The median value of accelerometer data can be used for the identification of postures.

$$\tilde{x} = \begin{cases} y[(n+1)/2], \text{ if } n \text{ is odd} \\ \frac{1}{2}(y[n/2] + y[1+n/2]), \text{ if } n \text{ is even} \end{cases} \tag{3}$$

2.1.2.4 Integral Method

Integral method was designed (Bouten et al. 1994, 1997) to give an estimate of energy expenditure using a triaxial accelerometer. The abbreviation IMA_{tot} stands for the Total Integral of Modulus of Accelerations, where modulus refers to the absolute value. In equation 4, a_x, a_y and a_z denote the three orthogonoal components of accelerations obtained from the triaxial accelerometer and t denotes time. T denotes a window length and t_0 denotes a window start time. The resulting integral is influenced by both the intensity and the duration of the movement activity. Equation 5 presents the formula in a discrete form, in which N denotes the window length. Before using the integral method, raw accelerometer data is bandpass filtered (0.5 … 11 Hz) to highlight voluntary human movements and reject changes caused by posture changes and high-frequency vibrations.

$$IMA_{tot} = \int_{t=t_0}^{t_0+T} |a_x| dt + \int_{t=t_0}^{t_0+T} |a_y| dt + \int_{t=t_0}^{t_0+T} |a_z| dt \tag{4}$$

$$IMA_{tot} = \sum_{i=1}^{N} |a_x[i]| + \sum_{i=1}^{N} |a_y[i]| + \sum_{i=1}^{N} |a_z[i]| \tag{5}$$

2.1.2.5 Peak Frequency

The frequency spectrum of discrete signal *x* is computed using the Discrete Fourier Transform (DFT). The DFT is defined as (Oppenheim et al. 1999)

$$X(k) = \sum_{n=0}^{N-1} x[n]\, e^{-j2\pi kn/N} \tag{6}$$

, where *X* denotes the frequency spectrum, *j* denotes the imaginary unit, *k* denotes the k^{th} Fourier coefficient in the frequency domain, *n* denotes the variable index, and *N* denotes the length of the sliding window. The frequency spectrum determines how much of each frequency component is required to synthesize the original time-domain signal *x* using complex sinusoidal components. The DFT can be efficiently computed using the Fast Fourier Transform (FFT) algorithm (Cooley & Tukey 1965).

The frequency spectrum is squared to obtain the Power Spectral Density (PSD) function $P(f_i)$

$$P(f) = \frac{1}{N}\, X(f)\, X^{*}(f) = \frac{1}{N}\, |X(f)|^2 \tag{7}$$

, where N denotes the number of frequency components obtained from the DFT.

Peak frequency is the frequency with the highest power of the computed power spectrum. The peak frequency of accelerometer data describes the dominant frequency of the activity. For example, if the sliding window contains accelerometer data from running, the frequency is higher than that from walking, because of a higher step rate during running.

2.1.2.6 Peak Power

Peak power is the maximum power in the PSD. It reflects the most prevalent sinusoidal component in the signal.

2.1.2.7 Spectral Entropy

As a physical concept in thermodynamics, entropy is proportional to the logarithm of the amount of microstates in a system and thus measures the amount of disorder in the system. In information theory, the entropy *H* of a random variable *X* was defined as (Shannon 1948)

$$H(X) = -\sum_{i=1}^{N} p(x_i) \log_2 p(x_i) \tag{8}$$

, where x_i is the state *i* of variable *x*, *N* is the number of states and $p(x_i)$ is the probability of state x_i. In addition to a time-domain, entropy has been computed also for frequency domain signals (Johnson & Shore 1984). This has been named the spectral entropy and it describes the irregularity of the signal. The spectral entropy has also been successfully used in activity recognition (Bao & Intille 2004, Lester et al. 2006). A method for the efficient computation of spectral entropy by combining time and frequency domain approaches was developed for anesthesia monitors (Viertiö-Oja et al. 2004).

The spectral entropy S_N is defined for a frequency band *[f1, f2]* as

$$S_N(f_1, f_2) = \frac{-\sum_{f_i=f_1}^{f_2} P(f_i)\log(P(f_i))}{\log(N_{f_1,f_2})} \tag{9}$$

, where $P(f_i)$ denotes the PSD component of frequency f_i. The entropy value is normalized to range between 0 (complete regularity) and 1 (maximum irregularity) by dividing the sum by $log(N_{f1,f2})$, where *N* refers to the number of frequency components in the defined frequency band.

2.1.2.8 Histogram Transformation

With many biomedical signals, the inter-individual variability is large, making it impossible to use common thresholds for classification. Histogram transformation is a method developed to transform the individual HR values to a common range from 0 to 100. The method was developed for the detection of surgical stress with people undergoing a surgery during general anaesthesia in order to control the amount of anesthetic medication during the surgery (Huiku et al. 2007). The histogram transformation combines HR data from two distributions, the individual and the group HR distributions into one distribution using a weighted sum. The combined distribution is used to form a cumulative sum function. The function is then used to transfer the individual HR values to the common range. Figure 3 shows an example of the method.

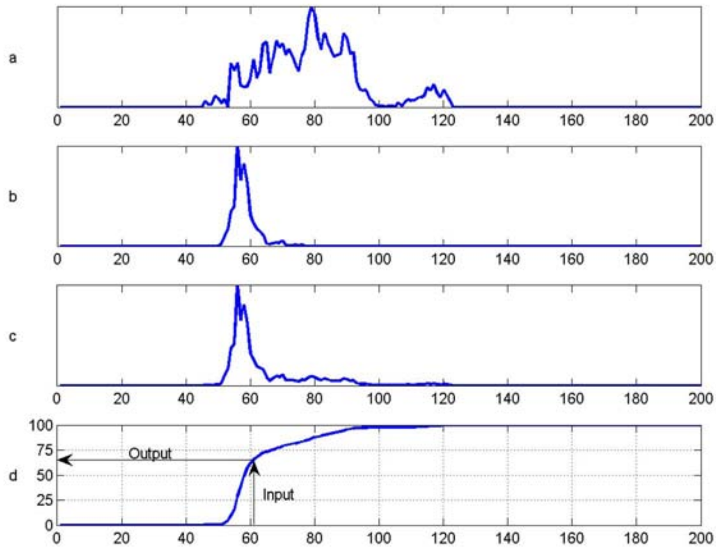


Figure 3. Histogram Transformation normalizes the HR with respect to the individual and group distributions: a) group distribution, b) individual distribution, c) combined distribution and d) the cumulative sum function. The x-axis represents the HR [bpm].

2.1.3 Feature Selection

Pattern recognition requires representative features, thus features that can be used to discriminate between patterns. If a classifier, using just a couple of features, does not provide accurate classification results, it is common to use more input features in the classifier. This often helps, but the number of input features is limited; after this limit the performance of the classifier starts to decrease. This phenomenon is called “the curse of dimensionality”. The performance degrades due to the fact that there are not enough training data to train the classifier well in the higher dimensional space. The number of training samples must grow exponentially with the number of input features. Thus, it is necessary to select a set of features to be used as inputs to the classifier. The collection of selected features is called a feature set. For effective classification, it is important to find features that have the optimal discriminative power between the classes. In case of activity recognition, well-selected features have very little variation between subjects and repetitions of the same activity, but they show large changes between different activities at the same time (Preece et al. 2009). The classification using a large feature set with heavily correlated or otherwise unnecessary features requires also more computational power and slows down the classification process.

Human eye is good at selecting features that show large changes between different activities. Although visual inspection may be an accurate method for finding the best features, it may be too laborious with large feature sets. Therefore, different methods have been developed for creating an optimum input feature set for classifiers. Methods referred to as dimensionality reduction create new features by combining the raw features (for example linear combination). Feature selection methods aim at selecting the best feature subset from the original features for classification.

Sequential Forward Search (SFS) (Whitney 1971) was one of the first feature selection methods and is widely used for feature selection also today. The SFS method starts with an empty feature set and it adds one feature, the one that alone has the highest classification accuracy. In next rounds, one feature, the one that together with the already selected features produces the highest recognition accuracy is selected. The process stops, when the given number of features is selected or no improvement is achieved by adding more features.

2.1.4 Classification

Classification aims at recognizing and assigning class labels to input patterns. The input patterns consist of the features computed from the input data. The classifiers can be divided into two classes: supervised and unsupervised classifiers. Unsupervised classifiers take the input data and search for clusters in data, without knowledge of the true class labels of the input patterns. Supervised classifiers are trained with the help of training data and training algorithms. In the training phase, a model of the training data is built for use in the classification phase. Many classifiers use training data to define decision borders in feature space. The decision borders divide feature space into distinct regions that represent the classes. A generalizable decision border is able to achieve a good accuracy of classification not only with training data, but also with unseen input patterns. An overly complex decision border overfits to training data and does not generalize well to unseen patterns. Training can be done offline, for example, on a PC, even if the classification phase takes place in an embedded device. In the classification phase, the aim is to use the trained classifier and recognize unseen input patterns effectively and with good accuracy.

The classifiers used with health-related data are the same as those generally used in machine learning and pattern recognition. Supervised methods are commonly used, because they find exact matches with real-world classes. This is not always the case with unsupervised methods. For long-term monitoring with wearable and embedded devices, the primary focus is on finding practical, computationally efficient methods. Classifiers with the efficient classification phase are preferred in particular. Training is supposed to be done elsewhere, not in the embedded device. Computationally more complex classifiers are used as references and to find which classifiers have the best performance in this application area. According to published literature (Duda et al. 2001, Lippmann 1989), the K-Nearest Neighbor classifier is computationally most demanding during the classification and decision trees are the computationally most efficient classification algorithms. Figure 4 shows the memory requirements for the most commonly used classifiers.

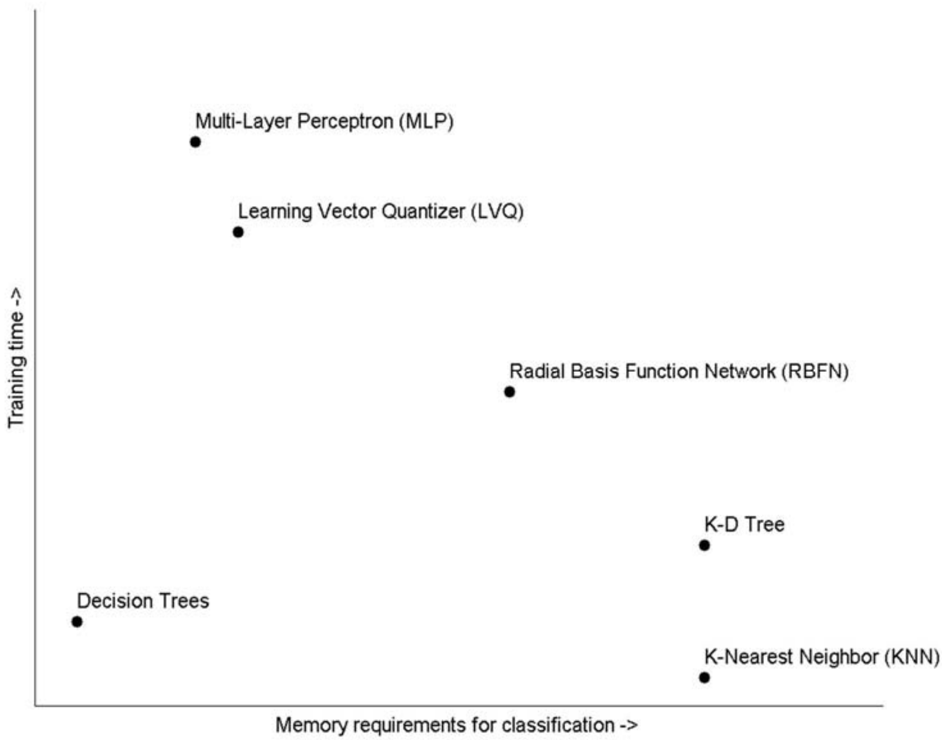
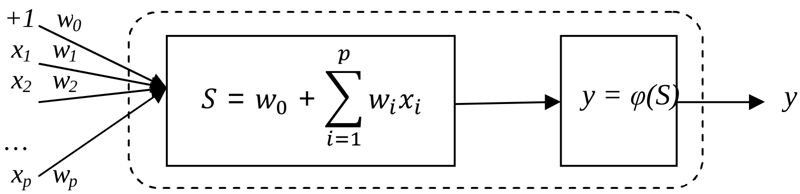


Figure 4. Classifier memory requirements versus training time ((Lippmann 1989), modified). Decision trees, Multi-Layer Perceptrons and K-Nearest Neighbor algorithms were used in the studies for this thesis.

2.1.4.1 Artificial Neural Network

From the class of classifiers known as Artificial Neural Networks (ANN), Multi-Layer Perceptron (MLP) network is one of the most widely used classifiers. An MLP is a feed-forward network that consists of artificial processing units (“neurons”) and connections between the neurons. The processing of one neuron is shown from inputs to one neuron’s output in equation (10)



(10)

, where x_1 … x_p denote the input features, w_0…w_p denote the weights of each input, S denotes the weighted sum of inputs and bias (*+1*), y denotes output and φ denotes the activation function (Haykin 1999, Lehtokangas 1995). The input signals are multiplied by weights and added to form the activation signal S, which is in turn fed to the activation function φ. The activation functions commonly scale the activation signal to some range. For example, the *tanh*-function scales the activation signal to range -1…1 and a logistic function scales the activation signal to range 0…1. The logistic activation function can be expressed as:

$$y = \frac{1}{1 + e^{-S}} \tag{11}$$

A multilayer perceptron (MLP) consists of several layers of such neurons. An example of a typical MLP structure is shown in Figure 5. The MLP in Figure 5 has an input layer, one hidden layer and an output layer. The input layer consists of 4 input neurons, a hidden layer of 5 hidden neurons and an output layer of 3 output neurons. No computations occur in the input layer. The neurons of the hidden and output layers contain the computations described in equation (10).

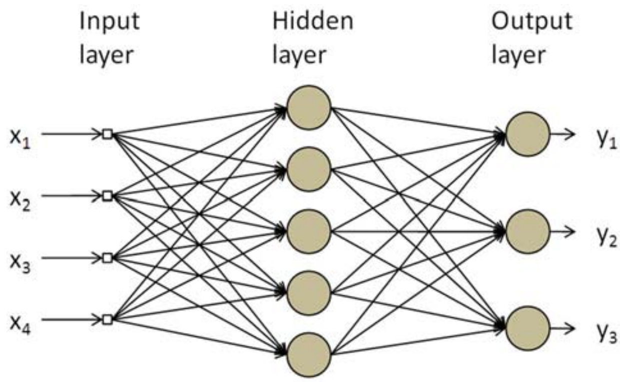


Figure 5. An example illustrating the multilayer perceptron (MLP) structure.

Training algorithms are used to learn from training data and build the knowledge into the ANN structure. MLP training is a supervised process, where each input pattern is accompanied with the desired output class. The gradient descent methods can be used to train the network. The Rprop training method (Riedmiller & Braun 1993) is especially suitable with large amount of training data. The Rprop training is computationally more efficient than that of, for example, the basic back-propagation algorithm. An MLP can be used to define nonlinear decision borders and thus very complex decision regions. One of the most difficult problems in constructing an MLP is to define a network of the suitable structure and size. If too complex a network is built, it slows down the training and classification phases and the resulting network does not generalize well for the use with new, unseen patterns. However, if too small a network is built, it cannot find the necessary decision regions for proper classification.

Although the training phase of an MLP is computationally demanding, the classification using an MLP is a straightforward process. After training the weights are fixed and classification requires only the operations described in equations (10) and (11) for each neuron. Thus, the structure is well applicable to long-term monitoring with battery-powered devices.

2.1.4.2 Decision Trees

Decision trees are widely used classifiers of several benefits to their long-term use in health and behavior monitoring: they are easy to interpret and a priori knowledge can be naturally incorporated (Duda et al. 2001). They also perform efficient classification compared to other classifiers (Figure 4) (Duda et al. 2001, Lee & Lippmann 1990). This is because they require only a few numerical comparisons for classification. Decision tree classification can be thought of as a series of questions, in which the next question depends on the answer to the previous question. After a series of questions, a class label is assigned to the sample. The decision tree decision boundaries in feature space form rectangular decision regions (Figure 6).

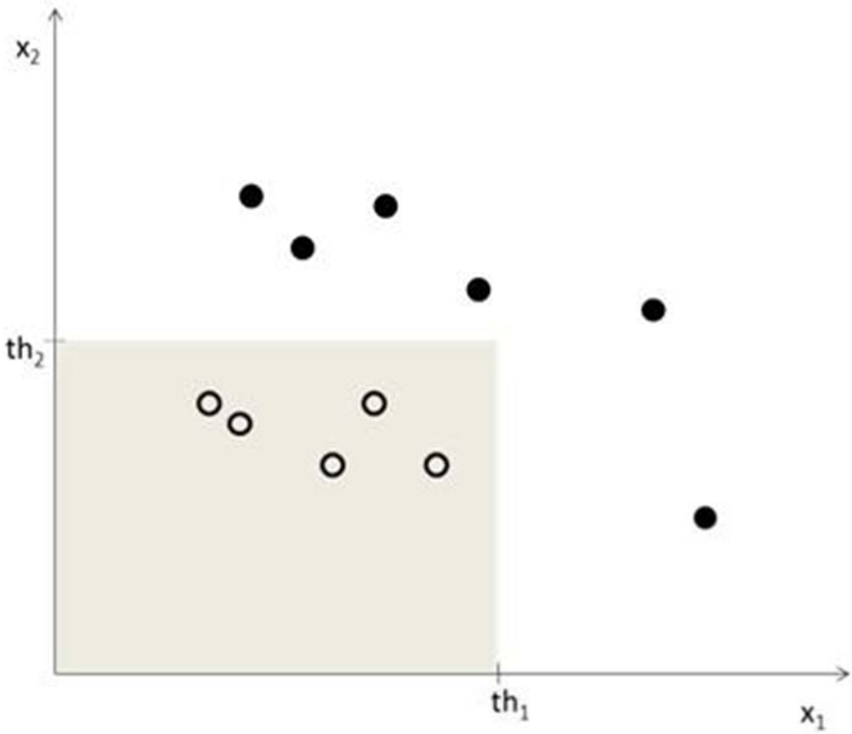


Figure 6. Binary decision tree decision regions in feature space. The classes are recognized by two features x_1 and x_2 and decisions $x_1 < th_1$ and $x_2 < th_2$, where th_1 refers to the threshold value of feature x_1.

A decision tree consists of *nodes*, *branches* and *leaves*. The first node is called the *root node* and it performs the first split of dataset into subsets. In each node, a question is asked (for example, is variance > 0.5?), which divides the input dataset into subsets. Depending on the answer, one of the branches is followed to the next node. Every decision tree can be represented using binary decisions (Duda et al. 2001). In binary decision trees, a dataset is divided into two subsets in each node. For example, the right branch is followed, when the sample value is > 0.5 and the left branch is followed when the sample value is <= 0.5. Terminal nodes are called leaves. In a leaf node, a class label is assigned to a sample. All samples arriving to the same leaf will get the same class label. When talking about subsequent nodes, the node closer to the root node is called the *parent node* and the node closer to leaves is called the *child node*.

The decision trees can be "custom", thus manually generated using a priori information and features selected after human reasoning, or they can be automatically generated with the help of training data. In automatic generation of a decision tree, the aim is to divide a dataset into as pure subsets as possible. The Gini impurity index is one of the most widely used methods for measuring the impurity of nodes. When using the Gini impurity index, datasets are divided into smaller subsets using divisions that cause the maximal decrease in impurity compared to that of a parent node. The Gini impurity is defined as

$$i(N) = \sum_{i \neq j} P(\omega_i)P(\omega_j) = \frac{1}{2}\left[1 - \sum_j P^2(\omega_j)\right] \tag{12}$$

, where $i(N)$ denotes impurity at node N, $P(\omega_j)$ denotes fraction of patterns at node N that are of class ω_j (Duda et al. 2001). This can be interpreted as the expected error rate at node N. The target Gini impurity is naturally 0 and it is reached, when all patterns belong to the same class: $i(N) = (1\text{-}1^2)/2 = 0$. Maximum impurity is reached when all classes are equally probable. For two-class case, the maximum impurity is: $i(N) = (1\text{-}0.5^2\text{-}0.5^2)/2 = 0.25$. The decrease in impurity from a parent node to a child node is defined as

$$\Delta i(N) = i(N) - P_L i(N_L) - (1 - P_L) i(N_R) \tag{13}$$

, where $i(N_L)$ and $i(N_R)$ are the impurities of left and right child nodes, and P_L is the fraction of patterns at node N that will go to the left branch, when one selected feature is used as a dividing criterion. When impurity decreases are computed for all feature candidates, the feature with the largest impurity decrease is selected for node N.

The decision tree obtained in the automatic decision tree generation is often overly complex and easily overfits to training data (Witten & Frank 1999). Thus, different pruning methods have been developed to reduce the size of such a tree. Pruning reduces complexity of the tree and improves its generalizability. In post-pruning, a complete tree is first generated and then pruned. Prepruning is used during the tree generation process to decide when to stop creating new subtrees. Most of automatic decision tree generation methods use postpruning, because they tend to find better combinations of features, and thus better performing trees. For example, this is the case when two separate features do not improve classification performance, but combined together they improve the performance.

2.1.4.3 K-Nearest Neighbor

The K-Nearest Neighbor (KNN) classifier (also called the “instance-based” classifier) is an intuitive method that classifies patterns based on their similarity to patterns in a training set. The method is also called ”lazy” learning, because in the training phase of the basic KNN algorithm, all training patterns are just stored for comparison in the classification phase and all computation is done during the classification phase. Because of these characteristics, the method is not efficient in the classification phase: it requires a lot of memory and a lot of computations compared to model-based methods that discard training patterns after the model creation. However, the accuracy of the KNN classifier is good when training data are representative and large enough (Duda et al. 2001). For this reason, it is often used as a reference classifier.

The algorithm requires only one parameter, the parameter K, which determines how many nearest neighbors are taken into account when comparing an unlabeled pattern to training patterns. In the case of K = 1, the method is simply called the “Nearest Neighbor method” or ”1-NN method”. When computing the distance from an unlabeled pattern to training patterns, different distance metrics can be used. The Euclidean distance is most commonly used

$$d = |\boldsymbol{x} - \boldsymbol{y}| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{14}$$

, where $\boldsymbol{x}$ is the unlabeled pattern in an n-dimensional feature space and $\mathbf{y}$ is a training pattern. To make the Euclidean distance work properly, the features have to be normalized. With 1-NN, an unlabeled pattern is assigned with the same class label as a training pattern with the smallest distance to the unlabeled pattern. The decision regions obtained in this procedure are shown in Figure 7. Such decision regions are also called the Voronoi tessellation. All points falling into the same cell are assigned the same class label.

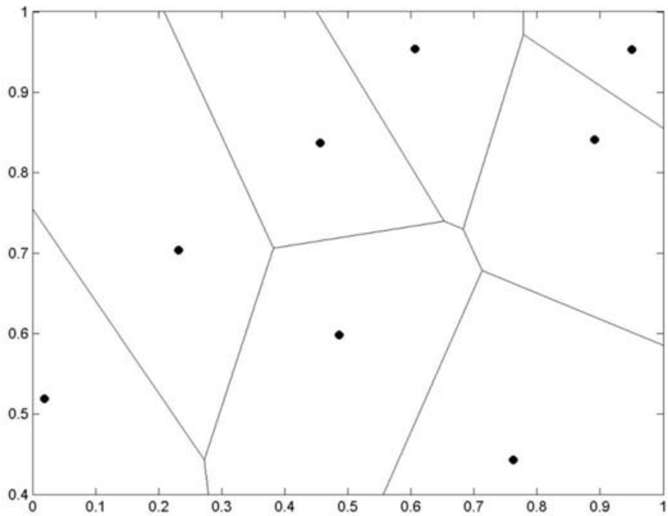


Figure 7. Decision regions for 1-NN classifier. The points represent training patterns and the lines describe the decision borders around each training pattern. An unlabeled pattern will be classified as the same class as a training pattern that is closest to the unlabeled pattern. The regions are also called the Voronoi tessellation. In three dimensions the regions become crystal-like regions.

With the selection K = 1, the classifier easily overfits to training data. By selecting a larger K, K nearest neighbors to the unlabeled pattern are searched and the class that appears most often among the K nearest neighbors is assigned to the unlabeled pattern. Odd K is preferred over even K to avoid a tie. The selection depends on the amount of training data and a possible overlapping of classes. Larger values of K achieve better generalization than small values of K. Large values of K also give probabilistic information of the decision, and thus obtain a more reliable estimate for the unlabeled pattern, but small values of K concentrate the search on neighbors closer to the unlabeled pattern. Thus, the selection of K is a compromise between obtaining a reliable estimate and localizing the search on suitably small region in feature space (Duda et al. 2001).

2.2 Activity Recognition

Today, physical inactivity is part of a normal lifestyle in industrialized countries. Level of physical activity required at work, to travel and at home is decreasing with sedentary work and technologies designed to ease home activities and traveling. Leisure physical activities are insufficient or too irregular to achieve an adequate level of physical activity per week. At the same time, the chronic diseases are becoming more and more prevalent. It has been shown that physical inactivity contributes to many chronic diseases, such as cardiovascular disease, type 2 diabetes, and possibly certain types of cancer and osteoporosis (Pate et al. 1995, WHO 2010).

2.2.1 Physical Activity Recommendations

Physical activity recommendations have been created to increase awareness of adequate levels of physical activity with a view of gaining health benefits. There are different recommendations for different target groups (such as adults, older adults, chronic disease groups) and for different ambition levels (to maintain good health, to get more health benefits).

The minimum recommendation of the World Health Organization (WHO) in respect of health maintenance is 30 minutes of moderate-intensity physical activity per day (Puska et al. 2004). This can be achieved through everyday activities like walking to work, shopping, gardening, cleaning, etc. The recommendation aims to achieve at least 1000 kcal energy expenditure per week. The only limitation is that the 30-minute period must be performed in continuous periods of minimum 10 minutes.

According to the WHO, at least 60% of the world's population fails to achieve the minimum recommendation (Puska et al. 2004). In a recent study (Chastin et al. 2009), compliance with physical activity recommendations was examined on a group of 78 postal workers in the UK. Only 10% of the participants succeeded in complying with the 30-minute-daily recommendation.

Physical activity guidelines for Americans (Leavitt 2008) were published to promote physical activity in the USA. The guidelines recommend 60 minutes of physical activity per day. The recommendation also emphasizes that the daily physical activity should include: 1) aerobic exercises, 2) muscle-strengthening exercises, and 3) bone-strengthening exercises.

The most recent recommendation "Global Recommendations on Physical Activity for Health" published by the WHO (WHO 2010) no longer advocates daily exercise, but weekly goals. The recommendation includes three targeted guidelines: 1) for children of 5 to 17 years of age, 2) for adults of 18 to 64 years of age, and 3) for older adults of 65 years of age and above. The new recommendation puts more emphasis on vigorous-intensity exercise. In addition, it recommends activities increasing muscle strength and bone strength.

The Finnish Current Care recommendations (Kesäniemi et al. 2010) contain physical activity guidelines for healthy adults as well as for adults suffering from different illnesses. Essentially, the Finnish recommendations are similar to the international ones, but adapted for the Finnish society. The Physical Activity Pie is a popular graphical visualization of the Finnish recommendations (Fogelholm et al. 2005).

2.2.2 Physical Activity Monitoring

Traditionally, the level of physical activity has been assessed using questionnaires or diaries. These are useful and cost-efficient methods for obtaining rough estimates of physical activity in large populations. However, for obtaining more objective estimates, direct measurements are needed.

By definition, physical activity is any bodily movement that results in energy expenditure and that is produced by skeletal muscles (Caspersen et al. 1985). Thus, it covers many types of physical activity, including both natural physical activity and intentional exercise. In order to adequately assess the activity profile of a person, both the recognition of energy expenditure and the types of activities are needed. Energy expenditure gives an overview, indicating whether the person performs enough physical activity. Activity type recognition enables profiling different categories (aerobic, muscle-strengthening and bone-strengthening activities). The need for activity and posture recognition has recently been identified also in studies comparing time spent in sedentary activities and cardiac risk factors. It has been found that reducing sedentary time and breaking long sedentary periods into several shorter ones reduce the cardiac risk factors (Healy & Owen 2010).

The first drawings of devices measuring steps and distance can be found in the drawings of Leonardo da Vinci (Gibbs-Smith 1985, Tudor-Locke 2003). Leonardo designed a pendulum-type pedometer, but as it did not work well for measuring distance, he designed an odometer. The odometer was a simplification of an ancient Roman machine, designed by the architect Vitruvius. It was a wheelbarrow-like device. The circumference of a wheel was known and the device dropped pellets to a box every few turns. The distance was estimated by calculating the number of pellets in the box. For example, the device was used to create maps.

A pedometer and the Manpo-kei program (or “10,000 steps meter”) in Japan in 1965 (Tudor-Locke 2003) were one of the first wearable methods intended for objectively measuring physical activity. The program used the pedometer developed by professor Hatano and aimed at reaching a daily goal of 10,000 steps. The program was a success in Japan and continued for decades there. A personalized version of the Manpo-Kei program was brought to the USA (Tudor-Locke 2003). This modified program highlights the importance of natural physical activity instead of intentional exercise and questions the 10,000 step goal for everyone. Instead it emphasizes the need to increase the amount of steps from the current level. At the beginning of the program, the baseline is measured over one week using a pedometer and the goal is defined individually for each person.

The pedometers used in the original Manpo-Kei program were based on a mechanical lever that deflected with vertical oscillations. Thus, the pedometer had to remain in a fixed position to detect steps correctly. Therefore, the first pedometers were attached using a belt clip on a waist, above the dominant leg. They were accurate in detecting steps during brisk walking and running, but could not detect steps accurately during slow walking. Although the first pedometers estimated also distance (the length of an individual's step had to be entered) and energy expenditure (required specifying weight, age and gender), these additional functions were not very accurate. Distance measurement did not work well with variable lengths of steps and with surfaces other than a flat one. Energy expenditure estimate often failed because of the fact that the pedometer did not measure the intensity of the activity, but the number of steps.

Today, electrical accelerometers that sense both the movements and their intensity are used. Despite the different construction, electrical accelerometers are also based on the spring-mass principle. In this system, a small mass is attached to a spring inside an accelerometer. When acceleration is applied to the mass, the spring either stretches or compresses. The displacement can be measured and used for computation of the applied acceleration. At present, the most popular accelerometers require an external power supply, but thanks to that, they respond to both accelerations caused by body movements and the static gravitational acceleration (Mathie et al. 2004). When the sensor is kept motionless, the resulting output signal is, in the case of a 1D accelerometer, a projection of the gravitational acceleration to the direction of the sensitive axis. With 3D accelerometers, the resultant vector of all three sensitive axes can be computed, representing the direction and magnitude of the acceleration. In movement, the measured acceleration is the vector sum of gravitational and movement accelerations. Artefacts are caused to the signal, for example, by a loosely attached sensor or by external accelerations, such as motor vibrations when traveling in a vehicle. The amount of artefacts can be minimized by a careful placement of the sensor and by signal filtering.

Table 1. Common accelerometer locations and applications ((Mathie et al. 2004), modified).

Location	**Application**	**References**
Ankle, thigh	Leg movement during walking	(Lafortune 1991, Bussmann et al. 2000, Aminian et al. 1999)
Ankle	Quantification of activities in stroke rehabilitation	(Hester et al. 2006)
Wrist	Parkinsonian bradykinesia	(Veltink et al. 1995)
Wrist	Activities of daily living	(Yang et al. 2008)
Arms, legs	Parkinsonian tremor	(Van Emmerik & Wagenaar 1996)
Ear	Quantification of activity and recovery from surgery	(Lo et al. 2007)
Chest	Coughing	(Fukakusa et al. 1998)
Waist	Detection of activity/rest	(Mathie et al. 2003, Karantonis et al. 2006)
Multiple instruments	Whole body movements	(Bao & Intille 2004, Fahrenberg et al. 1997, Foerster & Fahrenberg 2000, Uiterwaal et al. 1998, Veltink et al. 1996)
Center of mass (within pelvis)	Whole body movements	(Bouten et al. 1997, Smidt et al. 1971, Sekine et al. 2000)

The frequency range of voluntary human activities is between 0.3 and 3.5 Hz (Sun & Hill 1993). The use of 0.5 Hz … 11 Hz band-pass filters was suggested to reduce gravitational artefacts and allow recording of faster movements that occur in younger subjects (Van Someren et al. 1996).

The dynamical range required for the measurement depends on the application and measurement site. Generally, the accelerations measured during everyday activities like walking and running are largest from feet and smallest from the head (Mathie et al. 2004). Running produces vertical accelerations of 8.1g, and cycling produces 2.2g when measured from the ankle (Woodward & Cunningham 1993).

Accelerometry is used in many application areas. Table 1 shows examples of applications where accelerometers have been used. The most common application areas were summarized as follows (Mathie et al. 2004):

- longitudinal measurement of activities (activity recognition, functional status monitoring of an older adult or of a person in rehabilitation)

- estimation of long-term energy expenditure (EE profile of a day)
- circadian rhythms (sleep-wake patterns, EE estimation)
- event detection (falls, etc.).

Algorithms utilizing accelerometer data have most extensively focused on the detection of energy expenditure using a single accelerometer. However, with recent developments on electronics and sensors, studies on the use of several accelerometers and several other types of sensors together with accelerometers to detect the energy expenditure and activity type have been given more attention. According to the physical activity recommendations, it is possible to increase health benefits by performing, physical activities that improve 1) endurance, 2) bones and muscles, and 3) balance in addition to normal daily physical activity such as steps. Thus, methods are needed to profile human daily activity more accurately. Accelerometry is a suitable method for measurements in uncontrolled environments, because accelerometers have small size, light weight, long battery life and they accurately measure accelerations as a function of time.

Multisensory approaches as well as single-sensor approaches have been studied for activity recognition. Generally, multisensory approaches have produced more accurate results than those done with a single or just a few sensors. Bao and Intille (Bao & Intille 2004) compared the results obtained with a single sensor, a combination of two sensors and a combination of 5 sensors. They studied the automatic recognition of 20 everyday household activities in semi-supervised settings using 5 stand-alone data loggers with 20 subjects. The accelerometers were placed on ankle, thigh, hip, arm and wrist. The best results with a single sensor were obtained using the *thigh* sensor (-29%, as compared with the result obtained using 5 acceleration sensors), the second best with the *hip* sensor (-34%). The best results using a combination of two sensors were obtained with *thigh and wrist* sensors (-3%), second best with *hip and wrist* sensors (-5%). Generally, adding more sensors increases accuracy, but already with two sensors, one on the lower body and one on the upper body, accurate results can be achieved. The overall accuracy in this study was 84% and it was obtained using a decision tree classifier (Bao & Intille 2004). Table 2 shows examples of classifiers used for activity recognition and the accuracies obtained.

Table 2. Summary of studies that have compared different classifiers for automatic activity recognition. SVM stands for Support Vector Machine classifier, HMM stands for the Hidden Markov Model and GMM for the Generalized Markov Model ((Preece et al. 2009), modified).

Publication	**Sub-jects**	**Activ-ities**	**Activities**	**Accelerom-eter place-ments**	**Classification Accuracy**
(Bao & Intille 2004)	20	20	Walking, sitting, cycling, running, vacuuming, folding laundry, etc.	Shank, thigh, upper arm, wrist and hip	Decision tree (84%), kNN (83%), naïve Bayes (52%)
(Maurer et al. 2006)	6	6	Sitting, standing, walking, ascending stairs, descending stairs, running	Wrist	Decision tree (87%), Naïve Bayes (< 87%), kNN (< 87%)
(Pirttikangas et al. 2006)	13	17	Typing, watching TV, drinking, walking upstairs, cycling, etc.	Both wrists, thigh and necklace	ANN (93%), kNN (90%)
(Ravi et al. 2005)	2	8	Standing, running, sit-ups, vacuuming, brushing teeth, walking, etc.	Waist	Naïve Bayes (64%), SVM (63%), decision trees (57%), kNN (50%)
(Lester et al. 2005)	2	10	Walking, driving, jogging, ascending and descending escalator	Shoulder	Naïve Bayes (67%), HMM (47%), HMM and binary classi-fiers (95%)
(Allen et al. 2006)	6	8	Sitting, standing, lying, walking and four postur-al transitions	Waist	GMM (91%), decision tree (71%)
(Könönen et al. 2010)	12	9	Cycling, playing football, lying, Nordic walking, rowing, running, sitting, standing, walking	Hip and wrist	SVM (79%), kNN (77%), minimum distance classifier (73%)

2.3 Assessment of Energy Expenditure

The total daily energy expenditure (EE) of a human consists of 1) resting metabolic rate, 2) thermic effect of feeding and 3) posture, spontaneous and voluntary physical activity (Lagerros & Lagiou 2007). The resting metabolic rate is estimated to be between 60 and 75% of the total energy expenditure (Figure 8), and is reasonably constant from day-to-day and also between different individuals. The thermic effect of feeding represents the energy required by the body to digest, absorb, etc., after eating a meal. It is estimated to represent about 10% of the total EE. The component that influences the variation of the total EE the most is physical activity. It is estimated to represent 15 to 30% of the total EE.

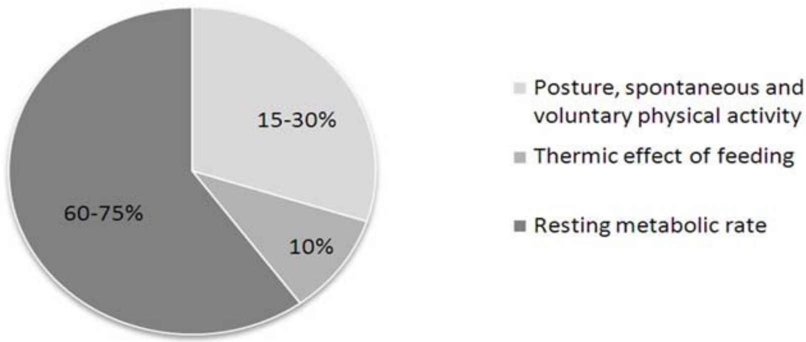


Figure 8. Components of the Total Energy Expenditure (Lagerros & Lagiou 2007).

Strenuousness of physical activity is referred to as intensity. The most common way to measure the intensity of physical activity is to use the metabolic equivalent or metabolic energy turnover (MET). MET is a measure, which tells the intensity of physical activity with multiples of a resting metabolic rate. Thus, 1 MET refers to the resting metabolic rate. This is traditionally measured during quiet sitting. Walking has an intensity of about 4 MET and jogging about 7 MET. Thus, the intensity of such jogging consumes 7 times the resting metabolic rate. The compendium of physical activities and their MET values was first published in 1993 and updated later in 2000 (Ainsworth et al. 1993, Ainsworth et al. 2000).

In energy metabolism, a calorie is the commonly used unit of energy, which corresponds to circa 4.19 Joule. When the intensity and duration of an activity as well as the weight of the subject is known, the energy expenditure can be computed by the multiplication of intensity [METs], duration [hours] and body weight [kilograms] (Figure 9). For example, for a person with a weight of 80 kilograms, jogging (7 MET) for 1 hour requires approximately 560 kcal. For comparison, one chocolate bar of 50 grams contains 200–250 kcal of energy. Stable body weight requires that energy intake and energy expenditure be balanced.

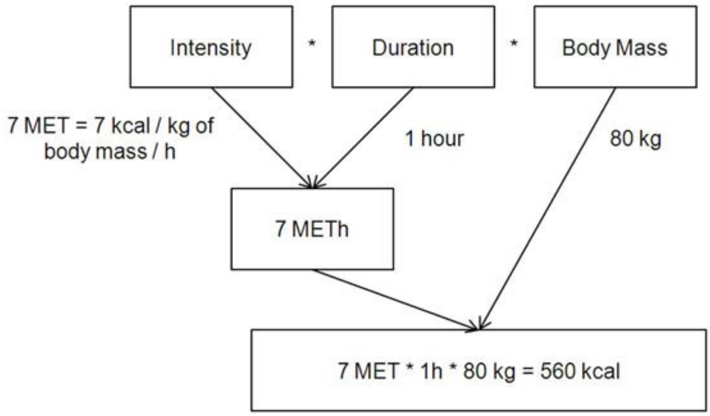


Figure 9. Energy Expenditure of an 80kg person jogging with the intensity of 7MET for 1 hour. Based on (Lagerros & Lagiou 2007).

The MET estimate standardizes the measurement of physical activity, but does not take into account varying conditions. The varying conditions can include age or sex of subject, efficiency of activity, weather, etc. Such standardization is useful, for example, when computing EE from self-assessed surveys that do not include measurements. Physical activities have been categorized into light, moderate and vigorous activities, which are represented by intensities < 3 MET, 3–5.99 MET, and > 6 MET respectively (Pate et al. 1995, Ainsworth et al. 2000). These categories are used in the physical activity guidelines for dosing physical activities. For example, moderate intensity activities have been found most helpful in increasing energy expenditure (Westerterp 2001). Since one activity can be light for one individual, but vigorous for someone else, the thresholds between light, moderate and vigorous activities can vary according to the conditions. For instance, walking is light physical activity for a healthy person in good condition, but it is a vigorous activity for someone with cardiac problems. It is possible to find people, who can increase their energy expenditure even to 100-fold above the resting metabolic rate for very short durations (Bouchard et al. 2007). Table 3 shows thresholds for middle-aged adults whose maximal oxygen consumption is either good (> 12 MET) or bad (< 5 MET) (Kesäniemi et al. 2010).

Table 3. Classification of MET intensities into light, moderate and vigorous. The values are computed for middle-aged adults whose maximal oxygen consumption is either good (> 12 MET) or bad (< 5 MET) ((Kesäniemi et al. 2010), modified)

Category of PA	**Relative Strenuousness [% of max HR]**	**Good condition [MET]**	**Bad condition [MET]**
Light	≤ 63	≤ 5.3	≤ 2.5
Moderate	64–76	5.4–7.5	2.6–3.3
Vigorous	≥ 77	≥ 7.6	≥ 3.4

There are many different methods of assessing energy expenditure (Figure 10). The methods can be divided into those measuring physical activity and those estimating energy expenditure. These are further divided into direct methods that measure the physical activity or energy expenditure directly, and indirect methods that provide a measurement that correlates with true physical activity or energy expenditure. Direct methods for the measurement of physical activity include motion sensors, direct observation, for instance by an assistant, and GPS tracking. All these methods measure the physical activity as it occurs and energy expenditure can be estimated based on the data. Direct methods for the measurement of energy expenditure include calorimetry and doubly labeled water. They are both rather accurate methods for the measurement of energy expenditure. Indirect measurements of energy expenditure include the measurement of oxygen uptake, heart rate, body temperature and ventilation. All these correlate with energy expenditure.

Surveys, questionnaires, recalls and logs are indirect methods that measure physical activity. They are the most practical methods for large epidemiologic studies, because they are low-cost methods. However, they do not provide accurate profiling of the daily EE. The problem with recalls is that they tend to put more emphasis on intentional exercise, and underestimate the unintentional physical activities.

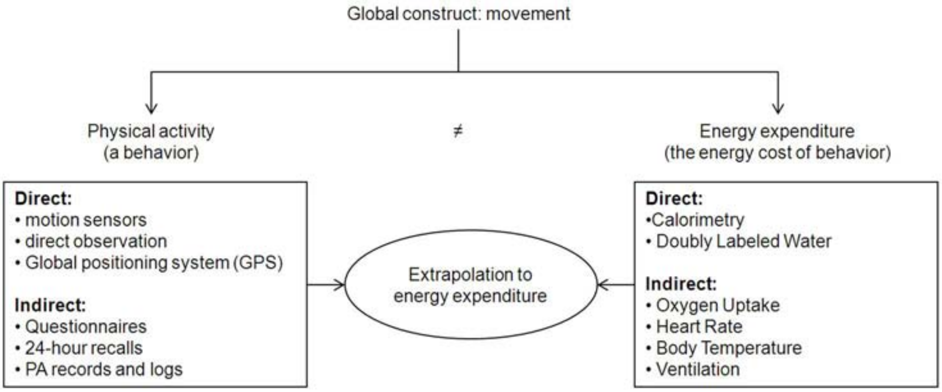


Figure 10. Conceptual framework for defining and assessing physical activity (PA) and energy expenditure (Ainsworth 2009).

Doubly labeled water is considered the gold standard in free-living energy expenditure measurements (Lagerros & Lagiou 2007). It achieves an error rate of 2–10% (Laporte et al. 1985). In this method, the subjects drink water containing isotopically labeled hydrogen and oxygen atoms. An overall estimate of EE is obtained by measuring the proportion of unmetabolized isotopically labeled water from urine. The measurement gives an overall estimate of EE over the measurement time, which is optimally 1 to 2 weeks. The doubly labeled water method is very practical, because it does not affect the activities of the subject. The drawback of the method is the high price of isotopes.

Direct calorimetry measures the production of heat. This can be measured in special chambers, where subjects spend the measurement period. Although the method is very accurate, with an error estimate of less than 1%, it is not applicable to measurements in free-living conditions.

Indirect calorimetry measures oxygen consumption which correlates with heat production. This method requires the subject to wear a face mask and the equipment that analyzes breathing gases. The error rate of this method is 2 to 3% (Laporte et al. 1985). Although this method is not applicable to everyday measurements, it is suitable for reference in (semi-) free-living measurements.

Pedometers, otherwise step counters, measure steps, not the intensity of walking or running. In general, pedometers assess the number of steps accurately, the distance less accurately, and the EE least accurately. 10 tested electronic pedometers estimated EE within +/- 30% at 5 different walking speeds on a treadmill as compared with indirect calorimetry (Crouter et al. 2003). Generally pedometers overestimate the true EE at all walking speeds (Crouter et al. 2003).

Modern accelerometers provide an accurate profile of daily physical activities with the intensity, duration and frequency of activities. The monitors do not interfere with physical activity and they can be considered socially acceptable thanks to their small size. In addition, the cost of such monitors has decreased to a level acceptable for large-scale studies (Troiano et al. 2008). If monitors can be made more accurate, more detailed information on the dose-effect of physical activity can be obtained.

Using accelerometry, correlations of 0.71–0.96 have been obtained between the EE estimate and doubly labeled water (Meijer 1990), indirect calorimetry (Bouten et al. 1994) and whole-room calorimetry using single regression models (one regression equation for the mapping of the activity integral values to METs). A two-regression model including separate regression models for irregular and regular (walking and running) activities achieved a correlation of 0.96 between the EE estimate and indirect calorimetry (Crouter et al. 2006). A recent study used a single triaxial accelerometer and different regression equations for each activity, obtaining the correlation of 0.71 between the EE estimate and doubly labeled water in free-living conditions (Bonomi et al. 2010). The results in all studies are influenced by the selection of the activities performed.

The motion sensors traditionally used for the estimation of energy expenditure include pedometers, actigraphy and only lately, accelerometers. Other movement sensors, such as angular rate sensors have not been studied extensively for the EE estimation. The validity and feasibility of EE estimation methods is summarized in Figure 11.

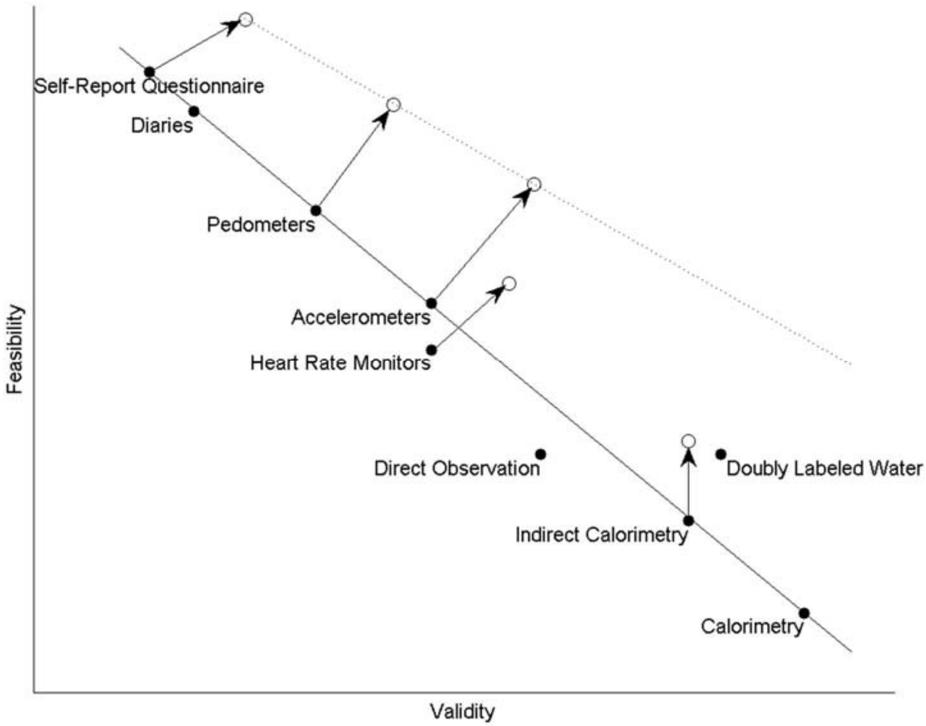


Figure 11. Feasibility versus validity of methods used for monitoring energy expenditure ((Esliger & Tremblay 2007), modified). Arrows indicate changes expected in the future.

2.4 Assessment of Mental Load and Stress

A psychophysiological phenomenon, stress, has become a major public health problem in industrialized countries. In Finland, about 7% of employees suffer from work-related burnout (Kalimo & Toppinen 1997). According to another study, 2.5% of employees in Finland suffer from severe burnout and 24% from mild burnout (Aromaa & Koskinen 2004). 7.4% of employees in Sweden and 4–7% in the Netherlands have been reported to suffer from severe burnout (Shirom 2005). Mental disorder was the most common reason for disability pension in Finland in 2008 (Hiltunen et al. 2008). One third of disability pensions were caused by mental health problems in 13 OECD countries (OECD 2009). In Finland, political decisions have been made in order to fight this trend by means of 1) improving employees' working capabilities and 2) by supporting employees to continue working (Gould et al. 2010).

Short-term mental load is healthy and only improves individual performance, but if a tolerance threshold is exceeded, one is said to have stress (Lindholm & Gockel 2000). In this thesis, the term *mental load* refers to short-term stress and the term *stress* refers to long term stress. Mental load is caused, for example, by intensive concentration on a task. It causes sympathetic responses such as the smaller heart rate variability (HRV) and higher blood pressure that help to cope with a difficult situation that requires high concentration or fast response. Normally, the sympathetic responses decrease as the stressful situation passes, for example, during sleeping or holidays. Prolonged mental load and insufficient recovery may lead to stress, allostatic load, burnout and physical illness (Honkonen et al. 2006). To deal with the prolonged physiological changes caused by stress, the body has to alter its physiological regulation and adapt to the constant stressors. The cumulative price of adaptation is called the allostatic load (Kinnunen 2005). Continued allostatic load may lead to a physical illness. This process may be insidious and the person affected may be unable to feel any changes before the diseases appear. During the allostatic load state, the person may feel symptoms of burnout: emotional exhaustion that does not disappear at leisure, depersonalization or cynicism, and a reduced sense of personal accomplishment (Maslach & Jackson 1981). The key factors affecting psychological ill health at work are: long hours worked, work overload and pressure and their effects on personal lives, lack of participation in decision-making, poor social support, and unclear management and work role (Michie & Williams 2003). The work environment has become mentally more burdening, for example, the work intensity and number of complex tasks at work have increased (Ahola et al. 2010).

Physiological variables, such as the heart rate variability (HRV) and blood pressure are known to be related to sympathetic stress reactions (Shapiro et al. 2001). The resting heart rate, resting-ECG and blood pressure have also been used to identify high sympathetic activity and stress. Similarly, the HRV has been found to be lower in older, ill or stressed individuals in contrast to younger individuals or individuals in good physical condition (Karemaker & Lie 2000). In addition to physiological variables, detected behavioral patterns, such as daily activity patterns, and sleep patterns may be relevant for modeling the user’s life status. Wellbeing of an individual includes physiological, psychological and social factors, all of which are interacting as determinants of health. Behavioral and social factors have been reported to contribute more than 50% to health outcomes (McGinnis et al. 2002).

Thus, there is a need for comprehensive health monitoring solutions that use physiological, psychological and behavioral monitoring. However, little is known about the mutual correlations of these variables in long-term settings or their relationship with changes in the stress status. One of the difficulties is the definition of stress. How could stress be monitored so that each individual would use a common stress scale? There is no gold standard for measuring stress. This has led to the fact that most studies report correlations between measured variables, but do not yet proceed to automatic classification of stress. Recent studies have demonstrated different measurement setups (Intille et al. 2003, Wilhelm et al. 2006) and developed user interfaces for representing the stress history to the user (Sanches et al. 2010), but the connection between the user interface developments, measured data and the actual stress states is currently not clear.

For the monitoring and profiling of the user's mental load and recovery, questionnaires such as the Bergen burnout indicator (BBI), Derogatis stress profile (DSP) and Maslach Burnout Inventory (MBI) have been used (Maslach & Jackson 1981). Long-term monitoring has been suggested for the identification of work-related stress (Van Amelsvoort et al. 2000). Automatic methods that use several data sources for the long-term monitoring of stress would help to identify stress early, and possibly allow earlier intervention when necessary.

Cortisol awakening rise is a widely used indicator of long-term stress. Cortisol samples are assessed based on saliva samples measured during the first hour after awakening. The timing of the sampling is important, because the cortisol level has a circadian rhythm: it is the lowest during the first half of night-time sleep and abruptly increases during the second half of sleep (Kudielka et al. 2006). The peak levels can be measured shortly after morning awakening. After the peak, the cortisol level decreases continuously during the day. Stress-related cortisol superimposes on the circadian cortisol rhythm. The measure is non-invasive, but it requires laboratory analysis, and thus is not applicable to everyday life. Cortisol levels have been shown to be higher in stress and after fragmented sleep (Ekstedt et al. 2004).

Wrist actigraphy has not been used to assess mental load in many studies, but it is used to assess sleep which is often affected under the high mental load. For instance, school children who were reported by teachers to have behavioral symptoms had shorter total sleep time than those without behavioral symptoms (Aronen et al. 2000). Psychiatric inpatients with major depressive disorder were reported to have higher nighttime motor activity than those with less depressive symptoms (Lemke et al. 1999). In home measurements, actigraphy is used instead of polysomnography (PSG), the gold standard for sleep analysis. PSG requires expensive equipment and is more obtrusive (for example nasal pressure, EEG, etc.) and is carried out in hospital sleep laboratories over one or a few nights. Actigraphy is unobtrusive and it can be recorded over multiple days and nights. Total sleep time assessed using actigraphy has been shown to correlate well with PSG in healthy adults with correlations r = 0.97 (Jean-Louis et al. 1996) and r = 0.722…0.836 (Weiss et al. 2010). Also minute-by-minute sleep-wake comparisons have shown good agreement with PSG in adults (91–93%) (Jean-Louis et al. 2001). However, actigraphy has been shown to be less accurate in specific measurements such as sleep offset and sleep efficiency. It has also been shown to be reliable in detecting sleep in healthy populations, but less reliable in detecting sleep, when sleep becomes more disturbed (Jean-Louis et al. 1996). In case of sleep-disordered patients, Kushida et al (Kushida et al. 2001) suggest using data from both the actigraphy and the subjective questionnaires. Artefacts of actigraphic measurements are caused by non-compliance (volunteer not wearing an actigraph), breathing movements, postural blocking of arm movements or vehicle movements when traveling in a vehicle (Pollak et al. 2001).

Previous studies on the automatic assessment of stress can be divided into two categories: 1) those aiming at automatic identification of mental load (short-term stress) or emotions, and 2) those aiming at the identification of long-term stress. The studies focusing on the identification of mental load deal with experiments, where the subject has to perform rather short-duration tasks with different stressors, for example, to give a public presentation, perform computations, etc. The measured signals are then compared, for instance, with those measured before and after the task. The studies focusing on the identification of long-term stress use questionnaires or, for example cortisol measurements as a reference.

2.4.1 Studies on Assessment of Mental Load

Pressure distribution on a chair was studied for the identification of mental load and high social evaluative stress states with 4-minute mental arithmetic tasks and evaluative feedback sessions (Arnrich et al. 2010). 74% accuracy in identifying stress from normal mental load was achieved using the XY-fused Kohonen network and self-organizing maps. Higher variance of sideward movements was recorded during periods with high stress in most volunteers.

ECG, electromyogram, skin conductance and respiration were studied during a 50-minute car-driving task in the city of Boston (Healey & Picard 2005).

5-minute intervals of representative data from rest, highway and city driving tasks were automatically classified with 97% accuracy using Fisher projection and a linear discriminant classifier. In most drivers examined, the skin conductance and heart rate correlated most with the driver's stress level.

Facial electromyogram, respiration, electrodermal activity and ECG were studied in car-racing simulations (Katsis et al. 2006). Five emotions: high stress, low stress, disappointment, euphoria and neutral face were recognized with 86% accuracy with experienced psychologist's manual scoring as a reference. A support vector machine was used as a classifier. Large between-subject variability was observed.

Electromyogram, ECG, skin conductance and respiration were studied during music listening for the purpose of emotion recognition (Kim & André 2008). Four emotional states: positive/high arousal, negative/high arousal, positive/low arousal and negative/low arousal were recognized for each song with 70% accuracy using a specially developed emotion-specific multilevel dichotomous classifier. 65% accuracy was obtained when using linear discriminant analysis.

HRV was studied during a 20-minute conference presentation as well as 30 minutes before and after it for the identification of stress (Kusserow et al. 2008). The talk period was best identified using the RR-interval duration. This feature was better for the identification of stress than, for example, the low-frequency/high-frequency (LF/HF) ratio and respiration frequency.

Electrodermal activity was studied for the identification of mental load and high social evaluative stress states with 4-minute mental arithmetic tasks and evaluative feedback sessions (Setz et al. 2010). Periods with stress were identified with 83% accuracy using linear discriminant analysis.

Electrodermal activity, pulse, pupil diameter and skin temperature were studied during the paced Stroop color test (Zhai & Barreto 2006). Relaxed and stressed states were recognized with 90% accuracy using a support vector machine classifier. Pupil diameter was found to be the best feature.

2.4.2 Studies on Assessment of Stress

HRV was studied for correlation with *self-reported mental strive*, physical strive and busyness (Antila et al. 2005). 12 subjects self-monitored HRV and made self-assessments of several stress-related variables for circa 10 weeks. Beat-to-beat HR was measured daily when awake. A behavioral diary with 12 different variables was filled in every day and every evening. Significant correlations between data from these two sources were found. Daytime “busyness” correlates positively, r = 0.143, with the average HR. Daytime “mental strive” correlates positively, r = 0.158, with the “stress time” that was computed from HRV.

Self-reported mental strain was studied for correlation with *HRV* (Kinnunen et al. 2006). 27 postal workers self-assessed mental strain and their RR-intervals were recorded during a working day, leisure time and sleep. 18 subjects had positive significant correlation between the “absolute stress vector”, a feature derived from HRV and self-reported mental strain. The features were computed using 5-minute time resolution.

HR, motor activity, blood pressure, weight, temperature, self-reported wellbeing, activities and alcohol consumption were studied for the prediction of *daytime diastolic blood pressure* during daily life for 2 to 3 months (Tuomisto et al. 2006). 14 volunteers participated. Self-reported psychological effort and alcohol consumption were found to be the best predictors for daytime diastolic BP. For two-predictor linear regression model the correlation was r = 0.14.

3. Objectives of the Thesis

The main objective of this thesis was to gain new knowledge on **using wearable sensors for long-term wellness management**. Large, annotated data libraries were first collected with unobtrusive sensors as part of everyday activities.

The data collected were then used to achieve the following specific objectives:

1. To identify the most useful sensors, sensor locations and signal interpretation methods for automatic **activity recognition** applications (**P1, P2,** and **P4**).
2. To identify the most useful movement sensors, sensor locations and signal interpretation methods for automatic **assessment of energy expenditure** (**P3**).
3. To identify the best sensors and signal interpretation methods for the automatic assessment of **mental load** and **stress** (**P5** and **P6**).

4. Outlines of the Studies

4.1 Automatic Activity Recognition using Data from Pre-Defined Scenario of Activities

The purpose of Study 1 (**P1**) was to take a data-oriented and empirical approach to automatic activity recognition and collect a large, annotated and well-documented data library of wearable sensor data of different everyday activities with several volunteers. The aim was to find the most information-rich sensors and the most useful signal processing and classification methods for automatic activity recognition. These could potentially be used to develop an "activity diary", which would show the user which activities he/she has performed during the day (or over a longer period) and how much sedentary activities are present. As a user sees this information, he/she can draw conclusions and adjust his/her behavior.

The target activities for the study were 1) lying, 2) sitting/standing, 3) walking, 4) Nordic walking, 5) running, 6) rowing using a rowing ergometer, and 7) bicycling using a bicycle ergometer. A scenario was written which included several activities in different places both the indoors and outdoors. The purpose of the scenario was to make sure each volunteer performed each activity. Total duration of one session was about 2 hours per volunteer. During this time, the volunteer wore several wearable sensors and an assistant followed the volunteer, annotating the activity start and end times on a PDA. Figure 12 shows the data collection system with sensors and their placements. Figure 13 shows the equipment used. Altogether 18 different quantities were measured: 1) altitude, 2) audio, 3) body position, 4) chest accelerations (3D), 5) chest compass bearings (3D), 6) ECG, 7) environmental humidity, 8) environmental light, 9) environmental temperature, 10) heart rate, 11) GPS location, 12) pulse plethysmographic waveform, 13) respiratory effort, 14) blood oxygen saturation (SpO_2), 15) skin conductance, 16) skin temperature, 17) wrist accelerations (3D) and 18) wrist compass bearings (2D). Figure 14 shows the annotation application used with a PDA. The accelerometers were placed on a chest (rucksack strap) and a wrist. Their sampling rate was 200 Hz on the chest and 40 Hz on the wrist. The dynamical range of the accelerometers was – 2 … + 2 g (g refers to gravitational acceleration, ~ 9.81 m/s^2).

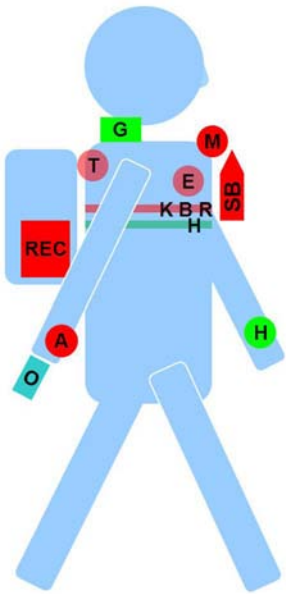


Figure 12. Study 1 sensors and equipment: G = GPS receiver, T = skin temperature sensor, M = microphone, SB = SensorBox (with a 3D accelerometer, 3D magnetometer, environmental temperature, humidity and illumination sensors), E = ECG, R = respiratory effort belt, B = body position sensor, K = skin conductivity, H = Heart rate and altitude monitors, A = SoapBox (with 3D acceleration and 2D magnetometer sensors), O = oximeter on a finger, REC = recorders (small PC and a signal recorder). Coloring indicates storage: data from red sensors are stored on a PC and signal recorder, green sensors store data locally.

16 volunteers (13 males and 3 females) took part in the collection of data. The mean age was 25.8 ± 4.3 years and the BMI was 24.1 ± 3.0 kg/m^2. Most of them were students from a local university.

Figure 13. Study 1 Data Acquisition System: on the left: equipment on the floor, on the right: equipment in the rucksack. On the left, top row: Rucksack. Second row: PC battery, PC, Embla recorder (white) and battery. Third row: Oximeter (gray), SoapBox (black), microphone and amplifier (black), SensorBox. Fourth row: Respiratory effort strap (blue), body position (blue, around strap on the left), ECG electrodes (white-blue), skin temperature (silver, on top of the blue strap), skin conductance (gray, around the blue strap), Suunto wrist-top computer and its heart rate strap below, second oximeter (black), GPS navigator.

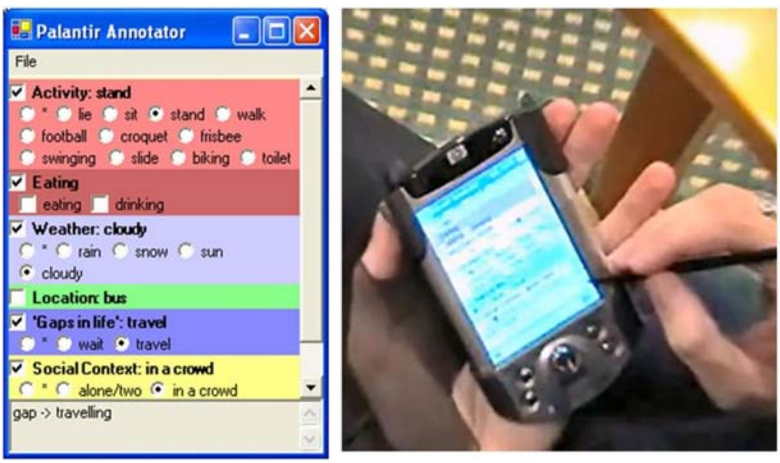


Figure 14. Annotation application on the PDA for storing the “true” activity.

The data were stored on two devices: a small PC and a sleep recorder (Figure 13). The data were used for offline analysis. Several feature signals with 1 Hz sampling frequency were computed from the raw data using a sliding window. The features included time-domain features like mean, variance, median, skewness, kurtosis, 25th percentile and 75th percentile and frequency-domain features like spectral centroid, spectral spread, peak frequency, peak frequency power and signal power in different frequency bands. 4- and 10-second sliding windows were used to compute the features. The best features were selected by the visual comparison of feature distributions between different activities. Three different classifiers: automatic decision tree, custom decision tree and artificial neural network were used for automatic activity recognition. The classifier training was performed by using the feature signals as inputs and the annotation as the target. The results were computed using leave-one-subject-out cross validation.

Custom decision tree represents a simplistic and computationally efficient decision tree. Custom made decision trees were built using a priori information and the input features were selected by human reasoning. First, the tree structure was designed, and second, the features for each node were selected. The thresholds for each node were selected so that they maximized the classification accuracy of training data.

Automatic decision tree refers to a tree that was automatically generated using the training data. Postpruning was used to find an optimal tree structure. Cross-validation (CV) was used by dividing the training set into 10 subsets. In each CV cycle, 9 subsets were used for tree generation and one for testing the obtained tree. The misclassifications of the 10 CV cycles were added together to find the total misclassification rate of each pruning level. The smallest tree with one standard error away in relation to the original, unpruned tree was selected to be the final decision tree.

4.2 Automatic Activity Recognition in Supervised and Unsupervised Conditions

The purpose of Study 2 (**P2**) was to collect data for automatic activity recognition with the updated equipment, both in supervised and in unsupervised conditions. In unsupervised conditions, the volunteers annotated the activities themselves, without the assistant’s presence. The same annotation application on a PDA was used as in Study 1. The target activities in Study 2 were 1) lying, 2) sitting/standing, 3) walking, 4) Nordic walking, 5) running, 6) bicycling, 7) bicycling with bicycle ergometer, 8) rowing with a rowing ergometer and 9) playing football (soccer). The total duration of one session increased from 2 hours to about 6 hours, including 2 hours of supervised time and 4 hours of unsupervised time, when the volunteer was free to do what he or she wanted, wearing the sensors and annotating the activities.

The equipment was updated so that all data were stored on the sleep recorder (Embla), and the PC (Databrick) was no longer used. The Piezo respiratory effort sensor was replaced with the respiratory inductance plethysmogram (RIP) sensor belts on the chest and abdomen. The sensor box with the 3D accelerometer, 3D magnetometer, environment light intensity, environmental temperature and environmental humidity sensors was moved from the chest (rucksack strap) onto the hip (belt). At the same time, the sampling rate of accelerometers and magnetometers was dropped from 200 Hz to 20 Hz to allow longer data collection. Accelerometer dynamical ranges were updated from 2g to 10g. Skin conductance, body position sensors were removed, because they did not produce useful signals for our purpose. The wrist magnetometer was removed, because the signal it provided was like a low-pass filtered version of the accelerometer signal. Thus, the accelerometer provided more accurate information on wrist movements for activity recognition. The differences between Study 1 and Study 2 sensor setups are shown in Table 4. Study 2 sensors and their locations are shown in Figure 15.

Table 4. Sensors, their locations and sampling rates in Study 1 (**P1**) and Study 2 (**P2**).

Signal	**Sensor site (P1)**	**Fs (P1) [Hz]**	**Sensor site (P2)**	**Fs (P2) [Hz]**
3D acceleration	chest & wrist	200 & 40	hip & wrist	20 & 20
3D magnetometer	chest	200	hip	20
2D magnetometer	wrist	40		
Environment light intensity	chest	200	hip	200
Environment humidity	chest	200	hip	1
Environment temperature	chest	200	hip	1
ECG (1-channel: 2 electrodes)	chest	200	chest	200
Respiratory effort	chest (Piezo)	200	chest & abdomen(RIP)	200
Skin temperature	lower neck	200	armpit	1
Body position (metal ball)	chest	200		
Skin conductance (chest)	chest	200		
Pulse wave (PPG) from oximeter			finger	75
Blood oxygen saturation (SaO2) from oximeter	finger & forehead	1 & 3	finger oximeter	3
Heart rate (from Suunto & oximeter)	chest & finger	0.5 & 1	finger oximeter	3
Altitude (barometer & GPS)	wrist & shoulder	0.5 & irregular	GPS on shoulder	1/20
Location, speed (GPS)	shoulder	irregular	shoulder	1/20
Audio (for annotation)	chest	22 000	wrist	8 000
Photos (for annotation)			chest	1/180
Annotation	PDA		PDA	

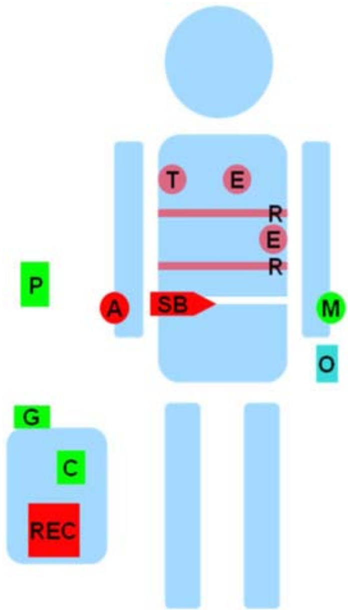


Figure 15. Study 2 Data acquisition system: T = temperature, E = ECG, R = Respiratory effort using RIP sensors, SB = Sensor box, A = accelerometer on the wrist, M = MP3 audio recorder, O = oximeter, P = PDA for annotations, G = GPS receiver, C = Camera, REC = Recorder for storing 19 channels of data.

12 volunteers (10 males and 2 females) took part in the collection of data. The mean age was 27.1 ± 9.2 years and the body mass index (BMI) was 23.8 ± 1.9 kg/m^2. Most of them were students from the local university. Table 5 summarizes the volunteer and recording characteristics of Study 1 (**P1**) and Study 2 (**P2**).

Table 5. Volunteer and recording characteristics in Study 1 and Study 2.

	Study 1 (P1, N = 16)					**Study 2 (P2, N = 12)**				
	Age (yrs)	Weight (kg)	Length (cm)	BMI (kg/m^2)	Recording length (hh:mm.ss)	Age (yrs)	Weight (kg)	Length (cm)	BMI (kg/m^2)	Recording Length (hh:mm.ss)
Min	19	53	160	20.4	1:27:12	19	60	167	21.5	5:49:18
Max	33	95	189	30.3	2:54:03	49	85	190	26.4	7:41:39
Mean	25.8	77.5	178.8	24.1	1:57:15	27.1	76.6	179.2	23.8	6:43:49
Std	±4.3	±12.7	±7.7	±3.0	±0:22:09	±9.2	±7.6	±6.2	±1.9	±0:32:38

Time-domain feature signals computed from the raw data using sliding windows included mean, variance, median, skewness, kurtosis, 25th percetile and 75th percentile. Frequency domain feature signals included peak frequency, peak power, and power on different frequency bands.

Four different classifiers were used for automatic activity recognition: a custom decision tree, automatic decision tree, artificial neural network and a hybrid classifier. The results were computed using leave-one-subject-out cross validation.

The hybrid classifier is a hybrid of a binary decision tree and an artificial neural network. It combines a priori knowledge with the nonlinear classification power of artificial neural networks. In this structure, each node of the custom decision tree was replaced with a small multilayer perceptron (MLP) network. Each MLP was given the same inputs that were given to the whole custom-made decision tree, but each MLP was required to make only one binary classification, the one that was relevant for each node.

4.3 Assessment of Energy Expenditure

The purpose of Study 3 (**P3**) was to assess two types of wearable sensors in the estimation of energy expenditure, and to find the best sensors, sensor locations and signal interpretation methods. Two different sensors, 3D accelerometers and 3D angular rate (gyro) sensors, were packaged into one data logger. The accelerometers had ± 18 g and the angular rate sensors had the ± 100 deg/s dynamical range. Three data loggers were placed in three different body points: 1) the wrist of a non-dominant hand, 2) a hip and 3) an ankle. The true energy expenditure was measured using a portable cardiopulmonary exercise testing system (indirect calorimetry). Figure 16 summarizes the study sensors and sensor locations.

Data were collected while performing common everyday tasks: hanging laundry, ironing, folding and putting away laundry, vacuuming, walking, using the stairs, walking and pushing a shopping cart, walking and carrying bags, running, bicycling with an ergometer, walking and running on a treadmill. Treadmill and bike ergometer resistance were adjusted to be approximately equally strenuous for users of different gender and age. 11 volunteers with the mean age of 38.6 ± 13.1 years took part in the measurements. Their mean length was 170.3 ± 8.5 centimeters, weight 67.5 ± 10.7 kilograms and BMI 23.2 ± 2.6 kg/m^2. Altogether 10 hours 7 minutes of data were collected, which is 55 minutes per volunteer on average. During the data collection, activities were annotated by a nurse using the same PDA annotation application that had been used in Studies 1 & 2 (**P1** and **P2**).

The features computed from the accelerometer and angular rate data included 1) time above threshold, 2) zero-crossings, 3) integral method. Estimates of MET were obtained using linear regression. 30-second median measured MET from the end of each activity were used as references.

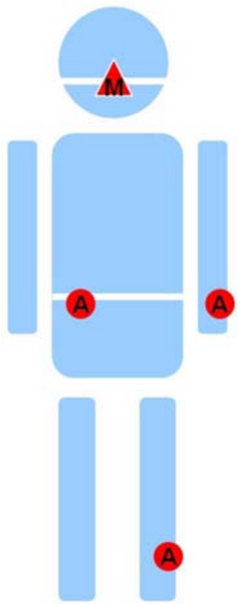


Figure 16. Study 3 data acquisition system: 3 data loggers (A) with a 3D accelerometer and 3D angular rate sensors on the body and the reference (M), a portable breathing gas analyzer.

4.4 Personalized, Online Activity Recognition

Study 4 (**P4**) presented a simplified, wireless and online version of the automatic activity recognition systems developed in Studies 1 and 2 (**P1** and **P2**). In Study 4 (**P4**), only ankle accelerometer data were used, because it was found to be the best sensor and the best sensor location in Study 3 (**P3**). The accelerometer sampling rate was 50 Hz and the dynamical range was ± 6 g. In contrast to Studies 1 and 2 (**P1** and **P2**), where data had been processed and classified offline on a PC, the data collected in Study 4 (**P4**) were analyzed and classified online on a PDA. In this study, the data were measured using wireless motion band sensors (Laurila et al. 2005). The data were transferred from the ankle accelerometer using wireless Bluetooth connection. The PDA application used for annotations in the previous studies was extended to allow user annotations, signal acquisition over Bluetooth, feature computation and classification of activity groups.

Study 4 (**P4**) presented a personalization algorithm that can be used to improve classifier accuracy by introducing individual data to the classifier and evaluating the difference between the default classifier and the personalized classifer performances. The default classifer was trained using leave-one-subject-out cross-validation. The features used for online classification were in time-domain: mean, variance, and in frequency domain peak power and spectral entropy. The classifier used was a custom decision tree.

A new data set was collected with the wireless online activity recognition system including a PDA and the 3D accelerometer on an ankle (Figure 17). The target activities included the most common everyday activities: 1) lying, 2) sitting/standing, 3) walking, 4) running, and 5) bicycling. The system was evaluated on 7 volunteers (6 males and 1 female). Their mean age was 23.6 ± 13.0 years (range 4 … 37) and the mean length was 158.4 ± 36.2 centimeters (range 92 … 187 cm).

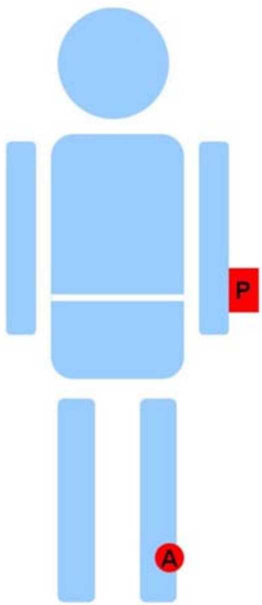


Figure 17. Study 4 data acquisition and activity recognition system: a 3D accelerometer (A) with wireless Bluetooth data transfer on an ankle and a PDA (P) with an application for annotation, data acquisition, feature computation and classification.

A personalization algorithm was developed and used with the custom decision tree. It was designed to keep the classifier structure intact, but update the threshold values with new training data. Thus, the nodes and division features were the ones selected with a priori information, but the personal threshold values were updated with the user's own data.

4.5 Recognition of Physical Activities and Mental Load

In Study 5 (**P5**), the focus was on automatic recognition of physical activities and mental load, identification of the most information-rich sensors and data interpretation methods. The target activities were 1) lying, 2) sitting with normal mental load, 3) sitting with heavy mental load, 4) walking, and 5) running. Sitting with the normal mental load meant sitting and reading comics. Sitting with the heavy mental load meant sitting and working on an IQ (Intelligence Quotient) test on the computer.

The data set collected in Study 2 with 12 volunteers was used. The duration of the IQ test was 20 minutes and the duration of reading comics was 5 minutes. The features computed include time and frequency domain features such as heart rate variability (HRV), acceleration (min, max, mean, variance, peak frequency, peak power, spectral entropy, energy expenditure), compass bearings and respiratory effort using the RIP sensor. Classification was done using three different classifiers: Custom decision tree, artificial neural network and K-Nearest Neighbor algorithm (with K = 5). The features for ANN and KNN were selected automatically using the Sequential Forward Search (SFS) algorithm.

Two different techniques, the histogram transformation (Huiku et al. 2007) and a new normalization technique using the information of the activity context were used to normalize the HR data. The activity context normalization normalizes the individual heart rates to a common range using median HRs during two different activities. This was done to allow identification of high and normal mental loads. The normalization was tested using two different activity pairs: sitting and walking, as well as sitting and running. The median HRs of the two activities are transformed to fixed values on the normalized range.

4.6 Assessment of Perceived Stress

The purpose of Study 6 (**P6**) was to study how different physiological and behavioral variables measured over a long period of time at home in uncontrolled conditions by participants themselves, or automatically by wireless sensors, were related to psychological self-assessments or data acquired by standard validated questionnaires. In addition, the aim was to find the best sensors and signal features that could be used in long-term psychophysiological wellness monitoring.

17 volunteers (14 females and 3 males) were recruited from vocational rehabilitation groups aimed at improving the working ability. They were white-collar workers: university employees and health care employees. The mean age was 54.5 ± 5.4 years. Participants of this program report increased levels of work exhaustion and long-term stress among other health and work-related complaints. Bergen Burnout Inventory (BBI) questionnaire was used to assess the initial level of burnout. The mean BBI of the volunteer group was 49.2 ± 12.0. The rehabilitation was paid by the Social Insurance Institution of Finland (KELA).

The data collection equipment is presented in Figure 18. The data were collected using wearable sensors, home-based sensors, self-assessments and questionnaires. The individual devices are shown in **P6**, Figure 1. In addition, the type of day (work / free / sick / rehabilitation) was written down on a paper form. Feature signals were computed from the raw data. The feature signals had a sampling interval of 1 day.

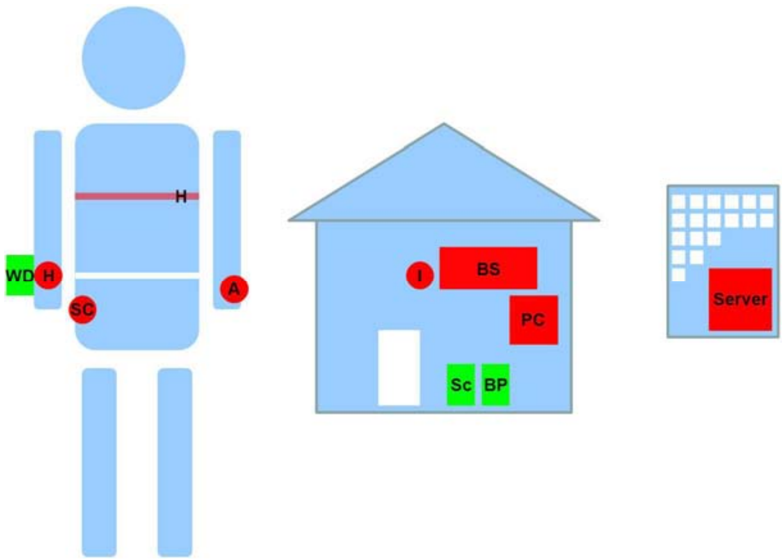


Figure 18. Study 6 data collection equipment: (A) IST wrist activity monitor (actigraph), (H) Suunto heart rate monitor, (WD) Nokia mobile phone with Wellness Diary application for self-assessments and measurement results from (SC) Omron pedometer, (BP) Omron blood pressure monitor and (Sc) weight scale, (BS) Emfit bed sensor, (I) bedroom temperature and illumination sensors, (PC) laptop PC and (Server) central server.

The study protocol (Figure 19) included a two-week rehabilitation in a rehabilitation center. The participants used the self-monitoring equipment for two weeks before the rehabilitation, during the two-week rehabilitation and for two months after the rehabilitation. In the beginning of the study, the participants filled in the BBI questionnaire. They filled the first DSP questionnaire, as the data collection equipment was installed into their homes. The second DSP was filled in after the rehabilitation, the third one a month after that, and the fourth one, a month later, at the end of the measurement period. Study protocol is shown in Figure 19.

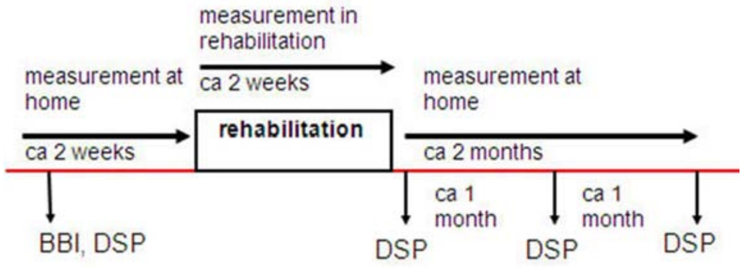


Figure 19. Study protocol: measurements and questionnaires, DSP = Derogatis Stress Profile, BBI = Bergen Burnout Inventory.

In addition to the DSP, stress assessments were performed daily by means of a Wellness Diary. The stress level was assessed every evening and typed into the application using the visual analog scale 0 … 10. The value given reflected the stress level of the whole day. The daily measurement and assessment routines are summarized in Table 6. Other measurements were carried out automatically and did not require user interaction.

Table 6. Daily measurements and self-assessments. WD stands for Wellness Diary mobile phone application.

Morning	**Actions needed**
Weight	Measure and fill weight form in WD
Blood pressure	Measure and fill blood pressure form in WD
Sleep	Estimate length and quality of previous night's sleep and fill in sleep form in WD
Steps	Place pedometer in pocket
Heart rate	Start HR measurement. Measure HR 3 days a week (2 on work days, 1 at the weekend)
Evening	**Actions before going to bed**
Heart rate	Stop measurement and transfer data to laptop
Blood pressure	Measure and fill blood pressure form in WD
Stress	Assess day's stress level and fill in stress form in WD
Steps	Check day's step count on pedometer and fill in steps form in WD
Others	
Exercise	Fill in exercise form in WD for each sports activity
Wellness Diary	Send measurement results from mobile phone to research server once a week

Features from the WD are values entered once per day. Bed sensor features were computed from presence and HR data that the sensor gave out once per minute. The computed features included the start time of overnight bed presence, the end time of overnight bed presence, the number of wakeups at night, bed time length, time in bed during daytime, mean bed time during last 3 nights, mean night HR. The activity features were computed from wrist activity count data that indicated the amount of movements. Sleep was identified based on low activity. The activity features included the mean night activity, standard deviation of night activity, day activity standard deviation of day activity, ratio of night activity and previous day activity, ratio of night activity and next day activity, sleep length and the number of sleep periods. For illumination data, the mean, median and variance of night data were computed. HRV features were computed using Firstbeat PRO Wellness Analysis Software (Firstbeat Technologies Ltd., Jyväskylä, Finland). The software splits the HR data into stationary segments of sports, stress and relaxation based on the HR, HRV and the indices derived from these. A medical doctor scored the sleep length and sleep quality using the activity monitor signal, bed presence signal and the WD self-reported sleep start and end times. The features were correlated with two targets: 1) daily self-assessment of stress level and 2) DSP questionnaire total stress score measured once per month for identification of the features reflecting changes in perceived stress level.

5. Results of the Studies

5.1 Automatic Activity Recognition using Data from Pre-Defined Scenario of Activities

In Publication **P1**, the target was to recognize seven activities: lying, sitting/standing, walking, Nordic walking, running, bicycling (using a bicycle ergometer) and rowing (using a rowing ergometer) using data recorded with wearable sensors. The data library collected with 16 volunteers contained more than 31 hours of annotated data from wearable sensors. Altogether 35 channels of data were collected. The data were synchronized, calibrated, re-sampled and converted into accessible formats. The data library was carefully documented, stored on a DVD and shared with collaborating companies.

For recognizing target activities, the accelerometer signals proved to be the most valuable signals as they reacted to activity changes with clear signal changes without delay. The magnetometer, environmental light intensity, GPS and audio also reacted immediately to activity or other context changes. Heart rate, respiratory effort and pulse plethysmograph signals can be used to assess the intensity of the activity, but they are not very suitable for recognizing the activity type. Heart rate and respiratory effort signals did not change immediately as the activity type changed (for instance, from running to sitting). The pulse plethysmogram on the other hand was prone to movement artefacts and produced a useful signal only in rest.

Feature signals were computed from the raw data and the best features were selected using the visual inspection of synchronized signals and annotations. The boxplots of feature distributions during different activities were also used for feature selection. The feature signals were computed to have a common sampling frequency of 1 Hz. Both time-domain and frequency domain features were found useful for activity recognition. For example, the FFT was used to extract the pace of periodic activities. This can be visualized in the form of a time-frequency plot or spectogram (Figure 20).

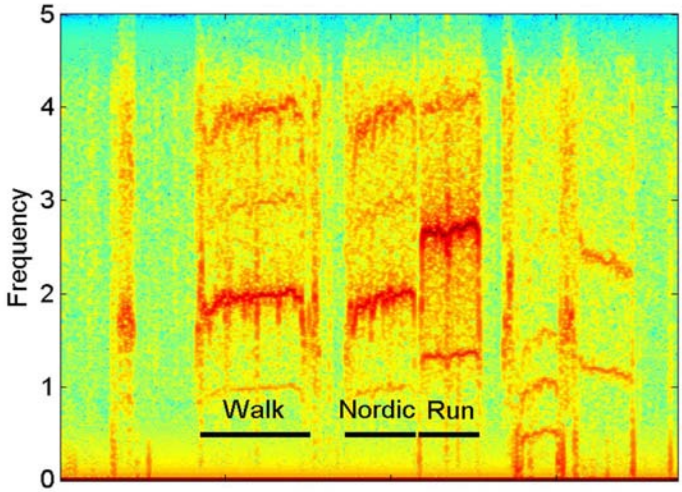


Figure 20. Spectogram of vertical acceleration on the chest during walking, Nordic walking and running activities.

The best features were selected by visualizing the features by 1) plotting and comparing the features and annotation in the time domain and by 2) plotting the distribution of feature values during each activity. The selected features are summarized in Table 7.

Table 7. Selected sensors, sensor placements, features and sensor dimensions for activity recognition.

Feature	**Sensor**	**Placement**	**Dimension**
Median	Accelerometer	Chest	Up-down
Variance	Accelerometer	Chest	Back-forth
Sum of 3D variances	Accelerometer	Wrist	3D
Peak Frequency	Accelerometer	Chest	Up-down
Peak Power	Accelerometer	Chest	Up-down
Power ratio: 1–1.5 Hz / 0.2–5Hz	Magnetometer	Chest	Left-right

Automatic classification was performed with 1 Hz resolution using three different classifiers: artificial neural network (Multi-Layer Perceptron), automatic decision tree and a custom-made binary decision tree. Table 8 summarizes the activity recognition results of the three different classifiers. All three classifiers were given the same six input features. The results were computed using leave-one-subject-out cross validation.

Table 8. Results of the automatic activity recognition using data collected in a supervised manner (assistant annotations as references)

	Lying	**Rowing (ergometer)**	**Cycling (ergometer)**	**Sitting/ Standing**	**Running**	**Nordic walking**	**Walking**	**TOTAL**
Custom Decision Tree	**87**	**69**	79	**96**	**97**	**90**	58	82
Automatic Decision Tree	83	56	**82**	95	**97**	72	78	**86**
Artificial Neural Network	74	59	75	**96**	22	52	**79**	82

Overall, activities containing periodic movements were detected with good accuracy. The best activities for automatic activity recognition are activities with periodic movements and distinct characteristic from other activities, for example, Nordic walking, where typical walking hip accelerations are accompanied with large wrist acceleration impulses as the pole hits the ground. Also static postures can be detected with good accuracy (for instance, lying). However, in this study, accelerometers on the chest and wrist did not allow discrimination between sitting and standing.

Annotations made by an assistant on a PDA application were found to be accurate except for certain occasions. For example when a volunteer and assistant went to a bus, paid their trip and walked to their seats, the volunteer sat down while the assistant was paying for his trip. Thus, an annotation error of some seconds was present. Generally, it can be said that the accuracy of annotations is in the order of ± 1 sec.

5.2 Automatic Activity Recognition in Supervised and Unsupervised Conditions

In Publication **P2**, a new data set was collected with an updated sensor setup and some new activities. The target activities included lying, sitting/standing, walking, Nordic walking, running, rowing, bicycling, bicycling using a bicycle ergometer, and playing football (soccer). In addition to the supervised activities, the volunteers spent 4 hours outside the laboratory, without assistant's supervision, annotating their activities on the PDA application themselves. The features used for the automatic recognition of the activities are summarized in Table 9. One more time, the accelerometer features were found the most information-rich signals for recognizing activity types. The speed computed from the GPS location data was also used for recognizing the football activity, where a lot of movement occurs within a restricted area. Moving the accelerometer from the chest to a hip did not help in the process of discriminating sitting and standing. Otherwise, features similar to chest accelerations were found useful, also by using the hip accelerations. Lowering accelerometer sampling rates from 200 Hz to 20 Hz did not dramatically degrade the recognition accuracy. Only the high impulses from Nordic walking would require higher sampling rates to be accurately recorded.

Table 9. Selected sensors, sensor placements, features and sensor dimensions for activity recognition.

Feature	**Sensor**	**Placement**	**Dimension**
Range	Accelerometer	Hip	Up-down
Mean	Accelerometer	Hip	Up-down
Sum of 3D variances	Accelerometer	Wrist	3D
Speed	GPS	Shoulder	
Peak Frequency	Accelerometer	Hip	Up-down
Peak Frequency	Accelerometer	Wrist	Back-forth
Spectral Entropy	Accelerometer	Hip	Up-down

The classification was performed using the same type of classifiers as in Study 1 and one additional classifier: the hybrid of a binary decision tree and an ANN.

The classification results are summarized in Table 10. The results were computed using leave-one subject-out cross validation and classifying each second of the data into one of the nine classes. All four classifiers were given the same seven input features.

Table 10. Results of automatic activity recognition using data collected in supervised and unsupervised settings (both phases: assistant-annotated and volunteer-annotated data)

	Lying	Rowing (ergometer)	Cycling (ergometer)	Sitting/ Standing	Running	Nordic walking	Walking	Football	Cycling	TOTAL
Custom Decision Tree	**98**	58	20	94	**91**	**85**	50	63	52	83
Automatic Decision Tree	96	84	**79**	53	83	66	62	55	**74**	60
Artificial Neural Network	**98**	85	4	96	90	66	67	47	67	87
Hybrid Classifier	97	**87**	18	**97**	89	70	**71**	**78**	72	**89**

5.3 Assessment of Energy Expenditure

In Publication **P3**, the target was to assess energy expenditure using a 3D accelerometer and 3D angular rate sensors placed in three different points on the body: ankle, hip and wrist. The reference MET values were measured with a breathing gas analyzer during several everyday activities: hanging, ironing, folding and putting away laundry, vacuuming, walking, using the stairs, walking and pushing a shopping cart, walking and carrying bags, running, bicycling with an ergometer, walking and running on a treadmill.

To find a steady-state period of each activity, the last 30 seconds of data collected for each activity were used. Three features were computed: time above a threshold, zero-crossings, and integral. The integral method proved to be the most accurate estimate as it measures both the time and intensity of movements. Before using the integral method, the 3D signals were rectified, summed, and band-pass filtered (0.5–11 Hz) to highlight the voluntary human movements.

Table 11. Selected sensors, sensor placements, features and sensor dimensions for the estimation of energy expenditure.

Feature	Sensor	Placement	Dimension
Integral	Accelerometer	Ankle	3D
Integral	Accelerometer	Hip	3D
Integral	Accelerometer	Wrist	3D
Integral	Gyro	Ankle	3D
Integral	Gyro	Hip	3D
Integral	Gyro	Wrist	3D

The results are summarized in Figure 21 and Table 12. In the case of the household activities in this study, the 3D acceleration sensor on the ankle provided the most accurate estimates of MET. The results obtained with the ankle accelerometer have a RMSE of 1.21 MET. Hip and wrist locations fail to provide a good resolution for low-intensity activities. Data from an angular rate sensor provided only slightly worse estimates of MET than the accelerometer sensors.

Task
1. Hanging laundry
2. Ironing laundry
3. Folding laundry
4. Putting away laundry (on a shelf)
5. Vacuuming
6. Walking up the stairs
7. Walking down the stairs
8. Walking and pushing a shopping cart
9. Walking and carrying bags
10. Walking (at a free pace)
11. Running (at a free pace)
12. Cycling a bike ergometer (65% of max)
13. Walking on a treadmill (35% of max)
14. Walking on a treadmill (45% of max)
15. Walking on a treadmill (55% of max)
16. Running on a treadmill (65% of max)
17. Running on a treadmill (75% of max)
18. Running on a treadmill (85% of max)

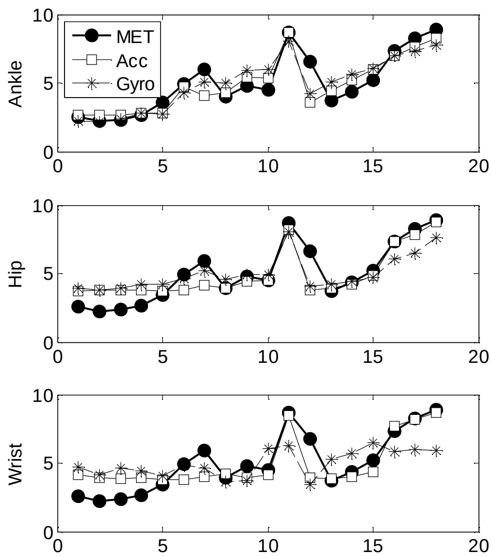


Figure 21. Task names, measured METs and estimates (task number on the x-axis).

Table 12. Number of tasks (N) included in the analysis, Pearson correlation (r) and root-mean-square error (RMSE) between the measured MET and estimates.

	N	r	RMSE [MET]
Acc Ankle	163	0.86	1.21
Gyro Ankle	163	0.84	1.32
Acc Wrist	178	0.81	1.40
Acc Hip	177	0.80	1.42
Gyro Hip	177	0.69	1.71
Gyro Wrist	178	0.48	2.09

When using wearable data loggers for estimating MET, higher-intensity activities (such as running) get better accuracy estimates than low-intensity activities (such as ironing laundry).

The annotation accuracy in this study was in the order of ± 1 sec. The same annotation tool as in Publication **P1** was used by an assistant.

5.4 Personalized, Online Activity Recognition

In Publication **P4**, a new data set was collected using an online activity recognition system on a PDA. Data were measured using a 3D accelerometer on the ankle that sent the data to the PDA using Bluetooth. The target was to recognize five activities: lying, sitting/standing, walking, bicycling and running. Additionally, a personalization algorithm was developed for improving classifier performance on individual data.

Only data from the ankle location was used in activity recognition. The features selected for the classification were computed from the up-down accelerations of the ankle accelerometer. The features selected are summarized in Table 13.

Table 13. Selected sensors, sensor placements, features and sensor dimensions for activity recognition.

Feature	**Sensor**	**Placement**	**Dimension**
Mean	Accelerometer	Ankle	Up-down
Variance	Accelerometer	Ankle	Up-down
Spectral peak power	Accelerometer	Ankle	Up-down
Spectral Entropy	Accelerometer	Ankle	Up-down

The binary decision tree classifier was used for online activity recognition. First, the results were computed with leave-one-subject-out cross validation, then individual data was introduced, one activity at a time and activity recognition results were computed after introducing each new activity.

Summary of the case-wise classification results before, during and after the personalization is presented in Table 14. The overall accuracy was 86.6% before personalization and 94.0% after personalization. The results were computed online on the PDA. The results before the personalization were computed using leave-one-subject-out cross validation and classifying each second of data. The annotation accuracy in this study was in the order of ± 1 sec. The same annotation tool as in Publication **P1** was used by an assistant.

Table 14. Individual classification accuracies in original stage and after personalization.

ID(sex, age, length)	**Original [%]**	**Personalized [%]**
Case 1 (M, 37, 180)	87	**99**
Case 2 (F, 37, 156)	80	89
Case 3 (M, 27, 180)	79	**99**
Case 4 (M, 28, 186)	90	90
Case 5 (M, 24, 187)	88	88
Case 6 (M, 4, 92)	74	77
Case 7 (M, 8, 128)	**95**	**99**
Overall	**87**	**94**

5.5 Recognition of Physical Activities and Mental Load

In Publication **P5**, the data set used in Publication **P2** was used for recognizing physical activities and mental load in supervised settings. The target activities included lying, sitting and the normal mental load, sitting and the heavy mental load, walking and running. Data used for the recognition were measured with accelerometers on the hip and wrist, magnetometers on the hip and wrist, 2 ECG electrodes, respiratory inductive plethysmogram belts on the abdomen and chest, and skin temperature and pulse plethysmogram using a finger oximeter.

Three different classifiers were used for automatic recognition: custom decision tree, 5-Nearest Neighbour (5-NN) and artificial neural network (Multi-Layer Perceptron). Features for the first two classifiers were selected using an automatic feature selection algorithm, the sequential forward search (SFS); features for the custom decision tree were selected manually. The selected features are summarized in Table 15.

Table 15. Selected sensors, sensor placements, features and sensor dimensions selected for the recognition of physical activities and the mental load.

Feature (Custom Decision Tree)	**Sensor**	**Placement**	**Dimension**
Mean	Accelerometer	Hip	Up-down
Normalized HR	2-electrode ECG	Chest	
Range	Accelerometer	Hip	Up-down
Peak power	Accelerometer	Hip	Up-down
Feature (SFS + 5-NN)	**Sensor**	**Placement**	**Dimension**
Maximum	Accelerometer	Hip	Up-down
Normalized HR	ECG	Chest	
Standard deviation	Respiratory Effort	Chest	
Minimum	Accelerometer	Hip	Up-down
Declination angle	Magnetometer	Hip	3D
Feature (SFS+ANN)	**Sensor**	**Placement**	**Dimension**
Maximum	Accelerometer	Hip	Back-forth
Estimate of EE	Accelerometer	Hip	3D
Minimum	Accelerometer	Hip	Up-down
Normalized HR	ECG	Chest	
Peak power	Accelerometer	Hip	Back-forth

The results were computed using the leave-one-subject-out cross validation and for each second of the data. Table 16 summarizes the recognition results. Raw heart rate data were not significantly different during the heavy versus normal mental load, but the normalized HRs and the standard deviation of respiratory effort signal were significantly different. The HR normalization was done utilizing information on the activity context. The inter-individual HR variability during walking was large compared to that during running. Thus, the running HR and sitting HR were selected as end-points for normalization. The respiratory effort signal was found to have lower variability during the heavy mental load than during the normal mental load.

Table 16. Recognition results for physical activities and the mental load.

	Lying	**Normal mental load**	**Heavy mental load**	**Walking**	**Running**	**TOTAL**
Custom Decision Tree	**98**	78	**93**	84	91	**89**
K-Nearest Neighbor	**98**	68	84	**95**	**100**	**89**
Artificial Neural Network	91	**94**	59	85	99	85

5.6 Assessment of Perceived Stress

In Publication **P6**, a new data set was collected using a wrist activity monitor (actigraph), a heart rate monitor, a bed sensor, environmental sensors and a mobile phone with the Wellness Diary application for self-assessments and measurement results from a pedometer, a blood pressure monitor and a weight scale. The target was to compare measurements with perceived stress. The perceived stress was assessed with two different measures: daily self-assessment on a mobile phone and a monthly DSP questionnaire.

Spearman correlations (Table 17) were computed between assessed busyness/stress/pressure and the measured variables. Only workdays were included in this correlation calculation, because the focus was on work-related stress, not on holiday stress. The personal average of the variable value was subtracted first, before pooling the data together.

Table 17. Statistically significant correlations to the daily self-assessment of stress. Only statistically significant overall correlations ($p < 0.05$) are shown.

Feature	Sensor	Placement	P	RHO
Sleep length	Self-assessed (WD)		0.00	-0.26
Sleep length	Actigraph (IST)	Wrist	0.00	-0.22
Number of weight entries	WD		0.00	0.13
Standard deviation during the day	Actigraph (IST)	Wrist	0.00	-0.15
Median night	Illumination	Bedroom	0.00	-0.14
Sport time	Suunto T6 HRM	Chest belt	0.00	-0.33
Variance at night	Illumination	Bedroom	0.00	0.13
Number of exercise entries	WD		0.00	-0.10
Sleep length	Scored		0.00	-0.12
Mean day activity	Actigraph (IST)	Wrist	0.00	-0.12
Sleep quality	Scored		0.00	0.21
Mean at night	Illumination	Bedroom	0.00	0.12
Sleep quality	Self-assessed (WD)		0.00	-0.10
Morning diastolic BP	Blood-pressure monitor (Omron)	Arm	0.01	0.10
Relaxation time	Suunto T6 HRM	Chest belt	0.01	0.15
Stress time	Suunto T6 HRM	Chest belt	0.01	0.15

Correlations to DSPtss points (Table 17) were obtained by computing the median of psychophysiological variables over 7 days (the day people filled in the DSP questionnaire and 6 days before). The personal average was not subtracted, before pooling the data from different cases together, because there were only maximum four DSPtss measurements per case.

Table 18. Statistically significant correlations (7-day medians) with the DSP total stress score. Only significant correlations ($p < 0.05$) are shown.

Feature	**Sensor**	**Placement**	**p**	**rho**
Mean night HR	Bed sensor	Below bed mattress	0.03	0.59
Standard deviation during the day	Actigraph (IST)	Wrist	0.03	0.35
Standard deviation at night	Actigraph (IST)	Wrist	0.03	0.34
Exercise duration	WD		0.04	-0.36
Weight change compared to the previous day	WD		0.04	0.30

The daily annotation of perceived stress was stored on the mobile phone application. The patients reported that it was rather difficult to assess one's stress level, and there might be changes in the way the patients performed the self-assessment over time. The full resolution 0 … 5 may thus not be reliable, but it was estimated that it was accurate with a tolerance of 1 unit on the visual analog scale (VAS).

6. Discussion

6.1 Results versus Objectives

The first objective was to identify the most useful wearable sensors, sensor locations and signal interpretation methods for automatic activity recognition.

This objective was achieved in Studies **P1**, **P2**, and **P4**. A large, annotated and well-documented data library of wearable sensor data was collected for the purpose of the development and validation of automatic activity recognition algorithms. Based on the collected data library, algorithms for automatic activity recognition were developed and evaluated, and the most information-rich sensors and signal interpretation methods were identified.

Based on the collected data sets, it was established that accelerometers were the most useful sensors for automatic activity recognition. As the activity changes, an accelerometer provides a clear signal change, without any delay. A magnetometer, environmental light intensity, GPS, and audio data also show immediate responses to activity and other context changes, and thus are potentially useful. The best locations of sensors are dependent on the activities to be recognized. Based on the data collected, movement sensors on the torso (chest and hip) provide data that represent the whole body movements and allow identification of many activities. A movement sensor on the ankle also provides useful data for the identification of everyday activities, as it picks a strong signal of leg movements. For the identification of more specific activities (such as Nordic walking versus walking), additional sensors placed on extremities are required. For signal interpretation, frequency-domain and time-domain features of movement sensor data were the most useful ones (for example, peak frequency, spectral entropy, mean and variance). The selection of the classifier did not play a critical role. Classification results of a computationally effective custom decision tree classifier were competitive with more complex classifiers. The custom decision tree classifier was implemented on a portable device and successfully used for wireless, on-line activity recognition on a PDA. A personalization algorithm for the custom decision tree was successfully implemented to improve the classification accuracies of different individuals' data.

The second objective was to identify the best movement sensors, sensor locations and signal interpretation for automatic assessment of energy expenditure.

This objective was achieved in Study **P3**. A data set was collected with compact data loggers on different body points during various everyday activities. As a result, the best sensors, sensor locations and data interpretation methods for the estimation of energy expenditure were identified. The estimate computed using the integral method on the ankle accelerometer data gave the best estimates of energy expenditure on the activity set selected for the study.

The third objective was to identify the best sensors and signal interpretation methods for automatic assessment of mental load and stress.

This objective was achieved in Studies **P5** and **P6**. The automatic assessment of the day-time mental load can be improved 1) by utilizing the activity context available from the automatic activity recognition and 2) by scaling the biosignals measured from individuals to a common range. In addition to measuring the day-time mental load, the measurement of recovery, for example, at night is useful for the automatic assessment of longer-term stress.

Recognition of the short-term mental load during tasks of a 6-hour data collection period was performed in Study **P5**. This study applies the recognized activity type for the automatic recognition of mental load and identifies the best sensors for the recognition of short-term mental load and physical activities. Heart rate and the standard deviation from respiratory effort were found to be the distinguishing features for the recognition of the heavy mental load. A new scaling method based on the HR during different activities was introduced for transforming the individual HRs to a common range.

Methods for assessing longer-term, perceived mental stress were developed using data obtained from wearable and fixed sensors in Study **P6**. A new data set was collected as part of a vocational rehabilitation program. The features with strongest correlations to two references: 1) self-assessed daily stress level and, 2) to the DSP questionnaire total stress score were identified. Sleep length estimated based on wrist actigraphy and bedroom illumination at night measured with an illumination sensor were found to be the best methods for the automatic assessment of perceived daily stress. An average night heart rate was found to be the best method for the automatic assessment of longer-term stress, as measured using the DSP questionnaire. Overall, the **P6** results showed modest, but significant pooled overall correlations between self-assessed stress level and physiological and behavioral variables. Strong, but sometimes conflicting correlations can were found in individual data.

6.2 Impact of Studies in Their Research Fields

In Studies **P1** and **P2**, an exceptionally large data library was collected in out-of-lab and realistic laboratory settings. They were one of the first studies on activity recognition based on such extensive data sets. Nonetheless, until now only a few studies have been published on automatic activity recognition that use annotated data collected in out-of-lab or realistic laboratory settings with several different wearable sensors and with several volunteers (Bao & Intille 2004, Foerster et al. 1999). When measurements are carried out in realistic settings, the variability of data increases and the classification accuracy decreases compared to the results obtained in laboratory environment (Foerster et al. 1999). Publication **P1** demonstrated the classification results in supervised settings and **P2** demonstrated their applicability also in unsupervised settings. The recognition accuracy decreased only slightly when the algorithms were applied to unsupervised data. The recognition accuracies reached in **P1** and **P2** are among the most accurate ones obtained in multi-class activity recognition studies (Preece et al. 2009). In addition, the classification results were computed with a higher frequency in **P1** and **P2** than in previous studies (Foerster et al. 1999). The target was inferred once per second using sliding windows in **P1** and **P2**, while block-windows of 20…40 sec were used in an earlier study (Foerster et al. 1999).

One of the best-known models for estimating energy expenditure from accelerometer data was presented by Crouter et al. (Crouter et al. 2006). Publication **P3** demonstrates results obtained using a single-linear-regression model. The single-regression model was found to perform better in the case of the activities in **P3**. Sensors other than accelerometers have not widely been studied in terms of energy expenditure estimation. Publication **P3** shows that EE estimates obtained with angular rate sensors perform almost as well as those obtained with accelerometers.

Publication **P4** demonstrates the use of an activity classification algorithm in an online environment on a PDA. This is one step in the process of developing and transforming research prototypes into commercial products. Publication **P4** evaluates the algorithms in a wireless, online environment and introduces a method for classifier (and product) personalization for the purpose of achieving a better accuracy of activity recognition. The effect of personalization has not been widely studied, but provides one way to improve the accuracy of a product. The study shows that default training gives good results for many individuals, but for some users, classifier personalization is needed to ensure equal performance.

The automatic classification of both physical activities and different levels of mental load in everyday life or realistic laboratory conditions has not been approached by many studies. Previous studies mainly concentrate on comparing feature signals during different activities, but do not yet infer classes from features (Kusserow et al. 2008). In Publication **P5**, such a classifier was developed and evaluated. The results identify feature signals and inter-individual scaling techniques for the recognition of mental load. Information from automatic activity recognition was found useful for improving the automatic recognition of mental load.

Previous studies on long-term stress, work exhaustion and burnout mainly adopt a psychological approach and use questionnaires as the primary tools of assessing the stress level (Honkonen et al. 2006, Maslach & Jackson 1981, Derogatis 1987). Study **P6** employed a more technological approach with the purpose of identifying the most information-rich sensors and signal interpretation methods for the assessment of stress. Publication **P6** was one of the first studies to evaluate the use of several wearable sensors in the assessment of long-term stress and burnout as part of everyday life and during a rehabilitation program. The results also show that people with long-term stress are skilled enough to use wearable monitoring devices and self-assessment tools. However, the devices may overburden people suffering from severe burnout. People react to increasing stress individually and changes can be seen in different variables. This highlights the need for the monitoring of personal trends of each individual.

6.3 Experiences with Sensors and Methods

Sensors that react fast to activity changes are particularly useful for the recognition of activities. They included for example accelerometer, magnetometer, angular rate, illumination, and GPS sensors.

Sensors that react slowly or poorly to activity changes are more difficult to use for activity recognition purposes. In this study, the following sensors represent this category: temperature, humidity, heart rate and respiratory effort. Sensors whose signal is easily disturbed by artefacts, for example because of movement or light are also difficult to use for activity recognition. The following sensors were prone to artefacts: an oximeter on finger and an oximeter on forehead.

Acceleration sensors are very useful for activity recognition as they react fast and robustly to changes in the activity and posture. The locations for the 3D acceleration sensors on the torso and the extremities were good. The signal from the torso sensor is useful for recognition of many activities such as walking, running and biking on an exercise bike. The signal from the wrist is useful in the detection of activities involving frequent hand movements such as Nordic walking. The separation of standing and sitting would require a signal from the lower parts of the body.

Magnetometer sensors are not as reliable in activity detection as acceleration sensors because there are lots of metallic objects in our everyday environment. They cause small changes to Earth’s magnetic field, and thus artifacts to a magnetometer signal. Even then, activities consisting of periodic movements can be detected reasonably well also from the magnetometer signal. Such activities include for example walking and running. In the analysis and calibration of a magnetometer signal, it must be noted that in Finland the strongest component is perpendicular to the earth surface. Thus, the sensor measures not only the direction, but also the posture of the person.

Lower sampling rates (20 Hz for acceleration and magnetometer signals) keep the volume of data reasonable, but the lower sampling rate degrades the recognition of activities that involve fast transients. For example, during Nordic walking there are spikes in the wrist accelerometer data, every time the pole hits the ground. With the lower sampling rate, the signal from activities involving fast transients is not so different from a signal of similar activities not involving such transients (for example, Nordic walking versus walking). Thus, the selection of a sampling rate is a compromise between a small volume of data and a good recognition performance. Lower sampling rate also ensures a longer battery life.

A 10g acceleration sensor on a wrist provided a range large enough for the measurement of Nordic walking. In Study **P1**, a 2 g acceleration sensor was used on the wrist; in some cases, the signal peaks were clipped during the Nordic walking because of too large accelerations. Thus, a 10g sensor is better for this purpose, but a 6g sensor would also have a range large enough as well.

Biosignals provide useful information for the measurement of the exertion level of an individual, but they do not react to changes in the type of activity. For example, the *heart rate* clearly shows the intensity with which an individual is performing a task, but it does not give any information as to whether the person is running or biking. There are also large differences between individuals, for example, the resting heart rates are different as well as changes in the heart rate after certain exertion. In addition, the heart rate remains high for a rather long time after the exertion stops, so the reaction to the activity change is slow. Heart rate can be measured with good reliability. However, the *respiratory effort* signal measured with a piezo belt contains too many artefacts on many occasions, for example because of steps. When a person is not moving, the respiration rate can be detected with a fairly good reliability. Regular breathing is available only when the subject is not talking, eating, coughing, etc.

Oximeter signals (*pulse*, *PPG* and SpO_2) often disappear because of movement artefacts. External light also interferes with the signals. The probe location on a forehead is slightly more immune to movements, but on the other hand, it can be difficult to find a good location on the forehead to get a signal powerful enough. A finger probe provides a more powerful signal, but is more prone to movement artefacts. The pulse and SpO_2 signals do not provide new useful information, but the PPG signal could be used as an estimate of relative changes in the blood pressure. However, the device should be firmly attached on skin and covered from external light for this to work in for ambulatory measurements. In this study, the signal quality was not good enough for such purposes.

A Skin conductance signal reacts only to heavy sweating, not to small changes in dry skin conductance. The heavy sweating made the material behind the sensor wet and the signal did not return from the saturated state during the rest of the measurement. Because of these problems with hardware, the signal was not used in activity recognition.

Skin temperature shows not only changes in skin temperature, but more clearly changes in the surrounding air temperature, for example, when moving between indoor/outdoor locations. Thus, the recognition of exercise types based on the skin temperature is difficult with the currently available signals and equipment. The signal can be useful in stable conditions (indoors), but the sensor used in studies **P1** and **P2** reacted to changes too slowly. The skin temperature sensor had a faster response than the environment temperature sensor that was part of a sensor board with a bigger mass.

A body position sensor did not work in the mobile measurements. The signal provides some information on movements, for instance whether the person is stable or not, but acceleration signals are more useful for this purpose.

An illumination sensor was a useful sensor for detecting light in general and artificial light in particular. The sensor has a fast response to changes in illumination level. In **P1** and **P2** it was used to detect indoor/outdoor locations, and in **P6** to detect waking at night.

Environment humidity and temperature sensors were not very useful in this data set, because they reacted too slowly to context changes (temperature - more than 10 minutes). However, they give an indication of the indoor/outdoor location. A humidity sensor measures not only the humidity of the environment, but also the humidity caused by sweating. In addition, when moving from the cold outdoors to the warm, but dry indoors, the moisture of the warm indoor air concentrates on the surface of the cold sensor and the signal saturates, showing ‘extremely humid’. The signal output corrects as the sensor warms up, but the delay is critical for detecting abrupt changes accurately.

Altitude and *location* (from the GPS) are useful signals that react fast to changes. Altitude was not very useful for the detection of the selected activities. However, the speed was calculated based on the location data and used for the recognition of the football activity.

Activities involving the periodic movements of the body can be detected particularly well using acceleration sensors. Magnetometer signals can also bring extra information in addition to acceleration signals. A good example of this process is the detection of rowing from the acceleration or magnetometer signals. High level activities that do not contain monotonic movements, but irregular combinations of many different movements (for example, football and other ball games, etc.) are more challenging for the activity recognition.

According to the **P6** field trial, people suffering from long-term stress accept self-assessments and measurements over a long period of time at home very well. It is important that the measurements are easy (preferably automatic) so that they are only a minimal burden to the subject. The compliance with measurements can be improved by providing the subject with useful feedback on the measurements and signal trends. Despite their stress status, most people are skilled enough to use the devices, except for people suffering from more severe burnout. The use of new devices can be easier, if they are introduced one by one, with time to learn new functions before another device is introduced.

In Study **P6**, the monitoring of recovery (sleep, exercise) was found to give best indications of stress. A note was taken of the system disturbing the privacy. The *bed sensor* did not disturb most of the subjects, but a few participants found it a little bit disturbing. The *wrist activity monitor* was found unobtrusive; generally, it was not brought up in discussions. A comment was made on its being slightly uncomfortable, because it had to be attached tightly around the wrist. The *Wellness Diary* on the mobile phone was found easy to use. The *heart rate measurement* using a wrist-top computer and a chest belt was found laborious and uncomfortable in long-term monitoring. This was mainly due to the fact that the chest belt was available in one size only. However, a comment was made on the heart rate measurement being especially useful. Some subjects saw the heart rate measurement more useful than others. The *pedometer* and *blood pressure monitor* were found easy to use. The subjects of **P6** estimated monitoring of the blood pressure, weight and exercise as most useful to their health.

Daily self-assessment of stress and the *total stress score from DSP* questionnaires once per month were used as the “ground truths” for measuring stress at home. The DSP was found to measure personal character rather than changes in the dynamic stress status. The daily self-assessment of stress was found to have more day-to-day variability, but difficult to perform over a long period of time.

In feature selection, a human eye is good at selecting features that show changes between different activities. By plotting the feature distributions during different activities (**P1**, Figure 4) or by plotting the feature signals along with annotations in the time domain (**P2**, Figure 3), it is possible to visually inspect, which features show great discriminatory power. It is also possible to arrange the computed features as a function of the target. If the plot still seems noisy, smoothing (moving median filtering in **P6**, Figure 5) can help see trends in data.

The complexity of the classification algorithm has an effect on the battery consumption of an embedded device. The lazy classifiers that perform many operations during the classification phase (such as the KNN classifier) consume more battery power than, for example, the decision tree algorithms that perform only threshold comparisons. Thus, decision tree algorithms can be regarded as energy-efficient algorithms in continuous use. In **P2**, the overall classification accuracy of the automatic decision tree algorithm was below those of the other classifiers. However, the automatic decision tree algorithm has the most stable performance, since it recognizes all the activities with reasonable accuracy, while the other classifiers fail to recognize cycling on an exercise bike. During training, for example, an ANN optimizes the overall performance, but an automatic decision tree algorithm uses a different optimization technique, which does not optimize the overall performance on training data but the classification accuracy of each activity separately. During the training phase, it is important to decide whether to use the same amounts of training data for each activity and ensure the optimal recognition of each activity, or to use the amounts of data that come from everyday activities and optimize the overall classification accuracy. The decision depends on the application.

6.4 Limitations of Studies

In Publication **P1**, the dynamic range of accelerometers was not large enough (±2g). This caused the clipping of the high-variability accelerometer data in some activities (for example, Nordic walking) performed by some volunteers and therefore degraded the classification accuracy of these activities.

In Publication **P2**, the sampling rates were lowered in relation to those used in **P1** to allow longer periods of data collection. However, the sampling rate of 20 Hz for accelerometers was too low. Too low sampling rate made the identification of vigorous activities more difficult, because the sharp peaks caused by impacts were rounded and less separable from less vigorous activities.

In Publications **P1** and **P2**, the absence of accelerometers on the lower body did not allow for the separation of sitting and standing from each other. The accelerometer on the chest was moved to a hip in order to achieve this separation. The move was inspired by a publication by Lee and Mase (Lee & Mase 2002). However, even the sensor location on the hip did not allow for the separation of the two activities, because the belt by which the accelerometer was attached was moving during activities and did not provide accurate enough data. Thus, the conclusion based on **P1** and **P2** is that an accelerometer on a thigh is needed for the automatic separation of sitting and standing.

In Publications **P1**, **P2**, **P3** and **P4**, large data sets with several volunteers and activities were collected and methods were developed based on the large data sets. But the applicability of the results in long-term monitoring in real-life conditions is still uncertain. Especially the results of Study **P4** were limited by the battery life of a wireless sensor, which limited its use to about 1 hour. Thus, it would be necessary to acquire data over full days or even several days with the participation of several different volunteers during different activities in order to prove the robustness of algorithms in end-user products.

In Publication **P5**, the data from physical activities and different levels of mental load were collected in experimental settings. The applicability of the results has not been tested in real-life conditions. The scaling of HR features to a common range would require individual data from the normal mental load in order to work properly. Therefore, the method would not work accurately if a device is taken into use only when already stressed and with altered HR profile.

In Publication **P6**, some hardware problems were experienced with the continuously measuring devices during the study. The problems encountered caused breaks in the process of measurement or data transmission (for example, a modem was damaged due to a thunder storm). Missing data may weaken the significance of the observed variable correlations. Two variables (perceived daily stress and long-term stress profile questionnaire) were used as “ground truth” references for stress. No gold standard reference exists for stress measurements. For the development of wearable stress monitors, a better definition of stress and an agreement on a gold standard reference for stress assessment are needed.

6.5 Future Directions

In Publications **P1**, **P2**, **P3**, **P4** and **P5**, the equipment did not allow the full profiling of people by recordings over the whole day or week, mainly due to limited battery life and memory capacity. In the future, however, this will be possible and recording over whole days and weeks will be possible. This will open possibilities for providing personalized information and guidance, for example, though the use of personal health records. For evaluating the performance of the developed algorithms in full-day and full-week recordings, new, longer measurements are needed.

In Publications **P1**, **P2**, **P4** and **P5**, both time- and frequency domain features derived from accelerometer data were found useful for activity recognition. Combinations of data from different sensor dimensions or locations were not used as inputs. For example, cross-covariance between axes has been found to be a useful feature (Atallah et al. 2010) for activity recognition. Thus, different features for the combining of sensor dimensions and sensor locations can provide useful information, for example, to find if different body parts work in synchrony or not during different activities. In the development of mental load and long-term stress, the adequacy of recovery plays an important role. Study **P6** showed that the assessment of recovery reflects the stress level with the strongest correlation of the measured variables. This is encouraging because for example sleep quality can be assessed unobtrusively, without user interaction, using several different sensors such as a bed sensor, actigraphy or bedroom illumination. In the future, further studies on such new measures of recovery may prove useful in the assessment of stress level.

A current lifestyle in industrialized countries drives to sedentarism and heavy mental load. Early history shows that the lifestyle required more physical activity and less mental load. The optimum would be to find a balance between the load and recovery in both the physical and mental load. The focus of current research is moving from the assessment of physical activity towards the measurement and assessment of the volume of sedentarism (Chastin & Granat 2010). This is because several studies have pointed out how the length of sedentary periods may affect several risk factors. Even short breaks during the long sedentary periods have been shown to decrease cardiometabolic risk factors (Healy & Owen 2010).

7. Conclusions

In this thesis, practical methods for the automatic recognition of physical activities and automatic assessment of energy expenditure as well as mental load and long-term stress using data from wearable sensors and other self-measurements were studied. Signal interpretation and classification methods were developed and the most information-rich sensors identified. The methods were developed based on large, realistic and annotated data libraries collected as part of the studies. Applicability of the methods was evaluated with the collected realistic data and state-of-the-art scientific results were obtained. The following conclusions can be drawn from the studies carried out within the framework of this thesis:

- Careful selection of sensors, sensor locations and input features plays a more critical role in successful classification than the selection of a classifier.
- Computational complexity of the classification phase of a classifier has an impact on the power consumption of the hosting mobile terminal.
- Accelerometer signal is the most useful signal for recognition of the type of activity as well as for estimating the energy expenditure with everyday activities.
- Automatic activity recognition enables more accurate automatic detection of high day-time mental load by focusing the analysis on sedentary periods.
- Normalized heart rate and variability of respiratory rate signals allow detection of high day-time mental load with reasonable accuracy.
- In addition to the measurement of the day-time mental load, the measurement of recovery at night is useful for the assessment of long-term stress.
- No single sensor is clearly better than others in monitoring long-term stress.
- Useful measures for automatic assessment of long-term stress are
    - sleep length estimated using wrist actigraphy,
    - night illumination level estimated using the bedroom illumination sensor and
    - average night heart rate estimated using the bed sensor.

References

Ahola, K., Väänänen, A., Koskinen, A., Kouvonen, A. & Shirom, A. 2010. Burnout as a predictor of all-cause mortality among industrial employees: A 10-year prospective register-linkage study. Journal of Psychosomatic Research, Vol. 69, No. 1, pp. 51–57.

Ainsworth, B.E., Haskell, W.L., Leon, A.S., Jacobs, D.R., Montoye, H.J., Sallis, J.F. & Paffenbarger, R.S. 1993. Compendium of physical activities: classification of energy costs of human physical activities. Medicine and Science in Sports and Exercise, Vol. 25, No. 1, pp. 71.

Ainsworth, B.E., Haskell, W.L., Whitt, M.C., Irwin, M.L., Swartz, A.M., Strath, S.J., O'Brien, W.L., Bassett Jr, D.R., Schmitz, K.H. & Emplaincourt, P.O. 2000. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and Science in Sports and Exercise, Vol. 32, No. 9 Suppl, pp. S498.

Ainsworth, B.E. 2009. How do I measure physical activity in my patients? Questionnaires and objective methods. British Journal of Sports Medicine, Vol. 43, No. 1, pp. 6–9.

Allen, F.R., Ambikairajah, E., Lovell, N.H. & Celler, B.G. 2006. Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. Physiological Measurement, Vol. 27, No. 10, pp. 935–951.

Amft, O. & Lukowicz, P. 2009. From Backpacks to Smartphones: Past, Present, and Future of Wearable Computers. IEEE Pervasive Computing, Vol. 8, No. 3, pp. 8–13.

Aminian, K., Rezakhanlou, K., De Andres, E., Fritsch, C., Leyvraz, P.F. & Robert, P. 1999. Temporal feature estimation during walking using miniature accelerometers: An analysis of gait improvement after hip arthroplasty. Medical and Biological Engineering and Computing, Vol. 37, No. 6, pp. 686–691.

Antila, K., van Gils, M., Merilahti, J. & Korhonen, I. 2005. Associations of psychological self-assessments and HRV in long-term measurements at home. Proceedings of the European Medical and Biological Engineering Conference 2005 (EMBEC'05), Prague, Czech Republic, 20–25 Nov 2005. Pp. 7–11.

Arnrich, B., Setz, C., La Marca, R., Tröster, G. & Ehlert, U. 2010. What does your chair know about your stress level? IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 2, pp. 207–214.

Aromaa, A. & Koskinen, S. 2004. Health and Functional Capacity in Finland – Baseline Results of the Health 2000 Health Examination Survey, National Public Health Institute, Helsinki.

Aronen, E.T., Paavonen, E.J., Fjällberg, M., Soininen, M. & Törrönen, J. 2000. Sleep and psychiatric symptoms in school-age children. Journal of the American Academy of Child and Adolescent Psychiatry, Vol. 39, No. 4, pp. 502–508.

Atallah, L., Lo, B., King, R. & Yang, G.Z. 2010. Sensor placement for activity detection using wearable accelerometers. Proceedings of the International Conference on Body Sensor Networks, BSN 2010. Pp. 24–29.

Bao, L. & Intille, S.S. 2004. Activity recognition from user-annotated acceleration data. Proceedings of the 2nd International Conference on Pervasive Computing (Pervasive 2004). A. Ferscha & M. Friedemann, (eds.). Springer, Berlin/Heidelberg, April 21–23, 2004. Pp. 1–17.

Bonomi, A.G., Plasqui, G., Goris, A.H.C. & Westerterp, K.R. 2010. Estimation of free-living energy expenditure using a novel activity monitor designed to minimize obtrusiveness. Obesity, Vol. 18, No. 9, pp. 1845–1851.

Bouchard, C., Blair, S.N. & Haskell, W.L. (eds.). 2007. Physical Activity and Health, Human Kinetics, Champaign, IL, USA.

Bouten, C.V.C., Koekkoek, K.T.M., Verduin, M., Kodde, R. & Janssen, J.D. 1997. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE Transactions on Biomedical Engineering, Vol. 44, No. 3, pp. 136–147.

Bouten, C.V.C., Westerterp, K.R., Verduin, M. & Janssen, J.D. 1994. Assessment of energy expenditure for physical activity using a triaxial accelerometer. Medicine and Science in Sports and Exercise, Vol. 26, No. 12, pp. 1516–1523.

Bussmann, J.B.J., Damen, L. & Stam, H.J. 2000. Analysis and decomposition of signals obtained by thigh-fixed uni-axial accelerometry during normal walking. Medical and Biological Engineering and Computing, Vol. 38, No. 6, pp. 632–638.

Carroll, R., Cnossen, R., Schnell, M. & Simons, D. 2007. Continua: An Interoperable Personal Healthcare Ecosystem. IEEE Pervasive Computing, pp. 90–94.

Caspersen, C.J., Powell, K.E. & Christenson, G.M. 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Reports, Vol. 100, No. 2, pp. 126–131.

Celler, B.G., Lovell, N.H. & Basilakis, J. 2003. Using information technology to improve the management of chronic disease. Medical Journal of Australia, Vol. 179, No. 5, pp. 242–246.

Chastin, S.F.M., Dall, P.M., Tigbe, W.W., Grant, M.P., Ryan, C.G., Rafferty, D. & Granat, M.H. 2009. Compliance with physical activity guidelines in a group of UK-based postal workers using an objective monitoring technique. European Journal of Applied Physiology, Vol. 106, No. 6, pp. 893–899.

Chastin, S.F.M. & Granat, M.H. 2010. Methods for objective measure, quantification and analysis of sedentary behaviour and inactivity. Gait & Posture, Vol. 31, No. 1, pp. 82–86.

Cleland, J.G.F., Louis, A.A., Rigby, A.S., Janssens, U. & Balk, A.H.M.M. 2005. Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: the Trans-European Network-Home-Care Management System (TEN-HMS) study. Journal of the American College of Cardiology, Vol. 45, No. 10, pp. 1654–1664.

Cooley, J. & Tukey, J.W. 1965. An Algorithm for the Machine Computation of the Complex Fourier Series. Mathematics of Computation, Vol. 19, pp. 297–301.

Crouter, S.E., Clowers, K.G. & Bassett Jr., D.R. 2006. A novel method for using accelerometer data to predict energy expenditure. Journal of Applied Physiology, Vol. 100, No. 4, pp. 1324–1331.

Crouter, S.E., Schneider, P.L., Karabulut, M. & Bassett Jr., D.R. 2003. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. Medicine and Science in Sports and Exercise, Vol. 35, No. 8, pp. 1455–1460.

Derogatis, L.R. 1987. The Derogatis Stress Profile (DSP): quantification of psychological stress. Advances in Psychosomatic Medicine, Vol. 17, pp. 30–54.

Duda, R.O., Hart, P.E. & Stork, D.G. 2001. Pattern Classification, Wiley, New York.

Ekstedt, M., Åkerstedt, T. & Söderström, M. 2004. Microarousals During Sleep Are Associated with Increased Levels of Lipids, Cortisol, and Blood Pressure. Psychosomatic Medicine, Vol. 66, No. 6, pp. 925–931.

Esliger, D.W. & Tremblay, M.S. 2007. Physical activity and inactivity profiling: The next generation. Applied Physiology, Nutrition and Metabolism, Vol. 32, No. SUPPL. 2E, pp. S195–S207.

Fahrenberg, J., Foerster, F., Smeja, M. & Müller, W. 1997. Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. Psychophysiology, Vol. 34, No. 5, pp. 607–612.

Foerster, F. & Fahrenberg, J. 2000. Motion pattern and posture: Correctly assessed by calibrated accelerometers. Behavior Research Methods, Instruments, and Computers, Vol. 32, No. 3, pp. 450–456.

Foerster, F., Smeja, M. & Fahrenberg, J. 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Computers in Human Behavior, Vol. 15, No. 5, pp. 571–583.

Fogelholm, M., Suni, J., Rinne, M., Oja, P. & Vuori, I. 2005. Physical activity pie: a graphical presentation integrating recommendations for fitness and health. Journal of Physical Activity & Health, Vol. 2, No. 4, pp. 391.

Fukakusa, M., Sato, T. & Furuhata, H. 1998. Use of an accelerometer to measure coughing. Journal of the Japanese Respiratory Society, Vol. 36, No. 4, pp. 343–346.

Gibbs-Smith, C. 1985. The Inventions of Leonardo da Vinci, Peerage Books, London.

Gould, R., Nyman, H. & Lampi, H. 2010. Masennukseen perustuvat työkyvyttömyyseläkkeet meillä ja muualla. In: Myöhemmin eläkkeelle – selvityksiä ja laskelmia. Finnish Centre of Pensions, Helsinki. Pp. 247–254. (In Finnish)

Haykin, S. 1999. Neural Networks: A Comprehensive Foundation (2nd Edition), Prentice-Hall International, London, UK.

Haynes, S.N. & Yoshioka, D.T. 2007. Clinical assessment applications of ambulatory biosensors. Psychological Assessment, Vol. 19, No. 1, pp. 44–57.

Healey, J.A. & Picard, R.W. 2005. Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems, Vol. 6, No. 2, pp. 156–166.

Healy, G.N. & Owen, N. 2010. Sedentary Behaviour and Biomarkers of Cardiometabolic Health Risk in Adolescents: An Emerging Scientific and Public Health Issue. Revista espanola de cardiologia, Vol. 63, No. 3, pp. 261–264.

Hester, T., Sherrill, D.M., Hamel, M., Perreault, K., Boissy, P. & Bonato, P. 2006. Identification of tasks performed by stroke patients using a mobility assistive device. Annual International Conference of IEEE Engineering in Medicine and Biology Society. P. 1501.

Hiltunen, M., Käkönen, K., Kannisto, J., Pellinen, M., Lybäck, K. & Goebel, R. 2008. Pensioners and Insured in Finland 2008. Joint publication of Finnish Centre for Pensions, The Local Government Pensions Institution and State Treasury, Helsinki.

Honkonen, T., Ahola, K., Pertovaara, M., Isometsä, E., Kalimo, R., Nykyri, E., Aromaa, A. & Lönnqvist, J. 2006. The association between burnout and physical illness in the general population-results from the Finnish Health 2000 Study. Journal of Psychosomatic Research, Vol. 61, No. 1, pp. 59–66.

Huiku, M., Uutela, K., van Gils, M., Korhonen, I., Kymäläinen, M., Meriläinen, P., Paloheimo, M., Rantanen, M., Takala, P., Viertiö-Oja, H. & Yli-Hankala, A. 2007. Assessment of surgical stress during general anaesthesia. British Journal of Anaesthesia, Vol. 98, No. 4, pp. 447–455.

Intille, S., Tapia, E., Rondoni, J., Beaudin, J., Kukla, C., Agarwal, S., Bao, L. & Larson, K. 2003. Tools for studying behavior and technology in natural settings. Lecture Notes in Computer Science, Vol. 2864, pp. 157–174.

Jean-Louis, G., Kripke, D.F., Mason, W.J., Elliott, J.A. & Youngstedt, S.D. 2001. Sleep estimation from wrist movement quantified by different actigraphic modalities. Journal of Neuroscience Methods, Vol. 105, No. 2, pp. 185–191.

Jean-Louis, G., Von Gizycki, H., Zizi, F., Fookson, J., Spielman, A., Nunes, J., Fullilove, R. & Taub, H. 1996. Determination of sleep and wakefulness with the actigraph data analysis software (ADAS). Sleep, Vol. 19, No. 9, pp. 739–743.

Johnson, R.W. & Shore, J.E. 1984. Which is the better entropy expression for speech processing: minus S LOG S or LOG S? IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-32, No. 1, pp. 129–137.

Kalimo, R. & Toppinen, S. 1997. Työuupumus Suomen työikäisellä väestöllä (Burnout in the Finnish Working-age Population). Finnish Institute of Occupational Health, Helsinki.

Karantonis, D.M., Narayanan, M.R., Mathie, M., Lovell, N.H. & Celler, B.G. 2006. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 1, pp. 156–167.

Karemaker, J.M. & Lie, K.I. 2000. Heart rate variability: A telltale of health or disease. European Heart Journal, Vol. 21, No. 6, pp. 435–437.

Katsis, C.D., Ganiatsas, G. & Fotiadis, D.I. 2006. An integrated telemedicine platform for the assessment of affective physiological states. Diagnostic Pathology, Vol. 1, No. 1, pp. 16–19.

Kesäniemi, A., Kettunen, J., Ketola, E., Kujala, U., Kukkonen-Harjula, K., Lakka, T., Rauramaa, R., Rauramo, I., Tikkanen, H. & Vuori, I. 2010. Aikuisten liikunta – Käypä

Hoito suositus (Physical Activity and Exercise Training for Adults in Sickness and in Health – Finnish Current Care Summary). Duodecim, Helsinki.

Kim, J. & André, E. 2008. Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 12, pp. 2067–2083.

Kinnunen, M.L. 2005. Allostatic load in relation to psychosocial stressors and health. University of Jyväskylä. (Dissertation)

Kinnunen, M., Rusko, H., Feldt, T., Kinnunen, U., Juuti, T., Myllymäki, T., Laine, K., Hakkarainen, P. & Louhevaara, V. 2006. Stress and relaxation based on heart rate variability: Associations with self-reported mental strain and differences between waking hours and sleep. Proceedings of the 38th Annual Congress of Nordic Ergonomics Society, NES 2006. Pp. 136–139.

Könönen, V., Mäntyjärvi, J., Similä, H., Pärkkä, J. & Ermes, M. 2010. Automatic feature selection for context recognition in mobile devices. Pervasive and Mobile Computing, Vol. 6, No. 2, pp. 181–197.

Korhonen, I., Pärkkä, J. & Van Gils, M. 2003. Health Monitoring in the Home of the Future. IEEE Engineering in Medicine and Biology Magazine, Vol. 22, No. 3, pp. 66–73.

Kudielka, B.M., Bellingrath, S. & Hellhammer, D.H. 2006. Cortisol in burnout and vital exhaustion: An overview. Giornale italiano di medicina del lavoro ed ergonomia, Vol. 28, No. 1, Suppl. 1, pp. 34–42.

Kushida, C.A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O. & Dement, W.C. 2001. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Medicine, Vol. 2, No. 5, pp. 389–396.

Kusserow, M., Amft, O. & Tröster, G. 2008. Analysis of heart stress response for a public talk assistant system. Lecture Notes in Computer Science, Vol. 5355, pp. 326–342.

Lafortune, M.A. 1991. Three-dimensional acceleration of the tibia during walking and running. Journal of Biomechanics, Vol. 24, No. 10, pp. 877–886.

Lagerros, Y.T. & Lagiou, P. 2007. European Journal of Epidemiology, Vol. 22, No. 6, pp. 353–362.

Laporte, R.E., Montoye, H.J. & Caspersen, C.J. 1985. Assessment of physical activity in epidemiologic research: Problems and prospects. Public Health Reports, Vol. 100, No. 2, pp. 131–146.

Lappalainen, R., Pulkkinen, P., van Gils, M., Pärkkä, J. & Korhonen, I. 2005. Long-term self-monitoring of weight: a case study. Cognitive Behaviour Therapy, Vol. 34, No. 2, pp. 108–114.

Laurila, K., Pylvänäinen, T., Silanto, S. & Virolainen, A. 2005. Wireless Motion Bands. The Seventh International Conference on Ubiquitous Computing, Ubicomp 2005 Workshop: Ubiquitous Computing to Support Monitoring, Measuring, and Motivating Exercise.

Leavitt, M.O. (ed.) 2008. 2008 Physical Activity Guidelines for Americans. U.S. Department of Health and Human Services, Washington D.C., the USA.

Lee, Y. & Lippmann, R.P. 1990. Practical characteristics of neural network and conventional pattern classifiers on artificial and speech problems. In: Advances in Neural Information Processing Systems 2, ed. Touretzky, D.S. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 168–177.

Lee, S.W. & Mase, K. 2002. Activity and location recognition using wearable sensors. IEEE Pervasive Computing, Vol. 1, No. 3, pp. 24–32.

Lehtokangas, M. 1995. Modeling with Layered Feedforward Neural Networks. Tampere University of Technology.

Lemke, M.R., Puhl, P. & Broderick, A. 1999. Motor activity and perception of sleep in depressed patients. Journal of Psychiatric Research, Vol. 33, No. 3, pp. 215–224.

Lester, J., Choudhury, T., Kern, N., Borriello, G. & Hannaford, B. 2005. A hybrid discriminative/generative approach for modeling human activities. Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. P. 766.

Lester, J., Choudhury, T. & Borriello, G. 2006. A practical approach to recognizing physical activities. Lecture Notes in Computer Science, Vol. 3968, pp. 1–16.

Lindholm, H. & Gockel, M. 2000. Stressin elinvaikutusten mittaaminen. Duodecim, Vol. 116, pp. 2259–2265. (In Finnish)

Lippmann, R.P. 1989. Pattern Classification using Neural Networks. IEEE Communications Magazine, Vol. 27, No. 11, pp. 47–64.

Lo, B., Atallah, L., Aziz, O., ElHew, M., Darzi, A. & Yang, G.Z. 2007. Real-time pervasive monitoring for postoperative care. 4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007). Pp. 122–127.

Lorig, K.R., Sobel, D.S., Ritter, P.L., Laurent, D. & Hobbs, M. 2001. Effect of a self-management program on patients with chronic disease. Effective Clinical Practice, Vol. 4, No. 6, pp. 256–262.

Lymberis, A. 2004. Research and development of smart wearable health applications: the challenge ahead. In: Wearable eHealth Systems for Personalised Health Management – State of the Art and Future Challenges IOS Press. The Netherlands. Pp. 155–161.

Martin, S., Schneider, B., Heinemann, L., Lodwig, V., Kurth, H.J., Kolb, H. & Scherbaum, W.A. 2006. Self-monitoring of blood glucose in type 2 diabetes and long-term outcome: an epidemiological cohort study. Diabetologia, Vol. 49, No. 2, pp. 271–278.

Maslach, C. & Jackson, S.E. 1981. The measurement of experienced burnout. Journal of Organizational Behavior, Vol. 2, No. 2, pp. 99–113.

Mathie, M.J., Coster, A.C.F., Lovell, N.H. & Celler, B.G. 2004. Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiological Measurement, Vol. 25, No. 2, pp. R1–R20.

Mathie, M.J., Coster, A.C.F., Lovell, N.H. & Celler, B.G. 2003. Detection of daily physical activities using a triaxial accelerometer. Medical and Biological Engineering and Computing, Vol. 41, No. 3, pp. 296–301.

Maurer, U., Rowe, A., Smailagic, A. & Siewiorek, D. 2006. Location and activity recognition using eWatch: A wearable sensor platform. Lecture Notes in Computer Science, Vol. 3864, pp. 86–102.

McGinnis, J.M., Williams-Russo, P. & Knickman, J.R. 2002. The case for more active policy attention to health promotion. Health Affairs, Vol. 21, No. 2, pp. 78–93.

Meijer, G.A.L. 1990. Physical Activity. Implications for human energy metabolism. Rijksuniversiteit Limburg, Maastricht, the Netherlands. (Dissertation)

Michie, S. & Williams, S. 2003. Reducing work related psychological ill health and sickness absence: A systematic literature review. Occupational and Environmental Medicine, Vol. 60, No. 1, pp. 3–9.

OECD 2009. Sickness, Disability and Work: Keeping on Track in the Economic Downturn – Background Paper, OECD.

Oppenheim, A.V., Schafer, R.W. & Buck, J.R. 1999. Discrete-Time Signal Processing. 2nd edn, Prentice-Hall International.

Pate, R.R., Pratt, M., Blair, S.N., Haskell, S.L., Macera, C.A., Bouchard, C., Buchner, D., Ettinger, W., Health, G.W., King, A.C., Kriska, A., Leon, A.S., Marcus, B.H., Morris, J., Pfaffenbarger, R.S., Patrick, K., Pollock, M.L., Rippe, J.M., Sallis, J. & Wilmore, J.H. 1995. Physical Activity and Public Health – A Recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. Journal of the American Medical Association, Vol. 273, pp. 402–407.

Pirttikangas, S., Fujinami, K. & Nakajima, T. 2006. Feature selection and activity recognition from wearable sensors. Lecture Notes in Computer Science, Vol. 4239, pp. 516–527.

Pollak, C.P., Tryon, W.W., Nagaraja, H. & Dzwonczyk, R. 2001. How accurately does wrist actigraphy identify the states of sleep and wakefulness? Sleep, Vol. 24, No. 8, pp. 957–965.

Preece, S.J., Goulermas, J.Y., Kenney, L.P.J., Howard, D., Meijer, K. & Crompton, R. 2009. Activity identification using body-mounted sensors – A review of classification techniques. Physiological Measurement, Vol. 30, No. 4, pp. R1–R33.

Puska, P., Benaziza, H. & Porter, D. 2004. Physical Activity, WHO [Online: http://www.who.int/dietphysicalactivity/media/en/gsfs_pa.pdf].

Ravi, N., Dandekar, N., Mysore, P. & Littman, M.L. 2005. Activity recognition from accelerometer data. Proceedings of the National Conference on Artificial Intelligence, pp. 1541–1546.

Riedmiller, M. & Braun, H. 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. Proceedings of the IEEE International Conference on Neural Networks. Pp. 586–591.

Sanches, P., Höök, K., Vaara, E., Weymann, C., Bylund, M., Ferreira, P., Peira, N. & Sjölinder, M. 2010. Mind the body!: designing a mobile stress management application encouraging personal reflection. Proceedings of the 8th ACM Conference on Designing Interactive SystemsACM, New York. Pp. 47–56.

Saranummi, N. 2009. In the spotlight: Health information systems: PHR and value based healthcare. IEEE Reviews in Biomedical Engineering, Vol. 2, pp. 15–17.

Sekine, M., Tamura, T., Togawa, T. & Fukui, Y. 2000. Classification of waist-acceleration signals in a continuous walking record. Medical Engineering and Physics, Vol. 22, No. 4, pp. 285–291.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G. & Ehlert, U. 2010. Discriminating stress from cognitive load using a wearable eda device. IEEE Transactions on Information Technology in Biomedicine, Vol. 14, No. 2, pp. 410–417.

Shannon, C.E. 1948. A Mathematical Theory of Communication. Bell System Technical Journal, Vol. 27, pp. 379–423.

Shapiro, D., Jamner, L.D., Goldstein, I.B. & Delfino, R.J. 2001. Striking a chord: Moods, blood pressure, and heart rate in everyday life. Psychophysiology, Vol. 38, No. 02, pp. 197–204.

Shirom, A. 2005. Reflections on the study of burnout. Work and Stress, Vol. 19, No. 3, pp. 263–270.

Smidt, G.L., Arora, J.S. & Johnston, R.C. 1971. Accelerographic analysis of several types of walking. American Journal of Physical Medicine, Vol. 50, No. 6, pp. 285–300.

Sun, M. & Hill, J.O. 1993. A method for measuring mechanical work and work efficiency during human activities. Journal of Biomechanics, Vol. 26, No. 3, pp. 229–241.

Tang, P.C., Ash, J.S., Bates, D.W., Overhage, J.M. & Sands, D.Z. 2006. Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. Journal of the American Medical Informatics Association, Vol. 13, No. 2, pp. 121–126.

Troiano, R.P., Berrigan, D., Dodd, K.W., Mâsse, L.C., Tilert, T. & Mcdowell, M. 2008. Physical activity in the United States measured by accelerometer. Medicine and Science in Sports and Exercise, Vol. 40, No. 1, pp. 181–188.

Tudor-Locke, C. 2003. Manpo-Kei: The Art and Science of Step Counting: How to Be Naturally Active and Lose Weight, Trafford, Victoria, Canada.

Tuomisto, M.T., Terho, T., Korhonen, I., Lappalainen, R., Tuomisto, T., Laippala, P. & Turjanmaa, V. 2006. Diurnal and weekly rhythms of health-related variables in home recordings for two months. Physiology and Behavior, Vol. 87, No. 4, pp. 650–658.

Uiterwaal, M., Glerum, E.B.C., Busser, H.J. & Van Lummel, R.C. 1998. Ambulatory monitoring of physical activity in working situations, a validation study. Journal of Medical Engineering and Technology, Vol. 22, No. 4, pp. 168–172.

Van Amelsvoort, L.G.P.M., Schouten, E.G., Maan, A.C., Swenne, C.A. & Kok, F.J. 2000. Occupational determinants of heart rate variability. International Archives of Occupational and Environmental Health, Vol. 73, No. 4, pp. 255–262.

Van Emmerik, R.E.A. & Wagenaar, R.C. 1996. Dynamics of movement coordination and tremor during gait in Parkinson's disease. Human Movement Science, Vol. 15, No. 2, pp. 203–235.

Van Someren, E.J.W., Lazeron, R.H.C., Vonk, B.F.M., Mirmiran, M. & Swaab, D.F. 1996. Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects. Journal of Neuroscience Methods, Vol. 65, No. 1, pp. 55–62.

Veltink, P.H., Bussmann, H.B.J., De Vries, W., Martens, W.L.J. & Van Lummel, R.C. 1996. Detection of static and dynamic activities using uniaxial accelerometers. IEEE Transactions on Rehabilitation Engineering, Vol. 4, No. 4, pp. 375–385.

Veltink, P.H., Olde Engberink, E.G., van Hilten, B.J., Dunnewold, R. & Jacobi, C. 1995. Towards a new method for kinematic quantification of bradykinesia in patients with Parkinson's disease using triaxial accelerometry. Proceedings of the Annnual International Conference of the IEEE Engineering in Medicine and Biology Society. Pp. 1303–1304.

Verberk, W.J., Kroon, A.A., Lenders, J.W.M., Kessels, A.G.H., Van Montfrans, G.A., Smit, A.J., Van Der Kuy, P.-M., Nelemans, P.J., Rennenberg, R.J.M.W., Grobbee, D.E., Beltman, F.W., Joore, M.A., Brunenberg, D.E.M., Dirksen, C., Thien, T. & De Leeuw, P.W. 2007. Self-measurement of blood pressure at home reduces the need for antihypertensive drugs: A randomized, controlled trial. Hypertension, Vol. 50, No. 6, pp. 1019–1025.

Viertiö-Oja, H., Maja, V., Särkelä, M., Talja, P., Tenkanen, N., Tolvanen-Laakso, H., Paloheimo, M., Vakkuri, A., Yli-Hankala, A. & Meriläinen, P. 2004. Description of the Entropy™ algorithm as applied in the Datex-Ohmeda 5/5™ Entropy Module. Acta Anaesthesiologica Scandinavica, Vol. 48, No. 2, pp. 154–161.

Webster, J.G. 2010. Medical Instrumentation: Application and Design, 4th edn, Wiley, USA.

Weiss, A.R., Johnson, N.L., Berger, N.A. & Redline, S. 2010. Validity of activity-based devices to estimate sleep. Journal of Clinical Sleep Medicine, Vol. 6, No. 4, pp. 336–342.

Westerterp, K.R. 2001. Pattern and intensity of physical activity. Nature, Vol. 410, No. 6828, pp. 539.

Whitney, A.W. 1971. Direct method of nonparametric measurement selection. IEEE Transactions on Computers, Vol. C-20, No. 9, pp. 1100–1103.

WHO 2010. Global recommendations on physical activity for health, [Online: http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf].

Wilhelm, F.H., Pfaltz, M.C. & Grossman, P. 2006. Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion. Interacting with Computers, Vol. 18, No. 2, pp. 171–186.

Witten, I.H. & Frank, E. 1999. Data mining: Practical machine learning tools and techniques with Java implementations, Morgan Kaufmann, San Francisco, CA, USA.

Woodward, M.I. & Cunningham, J.L. 1993. Skeletal accelerations measured during different exercises. Archive: Proceedings of the Institution of Mechanical Engineers. Part H, Journal of Engineering in Medicine 1989–1996, Vol. 207, No. H2, pp. 79–85.

Yang, J.Y., Wang, J.S. & Chen, Y.P. 2008. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. Pattern Recognition Letters, Vol. 29, No. 16, pp. 2213–2220.

Zhai, J. & Barreto, A. 2006. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. Proceedings of the Annual International Conference of IEEE Engineering in Medicine and Biology Society. Pp. 1355–1358.

Errata in Publications

Publication 1, page 3, Table 1 and page 4, paragraphs 3 and 4:
Correct abbreviation of blood oxygen saturation signal obtained using pulse oximeter is **SpO_2**.

PUBLICATION **P1**

Activity Classification Using Realistic Data from Wearable Sensors

In: IEEE Transactions on Information Technology in Biomedicine 2006. Vol. 10, No. 1, pp. 119–128.

Activity Classification Using Realistic Data From Wearable Sensors

Juha Pärkkä, Miikka Ermes, Panu Korpipää, *Member, IEEE*, Jani Mäntyjärvi, Johannes Peltola, and Ilkka Korhonen, *Member, IEEE*

***Abstract*—Automatic classification of everyday activities can be used for promotion of health-enhancing physical activities and a healthier lifestyle. In this paper, methods used for classification of everyday activities like walking, running, and cycling are described. The aim of the study was to find out how to recognize activities, which sensors are useful and what kind of signal processing and classification is required. A large and realistic data library of sensor data was collected. Sixteen test persons took part in the data collection, resulting in approximately 31 h of annotated, 35-channel data recorded in an everyday environment. The test persons carried a set of wearable sensors while performing several activities during the 2-h measurement session. Classification results of three classifiers are shown: custom decision tree, automatically generated decision tree, and artificial neural network. The classification accuracies using leave-one-subject-out cross validation range from 58 to 97% for custom decision tree classifier, from 56 to 97% for automatically generated decision tree, and from 22 to 96% for artificial neural network. Total classification accuracy is 82% for custom decision tree classifier, 86% for automatically generated decision tree, and 82% for artificial neural network.**

***Index Terms*—Activity classification, context awareness, physical activity, wearable sensors.**

I. Introduction

PHYSICAL inactivity is a health risk that many people in both developed and developing countries are facing today. According to World Health Organization (WHO), at least 60% of the world's population fails to achieve the minimum recommendation of 30 min moderate intensity physical activity daily [13]. The main reason for not achieving this basic level of physical activity is that the level of activity required in work, in travel, and at home is decreasing with sedentary work and with the advent of technologies that are designed to ease home activities and traveling. The physical activities on free time are insufficient or too irregular to achieve the weekly goal. Physical inactivity is known to contribute in many chronic diseases, such as cardiovascular disease, type 2 diabetes, and possibly certain types of cancer and osteoporosis [2], [3]. As the population is rapidly aging in many countries, promotion of a healthier lifestyle, especially for the elderly population, can provide substantial savings in future health care costs.

By following the minimum recommendation, many health benefits can be obtained, when compared with completely inactive people [4]–[6]. The basic level of physical activity helps, for example, in managing weight, in lowering blood pressure, in increasing the level of the good high-density lipoprotein (HDL) cholesterol, in improving sugar tolerance, and in changing hormone levels to a direction more suitable for preventing cancer [3], [7]. The basic level of physical activity can be achieved by everyday activities like walking at work, shopping, gardening, cleaning, etc. The 30-min daily physical activity targets to at least 1000 kcal energy expenditure weekly. The only limitation in achieving the goal is that the daily 30-min physical activity must be collected in continuous periods of a minimum 10 min.

Level of daily physical activity can be measured objectively by measuring energy expenditure. The accelerometer signal has been used previously to estimate energy expenditure, and the estimate has been shown to correlate well with true energy expenditure [8]. Although achieving the minimum recommendation of physical activity brings many health benefits, even more health benefits can be achieved by taking part in a more vigorous [5] and a wider spectrum of physical activities. For example, endurance-enhancing activities and activities maintaining flexibility and muscular strength bring health benefits that are not achieved with basic activity [3]. Endurance can be enhanced, e.g., with energetic walking, jogging, cycling, and rowing. Activities maintaining functions of the musculoskeletal system are, e.g., ball games, gym, and dancing. Thus, in addition to daily energy expenditure, activity types play an important role in overall well being and health.

Accelerometers have been shown to be adequate for activity recognition. The studies using accelerometry for monitoring human movement have been recently reviewed in [9] and [10]. In laboratory settings, the most prevalent everyday activities (sitting, standing, walking, and lying) have been successfully recognized with accelerometers [11]–[15]. However, applicability of these results to out-of-lab monitoring is unclear. For example, in [15] the recognition accuracy of nine patterns decreased from 95.8% to 66.7% as the recordings were shifted outside the laboratory. Also, recognition of different activities involving dynamic motion has not yet been studied thoroughly. In a few studies data have been collected outside the laboratory. In [15] 24 subjects spent approximately 50 min outside laboratory. Accelerometers were placed on sternum, wrist, thigh, and lower leg. Nine patterns (sitting, standing, lying, sitting and

Manuscript received December 21, 2004; revised June 16, 2005. This work was supported in part by National Technology Agency of Finland, Nokia, Suunto and Clothing+.

J. Pärkkä, M. Ermes, and I. Korhonen are with VTT Information Technology, 33101 Tampere, Finland (e-mail: juha.parkka@vtt.fi; miikka.ermes@vtt.fi; ilkka.korhonen@vtt.fi).

P. Korpipää, J. Mäntyjärvi, and J. Peltola are with VTT Electronics, 90571 Oulu, Finland (e-mail: jani.mantyjarvi@vtt.fi; panu.korpipaa@vtt.fi; johannes.peltola@vtt.fi).

Digital Object Identifier 10.1109/TITB.2005.856863

talking, sitting and operating PC, walking, stairs up, stairs down, and cycling) were recognized from presegmented data with an overall accuracy of 66.7%. In [16] five biaxial accelerometers attached to hip, wrist, arm, ankle, and thigh were used to recognize 20 everyday activities such as walking, watching TV, brushing teeth, vacuuming, etc. From 82 to 160 min of data were collected from 20 subjects and a decision tree classifier was used for classification. Recognition accuracies ranged from 41 to 97% for different activities.

Many research groups have recently studied activity recognition as part of context awareness research [16]–[22]. Context sensing and use of context information is an important part of the ubiquitous computing scenario [23]–[25]. Context sensing aims at giving a computing device (e.g., cellular phone, wrist-top computer, or a device integrated into clothes) senses, with which it becomes aware of its surroundings. With the senses the device is capable of measuring its user and environment and it becomes context aware. The context describes the situation or status of the user or device. Different devices can use the context information in different ways, e.g., for adapting its user interface, for offering relevant services and information, for annotating digital diary (e.g., energy expenditure), etc. Location and time belong to the group of the most important contexts and the use of these contexts has been studied extensively. However, to recognize the physical activities of a person, a sensor-based approach is needed.

Our vision in automatic classification of physical activities is to contribute to long-term monitoring of health and to a more active lifestyle. The application we have in mind can be called an "activity diary". The diary would show the user which activities he did during the day and what were the daily cumulative durations of each activity. When the user is shown this information, he can draw the conclusions himself and adjust his behavior accordingly. This model is called the *behavioral feedback model* [26]. This model is being successfully used, e.g., in weight management programs. On the other hand the activity diary information can be utilized by context-aware services and devices that offer adapted information or adapt their user interface (UI) based on the user's activity type.

In this work our aim was to study activity classification, which are the most information-rich sensors and what kind of signal processing and classification methods should be used for activity classification. We took a data-oriented and empirical approach and collected a large data library of realistic data. In this paper, we describe methods for automatic activity classification from data collected with body-worn sensors.

II. Methods

A. Data Collection

The goal of our data collection was to assess the feasibility and accuracy of context recognition based on realistic data. We collected a large data library of realistic context data with many different sensors (accelerometers, physiological sensors, etc.) and with many test persons. The collected data were then used in development of context recognition algorithms. A data collection system was developed for sensing and storing context-related data in real-life conditions. Only the sensors are small in size that were applicable to ambulatory measurements were used. The data were stored on a rugged, compact PC (Databrick III, Datalux Corporation, Winchester, VA, USA) and on a flash-card-memory-based, 19-channel recorder (Embla A10, Medcare, Reykjavik, Iceland). Additionally, two stand-alone devices were used: Global Positioning System (GPS) recorder (Garmin eTrex Venture, Garmin Ltd., Olathe, Kansas, USA) and wrist-top computer that measured heart rate and altitude (Suunto X6HR, Suunto Oy, Vantaa, Finland). The PC and recorder were placed into a normal rucksack (dimensions: 40 cm $\times$ 30 cm $\times$ 10 cm, weight 5 kg with the equipment) that the test persons carried during the measurement sessions. The sensors were put on the test person with help of an assistant before the start of the measurement session. The system measured 18 different quantities from the user and his environment (Table I). Some of the quantities were measured with multiple sensors, which resulted in altogether 22 signals and 35 channels of data.

During the measurement sessions, the test persons followed a scenario (Table II) that describes the tasks they should at least do and locations they should at least visit. The scenario consists of visits to several everyday places (bus, restaurant, shop, and library) and of several physical activities (lying, sitting, standing, walking, Nordic walking, running, rowing, cycling). Nordic walking is fitness walking with specifically designed poles to engage the upper body.

Because the signals have large interindividual difference in different activities, we recruited 16 volunteers (13 males, 3 females, age 25.8 ± 4.3 years, body mass index [BMI] 24.1 ± 3.0 kg/m^2) to gather a representative dataset for algorithm development. The volunteers were recruited by using bulletin board and news advertisements at a local university. The duration of each measurement session was about 2 h. The durations varied between measurement sessions because of the loose scenario, which was not supposed to be followed strictly. Because the goal was to collect realistic data, the test persons were given a lot of freedom during the measurement session. For example, they could choose the restaurant and shop they preferred. Also the order of places visited and time spent in each place depended on the test person.

The test person was accompanied by an annotator (same person for all cases), who used a personal digital assistance (PDA) to mark changes in context for reference purposes. An annotation application (Fig. 1) was written for a PDA using C#.NET. The annotation application provides a UI for visualizing and changing the currently selected and active contexts. In the UI, the contexts are organized hierarchically into upper level context types, e.g., *activity* and lower level context values, e.g., *lie* or *sit*. The context values are mutually exclusive. As a context value changes, the annotator taps on the name of the new context value with the PDA pen. The software stores the new state together with a timestamp on PDA

TABLE I
SIGNALS AND SENSORS OF DATA COLLECTION SYSTEM

Signal	Sensor	Measurement site	Fs
Altitude	Air Pressure (Suunto X6HR)	Wrist	0.5
Audio	Microphone (AKG C417)	Chest, on rucksack strap	22000, mono, 16 bit
Body Position	Metal ball moves between resistors (ProTech Position)	Chest	200
Chest Accelerations	3D acceleration (2 x Analog Devices ADXL202)	Chest, on rucksack strap	200
Chest Compass	3D compass (Honeywell HMC-1023)	Chest, on rucksack strap	200
EKG	Voltage between EKG electrodes (Blue Sensor VL, Embla A10)	Below left armpit, on breastbone	200
Environmental Humidity	Humidity (Honeywell, HIH-3605-B)	Chest, on rucksack strap	200
Environmental Light Intensity	Light sensor with two output dynamics (Siemens SFH 203P)	Chest, on rucksack strap	200
Environmental Temperature	Temperature sensor (Analog Devices TMP36)	Chest, on rucksack strap	200
Event Button	Switch (Embla XN Oximeter)	Chest, on rucksack strap	-
Heart Rate	IR light absorption (Embla XN oximeter)	Finger	1
Heart Rate	IR light reflectance (Nonin XPOD)	Forehead	3
Heart Rate	Voltage between chest belt electrodes (Suunto X6HR)	Chest	0.5
Location	GPS satellite receiver (Garmin eTrex Venture)	Shoulder, on rucksack strap	Based on location
Pulse Plethysmogram	IR light reflectance (Nonin XPOD)	Forehead	75
Respiratory Effort	Piezo sensor (Pro-Tech Respiratory Effort)	Chest	200
SaO2	IR light absorption (Embla XN Oximeter)	Finger	1
SaO2	IR light reflectance (Nonin XPOD)	Forehead	3
Skin Resistance	Resistance between two metal leads (Custom-made)	Chest	200
Skin Temperature	Resistive temperature sensor (YSI 409B)	Upper back, below neck	200
Wrist Accelerations	3D acceleration (Analog Devices, ADXL 202E)	Wrist, dominant hand	40
Wrist Compass	2D compass (Honeywell HMC-1022)	Wrist, dominant hand	40

TABLE II
SCENARIO FOR DATA COLLECTION

Location	Task
Home	Sitting at home
	Lying
	Sitting & reading newspaper
	Putting clothes on, going out
Bus	Walking to a bus stop
	Waiting for bus
	Traveling in bus
Restaurant	Walking to restaurant
	Queuing
	Eating, drinking, talking
Library	Walking to library
	Sitting in library, reading
Shop	Walking to shop
	Walking in shop, shopping
Home	Walking back home
Outdoor activities	Nordic Walking
	Running
Indoor activities	Rowing (rowing machine)
	Walking
	Bicycling (exercise bike)
	Sitting, drinking

Fig. 1. Annotation Software on PDA. Checkboxes on the left are used to expand and collapse between the title line and full view. Radio buttons are used to mark the active context value. Eating and Drinking can be active simultaneously. The asterisk is used to mark the context value "other."

memory. Data collection start and end markers were manually added to annotation data and all context data to allow synchronization of the data. The accuracy of manual markers is $\pm$ 0.5 s. In 2-h data collection this was considered an adequate accuracy.

B. Context Data Library

After the measurements, the data were synchronized, calibrated, re-sampled, and converted into suitable formats [27] for visualization. All the data (31 h) were collected into *Palantir Context Data Library 2003*.

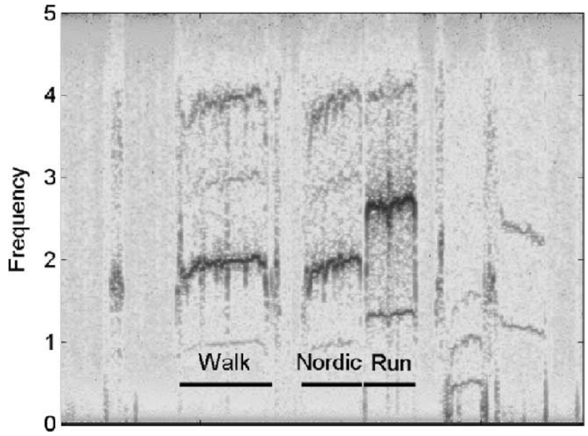


Fig. 2. Spectogram of vertical acceleration on chest during walking, Nordic walking and running. Horizontal axis is time.

C. Signal Processing and Feature Extraction

The goal in context recognition is to develop algorithms that can automatically infer the annotated context from the collected signals. The signals were first visually inspected and compared against the annotated contexts. This gave us the first impression on which signals are more useful than others. Feature signals (1-Hz sampling rate) were calculated from the raw data.

A priori information was used to select which features to calculate. For example, walking and running (measured in realistic circumstances) have constant frequency, which did not vary much between test persons either. Walking is seen as 2 Hz and running as 2.5–3 Hz oscillation in the signal (Fig. 2).

Time-domain features calculated were mean, variance, median, skew, kurtosis, 25% percentile and 75% percentile counted using a sliding window. *Frequency-domain features* were spectral centroid, spectral spread, estimation of frequency peak, estimation of power of the frequency peak, and signal power in different frequency bands. For acceleration both 4-s and 10-s windows were used. For blood oxygen saturation (SaO_2) the window was 10 s, for respiratory effort it was 60 s, and for all others it was 1 s.

Time-domain features were calculated for 1) body position; 2) humidity; 3) blood oxygen saturation SaO_2 ; 4) skin resistance; 5) skin temperature; and 6) environmental temperature. Both time and frequency domain features were calculated for 1) accelerations; 2) magnetometer signals; 3) environmental light intensity; and 4) respiratory effort.

In addition to the basic time- and frequency-domain features, the following features were calculated. Speech was detected from *audio* signal using a modified version of a speech/music discriminator [28]. Radius and two angles describing the vector of magnetic field as well as ratio between frequency bands 1–1.5 Hz and 0–5 Hz were calculated from *magnetometer signals*. R-peaks were detected and different features related to heart rate variability (e.g., R-R interval) were calculated from the electrocardiogram [29]. Speed was calculated from *GPS location data*. Power on frequency band 80–100 Hz was calculated from a *light-intensity signal*. Respiratory frequency, tidal volume, frequency and amplitude deviations, rate of ventilation and spectral entropies were estimated and calculated from the *respiratory effort signal*.

D. Feature Selection

Feature selection was based on visual and statistical analysis. The features were visually compared against annotation to find good candidate features. Distribution bar graphs of each feature signal during different activities were plotted for comparison (Fig. 4). The plots show how the distribution of each feature signal changes between different activities. The more the distribution moves between activities and the less the distributions overlap, the better it is for discrimination of activities.

A priori information was used in the quest for the best features. For example, during running there is more up-down movement and thus more energy in acceleration signal than during other activities. Based on *a priori* information, some new features were calculated from raw data. The best features were selected based on the distribution bar graphs. If there were more than one feature that could have been used for a specific decision, the feature with best discrimination power was selected.

As a result of the feature selection process, six features (Fig. 4) were selected for classification: 1) peak frequency of up-down chest acceleration F_{peak} *(chestacc,y)*; 2) median of up-down chest acceleration *Med(chestacc,y)*; 3) peak power of up-down chest acceleration P_{peak} *(chestacc,y)*; 4) variance of back-forth chest acceleration *Var(chestacc,z)*; 5) sum of variances of three-dimensional (3-D) wrist accelerations $\sum$ *Var(wristacc, 3 D)*; 6) power ratio of frequency bands 1–1.5 Hz and 0.2–5 Hz measured from left-right magnetometer on chest P_1 *(chestmagn,x)*.

E. Classification

During the feature selection process it was noticed that with the selected sensor setup, it was not possible to discriminate sitting and standing from each other (see Discussion for more details). Thus sitting and standing were combined into one class, resulting in seven target classes for classification: 1) lying; 2) sitting/standing; 3) walking; 4) Nordic walking; 5) running; 6) rowing (with a rowing machine); and 7) cycling (with an exercise bike). Three different classifiers were used in classification. All of them were given the same set of six features as inputs.

For classification, two decision trees were applied, namely a custom decision tree and an automatically generated decision tree. Also, an artificial neural network (ANN) was used as a reference classifier. Decision trees have been successfully applied to activity recognition earlier [16]. The custom decision tree was selected to represent a simple and transparent approach based on human rationalization. The automatically generated decision tree was selected to see how well the automatic tree generation algorithm performs compared with human-made rules. An advantage of the decision trees is that the problem of context recognition is divided in to smaller subproblems, which are tackled one by one very intuitively.

The recorded data were used for context recognition on a second-by-second basis by using the feature signals as inputs

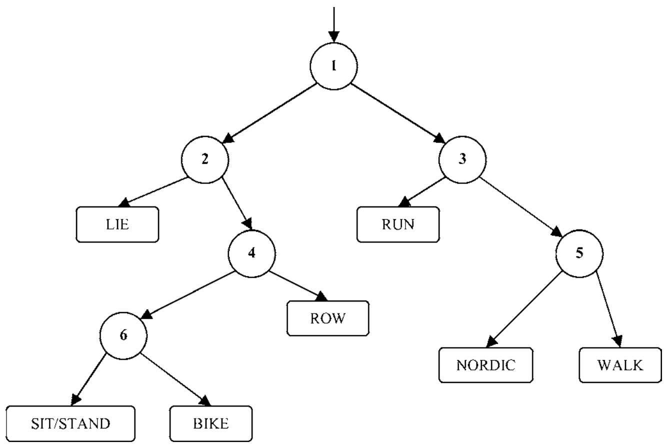


Fig. 3. Custom decision tree.

and PDA annotations as targets. For all three classifiers the results were acquired by 12-fold leave-one-subject-out cross-validation.

1) Custom Decision Tree: The custom decision tree (Fig. 3) was built by using domain knowledge and visual inspection of the signals. The tree has 13 nodes, 7 of which are leaf nodes and 6 of which represent a binary decision. The decisions can be named with questions: 1) footsteps?; 2) lying?; 3) running?; 4) rowing?; 5) Nordic walking?; 6) cycling?. The numbering refers to numbers of the six selected features. Leaf-nodes "sitting/standing" and "walking" can be considered as classes "other," because everything that is not recognized as any of the activities in upper levels of the tree falls into these categories. This is in line with the data, because the context value "other" was not used in annotations either. Fig. 4 depicts the decisions made in the nodes: it shows the distributions of feature data during each activity. The circled activities are relevant for the node; others have been ruled out in the upper level decisions. For each branch of the tree, the threshold value was defined by using a 12-fold leave-one-subject-out cross-validation. The threshold value for each node was chosen to be the average of the acquired 12 thresholds. The threshold values remained unchanged during the whole validation process.

2) Automatically Generated Decision Tree: An automatically generated decision tree was generated using a Matlab (MathWorks Inc, Natick, MA) Statistics Toolbox function called "treefit." The rule for splitting was Gini's index [30], which is one of the standard options. It progressively looks for the largest class in the data set and tries to isolate it from the rest of the data. The results were obtained by using leave-one-subject-out cross-validation resulting in separate training/validation sessions for each subject. In each training/validation session the tree was built using the training data (containing data from all but one subject), pruned to an optimum level (the level with the lowest error rate in the training set) using cross-validation within the training data, and the obtained tree was used to classify the data of the left-out subject. It should be noted that the size of the tree may be different in each training/validation session. In average the tree had 9.7 branches (minimum, 7; maximum, 14) and 10.7 leafs (minimum, 8; maximum, 15).

3) Artificial Neural Network: A multilayer perceptron with resilient backpropagation as the training algorithm was used as the artificial neural network classifier. The sizes of input, hidden and output layers were 6, 15 and 7, respectively. The output that had the highest value was selected as the classification result.

F. Postprocessing

Classification was made for each second of the data independently, and no temporal connections were considered. This resulted in rapid changes of the classification results especially at transitions between two activities. For instance, getting up from a sitting or lying position produced high acceleration peaks that caused misclassification. Activities that only last for a few seconds are not realistic. Thus, median filtering was used on the results of all three classifiers to use simple temporal logic to filter out short-duration misclassifications. The median filter replaces short activities with the surrounding longer duration activity. After several experiments, a median filter of 31 s was selected. A median filter this long may prevent the recognition of some short periods of activities (such as short walks) but it improves the overall classification. Both causal and anticausal versions were tested, and with the selected filter length their results were very close to each other. Anticausal filtering worked slightly better. Fig. 5 demonstrates the difference between filtered and unfiltered results.

III. Results

A. Data Quality

Data from 12 of 16 cases were used in classification. Data of four cases were left out because of missing wrist acceleration signals. The wrist acceleration signals were lost because of a hardware problem.

B. Classification Results

Tables III–V show the confusion matrices for the three different classifiers. In the tables, each sample represents 1 s. Table VI summarizes the classification accuracies of different activities.

IV. Discussion

We classified activities from realistic, out-of-lab context data using three different classifiers and six feature signals as inputs. Classification was done with 1-s time resolution; thus each second of the data was classified and compared with annotated data. Only a few previous studies have recognized activities from data measured in the out-of-lab environment. Rowing and Nordic walking have not been recognized in previous studies. Lying, sitting, standing, walking, running, and cycling have also been recognized in previous studies.

Bao and Intille [16] achieved recognition accuracy of 94.96% for lying down and relaxing, 94.78% for sitting and relaxing, 95.67% for standing still, 89.71% for walking, 87.,68% for running, and 96.29% for bicycling. Their data were measured in

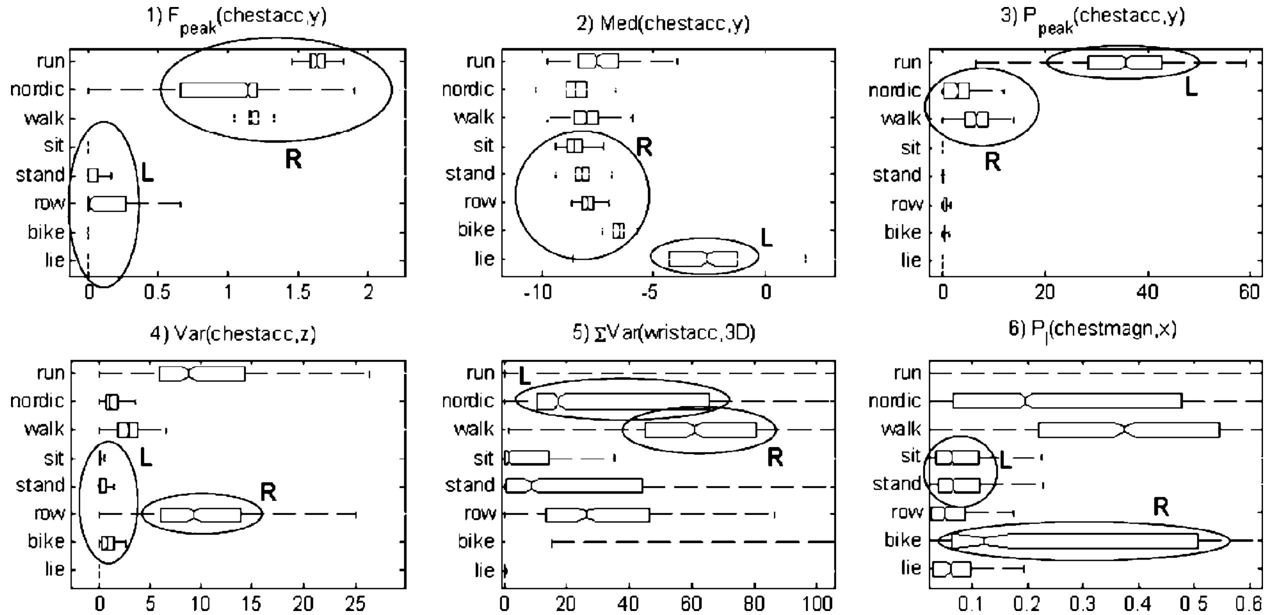


Fig. 4. Nodes of custom decision tree. Figure depicts distributions of feature signals during different activities. Activities marked with an R fall into right branch and activities marked an L fall into left branch of node. Circled activities are relevant for the node, others have been ruled out in upper level nodes.

Fig. 5. Classification results before (top) and after (bottom) median filtering. During black time intervals on the timeline sitting is classified as active $(\mathrm{sitting} = \mathrm{true})$. During white time intervals, sitting is not classified as active $(\mathrm{sitting} = \mathrm{false})$. Most of the sitting intervals that are shorter than 15 seconds are replaced with the dominant activity by median filtering.

TABLE III
CONFUSION MATRIX OF CUSTOM DECISION TREE

Annotation	Recognized Activity Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk
Lie	1417	0	0	205	0	0	0
Row	0	1646	0	717	0	0	23
ExBike	0	0	2461	612	0	0	29
Sit/ Stand	121	40	53	34083	4	340	962
Run	0	0	0	44	2284	21	5
Nordic walk	0	1	0	256	39	4507	194
Walk	0	16	4	5412	15	3964	12797

TABLE IV
CONFUSION MATRIX OF AUTOMATICALLY GENERATED DECISION TREE

Annotation	Recognized Activity Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk
Lie	1354	0	0	268	0	0	0
Row	0	1327	0	1028	0	0	31
ExBike	0	0	2552	508	0	0	42
Sit/ Stand	86	109	66	33928	8	13	1385
Run	0	0	0	45	2293	0	16
Nordic walk	0	0	0	250	36	3581	1130
Walk	0	27	4	4249	24	642	17262

TABLE V
CONFUSION MATRIX OF ARTIFICIAL NEURAL NETWORK

Annotation	Recognized Activity Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk
Lie	1206	0	0	357	0	0	59
Row	0	1414	0	874	63	0	35
ExBike	0	0	2336	561	0	0	205
Sit/ Stand	41	131	27	34032	0	22	1345
Run	0	250	0	40	517	1070	477
Nordic walk	0	2	0	210	0	2597	2188
Walk	0	18	4	4620	0	109	17457

a naturalistic environment, which is comparable to our setting. They used five acceleration sensors on the hip, wrist, arm, ankle, and thigh. They concluded that the thigh and wrist could be the ideal locations for activity recognition.

Absence of an accelerometer on the lower body is a limitation in our study. An extra accelerometer on the lower body would probably improve classification accuracy. Placing an accelerometer on the thigh was also considered in our study, but the thigh was not seen as a feasible sensor placement for a consumer product and this placement was ignored.

Foerster *et al.* [15] achieved recognition accuracy (subactivities combined) of 89% for lying, 100% for sitting, 88% for standing, 99% for walking, and 100% for cycling. Their data were collected in an out-of-lab environment. They segmented the data manually into 20 s or longer segments according to the behavior observation. Results were obtained by classifying the selected segments only (466 segments). About the segmentation they mention: "The classification can be improved by lengthening segments..." They used four sensor placements (chest, wrist, thigh, and lower leg). In our study, 1-s segments were used (72 272 segments).

TABLE VI
CLASSIFIER RESULTS [%]

	Custom Decision Tree	Automatic Decision Tree	Artificial Neural Network
Lie	**87**	83	74
Row	**69**	56	59
ExBike	79	**82**	75
Sit/ Stand	**96**	95	**96**
Run	**97**	**97**	22
Nordic walk	**90**	72	52
Walk	58	78	**79**
TOTAL	82	**86**	82

A. Confusions

Much of our classifiers' confusion seen in the results can be explained with transitions from one activity to another. The annotator was not given the choice to annotate "transition," but he had to switch from one activity to another instantly at some point during the transition. The transition is sometimes gradual, for example, when sitting changes to lying. The resulting inaccuracy is especially visible in the recognition of lying, which should be detected almost perfectly from the direction of gravity. Because lying periods were short, the uncertainty caused by transition periods in the beginning and end of lying became significant.

Lying is detected by the custom-made decision tree as a combination of decisions "no footsteps" and "lying." Duration of each lying period was only 2 min per case (total, 27 min) and confusion is 13 s per case (total, 3.5 min). The inaccuracy in annotation and duration of transition from sitting/standing to lying was in practice in this order. The artificial neural network additionally confuses lying with walking.

Recognition of *running* combines decisions "footsteps" and "running." Recognition of footsteps is rather clear (Fig. 4, node 1). The distributions of activities including footsteps and not including footsteps do not overlap much. The total amount of annotated running is about 39 min. The custom decision tree and the automatic decision tree recognize running very well. About 1 min of running is confused with standing and a few seconds with walking. Again, at least on part, the classifiers can be more accurate than the annotation and part of the confusion is not really confusion at all. Running started from the standing position and because of cars, slippery weather, etc. some walking and stops are included in the period annotated as "running." Artificial neural network confuses running heavily with other activities, especially with Nordic walking and walking.

Rowing is recognized as combination of decisions "no footsteps," "no lying." and "rowing." The custom decision tree recognizes 27 min of the total 40 min annotated as rowing. Because this includes data from 12 cases and rowing was started and ended by sitting, some sitting may indeed have been annotated as rowing. However, the amount of confusion toward sitting is rather large, so some classification error is also present. In addition, confusion with walking cannot be explained with annotation inaccuracy. The automatic decision tree similarly confuses rowing with sitting/standing and with walking. The artificial neural network commits the same error and further confuses 1 min of rowing as running.

Walking is one of the most common activities in this data set as in everyday life. Walking is recognized as combination of decisions "footsteps," "no running," and "no Nordic walking." Distributions of walking and Nordic walking partly overlap when using the feature in node 5. Both decision trees confuse walking mostly with Nordic walking and with sitting/standing. The artificial neural network confuses walking mostly with sitting/standing. Confusion with sitting/standing can be explained with inaccuracies in annotation. Activity annotated as walking often includes short periods of standing. Very short periods of walking between other dominant activities, even if annotated correctly and classified correctly by the decision tree, are replaced with dominating activity by the postprocessing method in the classification results. This degrades the performance when lots of short periods of walking are present.

Nordic walking was detected from increased arm motion. This approach is successful when the poles are used as effectively as they should be used. People not familiar with Nordic walking tend to use the poles very little and smoothly. Such use of the poles creates problems for recognition because the accelerations measured from the wrist have very low amplitude. This fact can be utilized, e.g., in teaching effective Nordic walking.

Sitting/standing is the most dominant activity in this data library and for most people in their everyday lives. It is recognized by combining decisions "no footsteps," "no lying," "no rowing," and "no cycling." All of the three classifiers classify sitting/standing rather well, mostly confusing them with walking. This is partly due to annotation inaccuracy. For example, in a library the activity annotated as standing includes very short periods of walking, which has not always been annotated correctly. Also, if annotated and classified correctly, very short periods of standing are replaced with the dominating activity by postprocessing.

Cycling with an exercise bike is detected from the left-right movement of chest by using the magnetometer signal. The distribution of cycling in this feature overlaps slightly with sitting/standing and thus some confusion with sitting/standing is inevitable. A small amount of annotation inaccuracy can be present mostly in the beginning and end of activities annotated as exercise biking. These are because the test person does not start or stop cycling exactly at the same time with annotation.

B. Classifiers

The custom decision tree treats the different activities more equally than the other classifiers because it optimizes performance of one node at a time, not the overall performance as the other classifiers do. That is why it has the best recognition accuracy for more than half of the activities, but the overall accuracy is not the best. The automatically generated tree had the best overall performance. This is in line with earlier studies [16]. For artificial neural network classification, the everyday data are rather noisy. Thus the artificial neural network easily overfits. Noticeable in neural network results is the poor recognition of running, which was well recognized by both of the decision tree classifiers.

C. Physiological Signals

Physiological signals such as heart rate and respiration were expected to have a larger role in activity recognition. Although they have been previously used together with accelerometers in ambulatory monitoring [14], they did not provide very useful data for activity recognition in our setup, in part, because they react to activity changes with a delay. The physiological signals correlate with the intensity level of the activity, but they do not reflect the type of activity (e.g., cycling versus walking), nor the duration of activity very accurately. With physiological signals (e.g., heart rate), the interindividual difference is also large, which creates extra challenge for algorithm development.

D. Sensors

In this study, accelerometers proved to be the most information-rich and most accurate sensors for activity recognition. They react fast to activity changes and they reflect well the type of activity. Placement of accelerometers in this study on rucksack straps and on wrists did not make it possible to separate sitting and standing from each other, because there was no clear change in the signal properties between these two activities. Different approaches were tried for detection of these activities. For example, it was assumed that the direction of a test person's body would stay more stable during sitting than during standing. However, the recorded data did not show such behavior. In the future, we will consider placing one accelerometer on the waist to enable discrimination of sitting from standing.

Even though gravity and magnetic flux are fundamentally different measures (e.g., direction), our data showed that for activity recognition, the information content of accelerometer and magnetometer signals is similar. Our 3D magnetometer and 3D accelerometer were located in one box, attached on a rucksack strap. When visually comparing the signals recorded during different physical activities, the magnetometer signal looks like a low-pass-filtered version of the accelerometer signal.

E. Temporal Connections Between Activities

Temporal connections between activities were not thoroughly studied in this work. In this study a median filter was used to remove very short activities from classifier results. Use of median filtering degrades the classification accuracy of the short-duration activities, which may be a problem in some applications. However, when aiming for a daily summary of activities, this is not a major problem. Utilizing the temporal history of activities might improve accuracy of activity recognition. Probabilistic models can be used to help in classification process, especially in transition from one recognized activity type to another. In [20] Markov chains have been used to assign probabilities to state transfers from one activity to another. The model is used to inhibit class change based on raw data only. If the transition has low probability, more requests from raw data classification are required before the change is accepted by the overall classification system. The drawback of this approach is that it requires a lot of realistic training data and probably also user-specific training data. However, in the future we will consider using a probabilistic model to reduce the number of (short-duration) false recognitions.

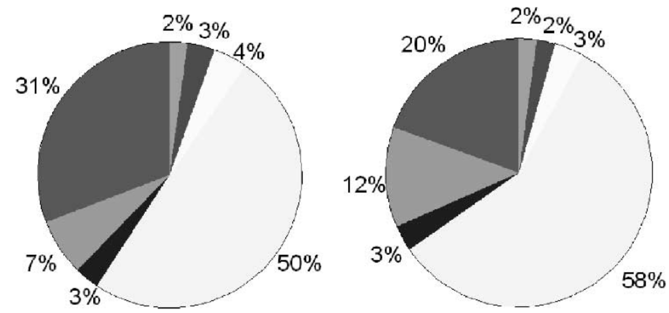


Fig. 6. Portions of activities in annotation (left) and results of custom decision tree (right). Activities clockwise from 12 o'clock: lying, rowing, cycling, sitting/standing, running, Nordic walking, walking.

F. Rucksack

Weight of the rucksack with the equipment was approximately 5 kg. This felt like a normal rucksack. Before selecting the rucksack, we also tried a belt bag, but compared with the rucksack, it felt uncomfortable with the equipment. In the data collection, the rucksack may have some effect on the activities, but it was not considered disturbing by the volunteers. Note that in this study placement of the chest acceleration sensor on the rucksack strap may affect the signal during dynamic activities, like running, because the rucksack moves slightly. However, in overall activity classification, the effect caused by rucksack movement is not significant.

G. Application: Activity Diary

Automatic classification of everyday activities can be used for promotion of a healthier lifestyle, e.g., with an "activity diary." The user could, e.g., in the evening check what kind of activities he has done during the day and how much time he has spent on each activity. Fig. 6 depicts the portions of our data in the form of an "activity diary."

V. Conclusion

Results of activity recognition were encouraging. With careful selection and placement of sensors, several everyday activities can be automatically recognized with good accuracy by using feature extraction and classification algorithms. Information about the daily activities can be used in consumer products to show the user his daily activity diary. This would increase the user's awareness of his daily activity level and promote a more active lifestyle.

Acknowledgment

The authors would like to thank A. Ylisaukko-oja, L. Cluitmans, J. Vilmi, P. Välkkynen, M. van Gils, and S.-M. Mäkelä for their work and help during the project. The authors would also like to thank Finnish Institute of Occupational Health for letting us use some of their equipment. Special thanks also to all volunteers for collecting a valuable data library with us.

References

[1] P. Puska, H. Benaziza, and D. Porter (2004, Oct. 15). Physical Activity, WHO Information Sheet on Physical Activity, WHO [Online]. Available: http://www.who.int/ dietphysicalactivity/media/en/gsfs_pa.pdf

[2] H. Hagendoorn, I. Vuori, and P. Oja (2004, Oct. 15). Guidelines for the development of national policies and strategies for promoting health through physical activity, The European Network for the Promotion of Health-Enhancing Physical Activity [Online]. Available: http://www.who.int/gb/ebwha/pdf_files/WHA57/A57_R17-en.pdf

[3] R. R. Pate, M. Pratt, S. N. Blair, W. L. Haskell, C. A. Macera, C. Bouchard, D. Buchner, W. Ettinger, G. W. Health, A. C. King, A. Kriska, A. S. Leon, B. H. Marcus, J. Morris, R. S. Paffenbarger, K. Patrick, M. L. Pollock, J. M. Rippe, J. Sallis, and J. H. Wilmore, "Physical activity and public health: a recommendation from the centers for Disease Control and Prevention and the American College of Sports Medicine," *JAMA*, vol. 273, no. 5, pp. 402–407, Feb. 1995.

[4] N. C. Barengo, G. Hu, T. A. Lakka, H. Pekkarinen, A. Nissinen, and J. Tuomilehto, "Low physical activity as a predictor for total and cardiovascular disease mortality in middle-aged men and women in Finland," *Eur. Heart J.*, vol. 25, pp. 2204–2211, 2004.

[5] J. E. Manson, P. Greenland, A. Z. LaCroix, M. L. Stefanick, C. P. Mouton, A. Oberman, M. G. Perri, D. S. Sheps, M. B. Pettinger, and D. S. Siscovick, "Walking compared with vigorous exercise for the prevention of cardiovascular events in women," *N. Engl. J. Med.*, vol. 347, no. 10, pp. 716–725, Sep. 2002.

[6] M. Martinez-Gonzalez *et al.*, "Physical inactivity, sendentary lifestyle and obesity in the European Union," *Int. J. Obes. Relat. Metab. Disord.*, vol. 23, no. 11, pp. 1192–1201, Nov. 1999.

[7] M. Fogelholm, P. Oja, M. Rinne, J. Suni, and I. Vuori, "Riittääkö puoli tuntia kävelyä päivässä" *Suomen Lääkärilehti*, vol. 59, no. 19, pp. 2040–2042, May 2004.

[8] C. Bouten, A. Sauren, M. Verduin, and J. Janssen, "Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking," *Med. Biol. Eng. Comput.*, vol. 35, no. 1, pp. 50–56, Jan. 1997.

[9] M. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiol. Measure.*, vol. 25, no. 2, pp. R1–R20, Apr. 2004.

[10] K. Aminian and B. Najafi, "Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications," *Comput. Animation Virtual Worlds*, vol. 15, no. 2, pp. 79–94, May 2004.

[11] M. Mathie, A. Foster, N. Lovell, and B. Celler, "Detection of daily physical activities using a triaxial accelerometers," *Med. Biol. Eng. Comput.*, vol. 41, no. 3, pp. 296–301, May 2003.

[12] K. Aminian, Ph. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon, "Physical activity monitoring based on accelerometry: validation and comparison with video observation," *Med. Biol. Eng. Comput.*, vol. 37, no. 3, pp. 304–308, May 1999.

[13] F. Foerster and J. Fahrenberg, "Motion pattern and posture: Correctly assessed by calibrated accelerometers," *Behav. Res. Methods Instrum. Comput.*, vol. 32, no. 3, pp. 450–457, Aug. 2000.

[14] J. Ng, A. Sahakian, and S. Swiryn, "Accelerometer-based body-position sensing for ambulatory electrocardiographic monitoring," *Biomed. Instrum. Technol.*, vol. 37, no. 5, pp. 338–346, Sep. 2003.

[15] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Comput. Hum. Behav.*, vol. 15, no. 5, pp. 571–583, 1999.

[16] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. 2nd Int. Conf. Pervasive Computing*, 2004, pp. 1–17.

[17] V.-M. Mäntylä, J. Mäntyjärvi, T. Seppänen, and E. Tuulari, "Hand gesture recognition of a mobile device user," in *Proc. Int. IEEE Conf. Multimedia Expo (ICME)*, 2000, pp. 281–284.

[18] J. Mäntyjärvi, J. Himberg, and T. Seppänen, "Recognizing human motion with multiple acceleration sensors," in *Proc. Int. IEEE Conf. Systems, Man Cybernetics (SMC)*, 2001, pp. 747–752.

[19] S.-W. Lee and K. Mase, "Activity and location recognition using wearable sensors," *Pervasive Computing*, vol. 1, no. 3, pp. 24–32, Jul.–Sep. 2002.

[20] K. van Laerhoven and O. Cakmakci, "What shall we teach our pants?," in *Proc. 4th Int. Symp. Wearable Comput.*, 2000, pp. 77–83.

[21] P. Korpipää, M. Koskinen, J. Peltola, S.-M. Mäkelä, and T. Seppänen, "Bayesian approach to sensor-based context awareness," in *Personal Ubiquitous Comput. J.*, vol. 7, no. 4, pp. 113–124, Jul. 2003.

[22] A. Flanagan, J. Mäntyjärvi, and J. Himberg, "Unsupervised clustering of symbol strings and context recognition," in *Proc IEEE Int. Conf. Data Mining (ICDM'02)*, 2002, p. 171.

[23] A. Dey, G. D. Abowd, and D. Salber, "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications," *Human-Computer Interaction*, vol. 16, no. 2–4, pp. 97–166, 2001.

[24] N. Streitz and P. Nixon, "The disappearing computer," *Commun. ACM*, vol. 48, no. 3, pp. 32–35, Mar. 2005.

[25] J. Coutaz, J. L. Crowley, S. Dobson, and D. Garlan, "Context is key," *Commun. ACM*, vol. 48, no. 3, pp. 49–53, Mar. 2005.

[26] J. Pärkkä, M. van Gils, T. Tuomisto, I. Korhonen, and R. Lappalainen, "A wireless wellness monitor for personal weight management," in *Proc. IEEE ITAB-ITIS2000*, Arlington, VA, 2000, pp. 83–88.

[27] J. Pärkkä, L. Cluitmans, and P. Korpipää, "Palantir Context Data Collection, Annotation and PSV File Format," in *Pervasive 2004 Workshop Proc.: Benchmarks and Database Context Recognition*, 2004, pp. 9–16.

[28] J. Penttilä, J. Peltola, and T. Seppänen, "A speech/music discriminator-based audio browser with a degree of certanity measure," in *Proc. Infotech Oulu Int. Workshop Information Retrieval (IR-2001)*, 2001, pp. 125–131.

[29] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger, F. Lombardi, A. Malliani, A. J. Moss, J. N. Rottman, G. Schmidt, P. J. Schwartz, and D. H. Singer, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use," *Eur. Heart J.*, vol. 17, pp. 354–381, Mar. 1996.

[30] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, MAr. 1984.

Juha Pärkkä received the M.Sc. (Tech.) degree in information technology (digital signal processing) from Tampere University of Technology in Tampere, Finland, in 1997.

Currently, he is working as a research scientist at VTT Information Technology in Tampere, Finland. His main research interests are biomedical signal processing and ubiquitous computing.

Miikka Ermes is studying for the M.Sc. degree in Tampere University of Technology. He is majoring in signal-processing methods and he is currently finishing his M.Sc. thesis concerning activity recognition.

Currently, he is working as a Research Trainee at VTT Information Technology in Tampere, Finland.

Panu Korpipää (M'05) received his M.Sc. (Tech.) degree in electrical engineering (software engineering) from the University of Oulu, Finland, in 1996.

Currently, he is working as a senior research scientist at VTT Electronics, Advanced Interactive Systems, Oulu, Finland. His professional interests include mobile and context aware computing, novel user interfaces, and visualization.

Jani Mäntyjärvi received the M.Sc. degree in biophysics and the Ph.D. degree in information processing from the University of Oulu in 1999 and 2004, respectively.

Currently, he is a senior research scientist at the Technical Research Center of Finland, VTT Electronics. His current professional interests include pervasive and context-aware computing for handheld devices, wearable sensing, and technologies for adaptive interaction.

Johannes Peltola received the M.Sc. (Tech.) degree in electrical engineering from the University of Oulu, Finland, in 1998.

Currently, he is working as a research scientist in VTT Electronics, Oulu. His research topics are multimedia signal-processing algorithms for source compression, content analysis, and joint source channel coding. He is also leading a research team in the field of these topics.

Ilkka Korhonen (M'98) received the M.Sc. and Dr.Tech. degrees in digital signal processing from Tampere University of Technology in 1991 and 1998, respectively.

Currently, he is currently working as a research professor for intuitive information technology at VTT Information Technology. He is a docent in medical informatics (with a speciality on biosignal processing) in Ragnar Granit Institute at Tampere University of Technology. His main research interests include biosignal interpretation methods and pervasive health care technologies, especially their application in critical care patient monitoring, wearable biomedical monitoring, and home health monitoring. He has published more than 60 original papers in international scientific journals and conference proceedings.

PUBLICATION **P2**

Detection of Daily Activities and Sports with Wearable Sensors in Controlled and Uncontrolled Conditions

In: IEEE Transactions on Information Technology in Biomedicine 2008. Vol. 12, No. 1, pp. 20–26.

Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions

Miikka Ermes, Juha Pärkkä, *Member, IEEE*, Jani Mäntyjärvi, and Ilkka Korhonen, *Member, IEEE*

***Abstract*—Physical activity has a positive impact on people's well-being, and it may also decrease the occurrence of chronic diseases. Activity recognition with wearable sensors can provide feedback to the user about his/her lifestyle regarding physical activity and sports, and thus, promote a more active lifestyle. So far, activity recognition has mostly been studied in supervised laboratory settings. The aim of this study was to examine how well the daily activities and sports performed by the subjects in unsupervised settings can be recognized compared to supervised settings. The activities were recognized by using a hybrid classifier combining a tree structure containing *a priori* knowledge and artificial neural networks, and also by using three reference classifiers. Activity data were collected for 68 h from 12 subjects, out of which the activity was supervised for 21 h and unsupervised for 47 h. Activities were recognized based on signal features from 3-D accelerometers on hip and wrist and GPS information. The activities included lying down, sitting and standing, walking, running, cycling with an exercise bike, rowing with a rowing machine, playing football, Nordic walking, and cycling with a regular bike. The total accuracy of the activity recognition using both supervised and unsupervised data was 89% that was only 1% unit lower than the accuracy of activity recognition using only supervised data. However, the accuracy decreased by 17% unit when only supervised data were used for training and only unsupervised data for validation, which emphasizes the need for out-of-laboratory data in the development of activity-recognition systems. The results support a vision of recognizing a wider spectrum, and more complex activities in real life settings.**

***Index Terms*—Activity classification, context awareness, physical activity, wearable sensors.**

I. Introduction

CHRONIC noncommunicable diseases (NCDs) cause 60% of global deaths and the figure is expected to rise to 73% by 2020 [1]. Such diseases include, for example, cardiovascular diseases, diabetes, osteoporosis, and certain types of cancer. Physical inactivity is a major risk factor for these deaths, and it is estimated to cause 2 million unnecessary deaths per year. There is, thus, an urgent need to promote more active lifestyle.

There is strong evidence that regular physical exercise decreases the risk of cardiovascular disease (e.g., [2]), which is the leading cause of death in many developed countries. Risk factors associated with cardiovascular diseases include smoking, obesity, and high blood pressure, the last two of which are closely related to physical inactivity. Type II diabetes is strongly associated with obesity that, in turn, has a well-known relation to physical inactivity [3]. There is evidence that exercise improves the physiological control of glucose metabolism [4]. Falls are also a major health hazard to elderly people resulting often in hip fracture requiring surgical operation and long rehabilitation. It is suggested that muscle strength, neuromuscular coordination, postural stability, steadiness of gait, and the structural properties of bone all influence fall frequency [5]. Each of these can be directly enhanced by physical training.

Estimating energy expenditure is a common way to assess the activity level of a subject. Traditional devices for the estimation of energy expenditure are not suited for unobtrusive ambulatory monitoring. Recently, wearable devices have become available for that purpose and studies have shown that accelerometer-based estimation of energy expenditure can be obtained with relatively good accuracy [6], [7]. However, energy expenditure is only one important aspect of physical activity. An international consensus statement regarding physical activity, fitness, and health [8] identifies six areas affected by physiological effort: body shape, bone strength, muscular strength, skeletal flexibility, motor fitness, and metabolic fitness. All of these have their own distinct impact on an individual's general well-being, and thus, estimating energy expenditure alone is not sufficient in order to assess the overall impact of the physical activities on the individual's well-being.

A more detailed analysis of physical effort can be obtained by activity recognition, i.e., by detecting the exact form of activity the subject is performing. Previous studies have applied activity recognition, e.g., for elderly care [9]. We believe that another important application domain for the activity recognition lies in preventive healthcare (prevention of NCDs). In order to avoid the vicious circle of illnesses and related reduced ability to perform physical activities, the monitoring of the changes in physical activity needs to start before the physical ability of an individual starts to decline.

Accelerometers are currently among the most widely studied wearable sensors for activity recognition, thanks to their accuracy in the detection of human body movements, small size, and reasonable power consumption [10]. Recent reviews have described the use of accelerometers in movement and activity detection [10], [11]. In laboratory settings, the most prevalent everyday activities (sitting, standing, walking, and lying) have

Manuscript received October 27, 2006. This work was supported by the Tekes (National Technology Agency of Finland) Fenix programme by VTT, Tekes, Nokia, Clothing+, and Suunto.

M. Ermes, J. Pärkkä, and I. Korhonen are with the VTT Technical Research Centre of Finland, 33101 Tampere, Finland (e-mail: miikka.ermes@vtt.fi; juha.parkka@vtt.fi; ilkka.korhonen@vtt.fi).

J. Mäntyjärvi is with VTT Technical Research Centre of Finland, 90571 Oulu, Finland (e-mail: jani.mantyjarvi@vtt.fi).

Digital Object Identifier 10.1109/TITB.2007.899496

been successfully recognized with accelerometers [12]–[18]. However, the applicability of these results to out-of-laboratory monitoring is unclear. Long-term out-of-laboratory monitoring often means less-controlled user-annotated data collection, which introduces several challenges such as the following.

1) Annotations of the data are more unreliable causing difficulty in classifier training and also degrading classification results.
2) People perform activities in many different ways that are hard to categorize. For example, a person may lie down on a sofa in a way that cannot be said to be either sitting or lying.

In a few studies, data have been collected outside the laboratory. In [16], 24 subjects spent approximately 50 min outside the laboratory. Accelerometers were placed on sternum, wrist, thigh, and lower leg. Nine patterns (sitting, standing, lying, sitting and talking, sitting and operating PC, walking, stairs up, stairs down, and cycling) were recognized from presegmented data using similarity measures with a total accuracy of 66.7%. In the same study, the accuracy in laboratory settings was 95.8% illustrating the difficulties introduced by out-of-laboratory settings. In [19], five biaxial accelerometers attached to hip, wrist, arm, ankle, and thigh were used to recognize 20 everyday activities such as walking, watching TV, brushing teeth, vacuuming, etc. Data were collected for 82–160 min from 20 subjects. Four different classifier structures were used of which decision tree provided the best results accuracies ranging from 41% to 97% for different activities.

In our previous study on activity recognition [20], 16 test persons went through an approximately 2 h recording session with a supervisor during which the following activities were executed: lying, rowing (with a rowing machine), cycling (with an exercise bike), sitting, standing, running, Nordic walking, and walking. The recognition accuracies for different activities varied for the best classifier between 58% and 97%.

In this study, new data were collected that also contained unsupervised out-of-laboratory period. Our aim was to study the effects of such environment to the classification accuracy.

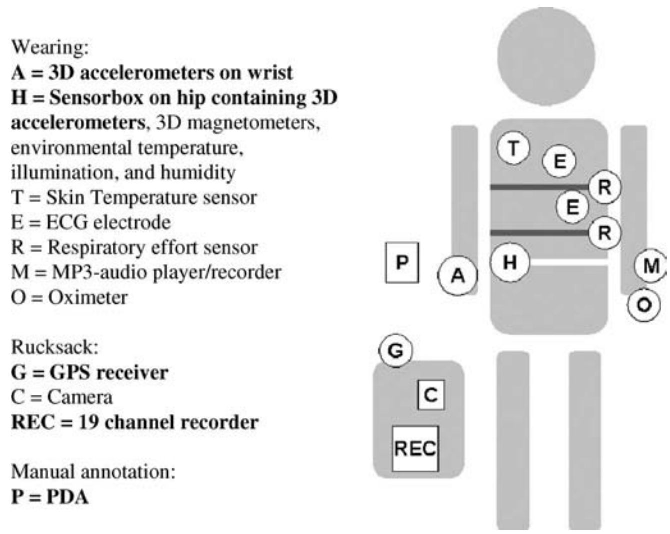


Fig. 1. Data collection and annotation system. The sensors and devices relevant for this study are printed in bold.

TABLE I
TEST PROTOCOL FOR DATA COLLECTION

Supervised indoor activities	**Time [min]**
Lying	3
Sitting and doing computer work	20
Lying	3
Sitting and reading comics	5
Lying	3
Cycling with an exercise bike	5
Lying	3
Exercising with a rowing machine	5
Lying	3
Standing and reading a poster	5
Supervised outdoor activities	**Time [min]**
Cycling with a regular bike	5
Walking to a park	5
Playing football	5
Walking back from park	5
Nordic walking to park & back	5
Running to park & back	5
Drinking water indoors	2
Unsupervised free period	**Time [min]**
Changing clothes	5
Free period (min. duration)	240

II. METHODS

A. Data Collection

The devices used in the data collection and their locations on the body are illustrated in Fig. 1. Although this figure includes sensors from which data are not used in this study, they are shown in the figure for more complete picture of the data-collection system.

Acceleration signals were measured with ADXL202 accelerometers (Analog Devices, Norwood, MA), and were stored on a flash-card-memory-based, 19-channel recorder (Embla A10, Medcare, Reykjavik, Iceland). Sampling frequency was 20 Hz, and the range of the sensor output was ± 10 g. Location information was stored on a Garmin eTrex Venture GPS receiver (Garmin Ltd., Olathe, KA) once per 20 s. Accelerometers were attached to their data-storage unit by cables. Cables were taped to the body so that they did not restrict normal movements. Also, the cables were placed so that it was possible to place the rucksack with the data-storage unit on the floor, for example, when sitting down.

Twelve subjects [aged 27.1 ± 9.2 years, body mass index (BMI) 23.8 ± 1.9, ten males, two females] were recruited by advertisements at a local university. A written consent was obtained from each volunteer. Approximately 6 h of data were collected with each subject. The 6 h measurement session was further divided into two phases: 1) a supervised period with exact scenario and accurate supervisor-made annotations and 2) an unsupervised period with subject-made annotations. The test protocol is described in Table I.

During the supervised data-collection session, the subject was accompanied by a supervisor, who used a personal digital

Fig. 2. Annotation application on PDA.

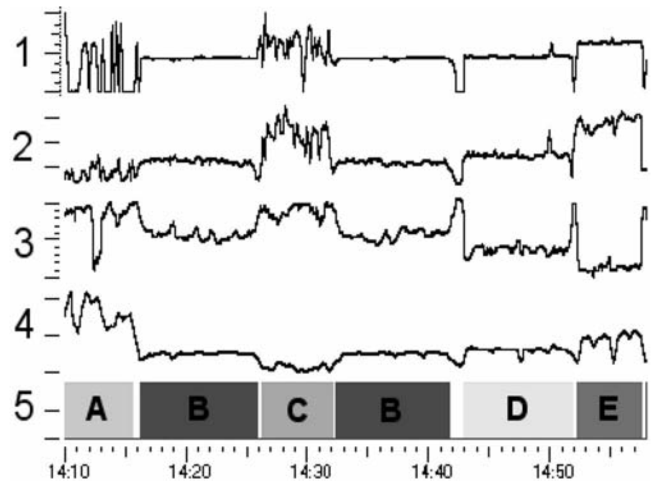


Fig. 3. Selected signal features during different activities in excerpt from the supervised data. Panels from top to bottom: 1) peak frequency of up-down acceleration (feature A); 2) range of up-down acceleration (feature B); 3) spectral entropy of up-down acceleration (feature F); 4) speed; 5) activity: A) cycling; B) walking; C) playing football; D) Nordic walking; E) running.

assistant (PDA) and a custom-made application to mark changes in activity and context for reference purposes (Fig. 2). After the supervised phase, the use of the PDA application was instructed to the test person, and he/she made the annotations himself/herself throughout the unsupervised period.

Exercise bike and *rowing machine* were each used for 5 min indoors. The user was given the freedom to choose a comfortable pace in each activity. *Cycling* was performed outdoors with real bikes. Most test persons used their own bikes. *Football* was played in a nearby park. In practice, this meant kicking a ball with the supervisor and running after the ball every now and then, not real football game with 22 players. *Nordic walking* is an activity that has recently become popular in northern and middle Europe. In short, it is fast-pace walking using poles similar to skiing poles. It also enables the training of the upper body during walking. During the unsupervised period, many people went to work or attend lectures. Some people performed different activities such as bowling, driving a car, walking to different places like library, cottage, etc. One person went home and took a nap.

Fig. 2 depicts the annotation application [21]. In each panel, the options are exclusive. The context value was changed by tapping another value. “Activity” panel was used to mark the true activity of the subject. “Location” was used to tell whether the subject was indoors, outdoors, or in a vehicle. “Eating” described eating and drinking in general. “Annotator” is “assistant” during the supervised activities and “self” during the unsupervised period. “Sync” was used to mark the start and end markers for synchronization. For some annotated context, there was also an option for transition (“$*$” in the application UI) such as the transition from sitting to standing, which was used only by the supervisor. It was not in use during the free period, as marking the transitions from one activity to another was considered too challenging to be done by the subject alone while performing the activities.

B. Signal Processing

Signal features were calculated for each second of the data collection. Time-domain features calculated were mean, variance, median, skew, kurtosis, 25% percentile, and 75% percentile. Frequency-domain features included the estimation of power of the frequency peak and signal power in different frequency bands. Speed was calculated from GPS location data. Spectral entropy S_N [22] of the acceleration signals for the frequency band 0–10 Hz was calculated as

$$S_N(f_1, f_2) = \frac{-\sum_{f_i=f_1}^{f_2} P(f_i)\log(P(f_i))}{\log(N[f_1, f_2])} \qquad (1)$$

where $P(f_i)$ represents the power spectral density (PSD) value of the frequency f_i. The PSD values are normalized so that their sum in the band $[f_1,\, f_2]$ is 1. $N[f_1,\, f_2]$ is the number of frequency components in the corresponding band in PSD.

The feature selection proceeded by identifying for each activity the feature having the best performance in discriminating the corresponding activity from other activities. The performance of each feature was evaluated by the area under the receiver operator characteristic (ROC) curve.

Figs. 3 and 4 show examples of how different signal features behave during different activities. The following signal features were selected for activity classification:

1) peak frequency of the up–down acceleration measured from the hip;
2) range of the up–down acceleration measured from the hip;
3) mean value of the up–down acceleration measured from the hip;
4) peak frequency of the horizontal acceleration measured from the wrist;
5) sum of variances of 3-D acceleration measured from the hip;
6) spectral entropy of the up–down acceleration measured from the hip;
7) speed measured from the GPS.

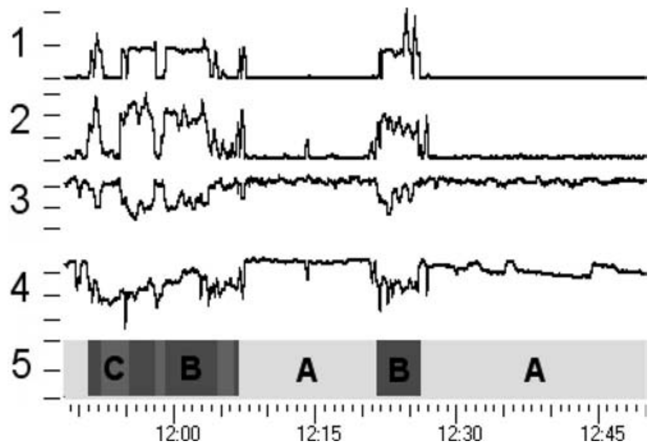


Fig. 4. Selected signal features during different activities in excerpt from the unsupervised data. Panels from top to bottom: 1) peak frequency of up-down acceleration (feature A); 2) range of up-down acceleration (feature B); 3) spectral entropy of up-down acceleration (feature F); 4) mean value of up-down acceleration (feature C); 5) activity: A) sitting; B) walking; C) standing. Also, shorter segments with corresponding shades of gray represent the same activities.

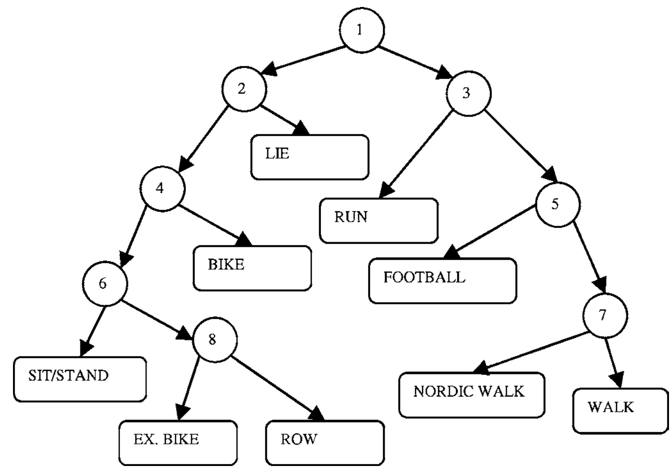


Fig. 5. Structure of the custom decision tree and hybrid model.

The following nine target classes were used for the activity recognition: 1) lying; 2) sitting and standing (see Section IV for the reason why these activities were combined into a single group); 3) walking; 4) running; 5) Nordic walking; 6) rowing with a rowing machine; 7) cycling with an exercise bike; 8) cycling with real bike; and 9) playing football. Compared to our earlier study, two activities were novel: cycling with a real bike and playing football. Although cycling with a regular bike and with an exercise bike are very similar activities, we wanted to keep them as separate classes because in everyday life they can be performed with different purposes: exercise bike is used only for exercising aerobic fitness, whereas regular bike is often used for transportation. Football was included to test the feasibility of the system to detect a more complex type of activity, as football comprises walking, standing, running, kicking the ball, etc.

Four different classifiers were used: 1) custom decision tree; 2) automatically generated decision tree; 3) artificial neural network (ANN); and 4) hybrid model. Classifiers 1–3 were included mainly for comparison purposes to evaluate the performance of the classifier 4. For all classifiers, results were acquired by 12-fold leave-one-subject-out cross validation. Each classifier had the same seven-signal features at their disposal. Data sample order was randomized before the training phase. The following describes the classifier structures in detail.

1) *Custom decision tree*: In custom decision tree, each decision is made by a simple thresholding mechanism [20]. The structure of the tree was built using *a priori* knowledge and our own intuitive modeling of different activities. The obtained tree had eight binary decision nodes. The structure of the tree is depicted in Fig. 5. Specific questions can be assigned to each of the numbered decision nodes: a) footsteps? b) lying? c) running? d) cycling? e) playing football? f) doing indoor exercise? g) Nordic walking? h) rowing? The tree has been built so that "walk" and "sit/stand" are default groups for any activity the decision tree is not familiar with. For example, if footsteps are detected, but not the characteristics of running or Nordic walking, the activity falls through the tree to a class "walk". Similarly, if no footsteps are recognized and also none of the characteristics of lying, cycling, cycling on exercise bike, or rowing, the activity falls to "sit/stand".
2) *Automatically generated decision tree*: An automatically generated decision tree was used in order to compare how well the human-made rules and tree structure perform compared to automatic classification. The tree was generated using a Matlab (MathWorks, Inc., Natick, MA) Statistics Toolbox function "treefit."
3) *Artificial neural network (ANN)*: A multilayer perceptron with a hidden layer of 15 nodes and with resilient back propagation as the training algorithm was used as the ANN classifier.
4) *Hybrid model*: As a novel method, we combined the best qualities of the custom decision tree model and neural networks. Our observations suggested that though implementing *a priori* knowledge into a classifier structure improved the results in general, it also resulted in simpler rules, which degraded the recognition accuracy in some aspects. Thus, the purpose was to achieve a model that could combine the best properties of the human *a priori* knowledge of the activities with the accurate nonlinear classification properties of the ANNs. In the hybrid model, the simple thresholding decisions made in each decision node of the custom decision tree (Fig. 5) were replaced by small multilayer perceptron networks (size 7:5:1). Each node gave as output a value between 0 and 1. A value of 0.5 was considered the decision boundary when selecting which branch of the tree to proceed.

III. RESULTS

The total amount of data used for the analysis was 68:28:32 (hh:mm:ss), 21:08:57 of which were supervised data and 47:19:35 were unsupervised. The data consisted of the following percentages of activities: 1) lying 7.3%; 2) rowing 1.5%; 3) cycling with an exercise bike 1.4%; 4) sitting and standing

TABLE II
SUMMARY OF THE ACTIVITY-RECOGNITION RESULTS USING BOTH SUPERVISED AND UNSUPERVISED DATA FOR TRAINING AND TESTING OF THE CLASSIFIERS (PERCENTAGES)

	Custom Decision Tree	Automatic Decision Tree	Artificial Neural Network	Hybrid Model
Lie	**98**	96	**98**	97
Row	58	84	85	**87**
ExBike	20	**79**	4	18
Sit/ Stand	94	53	96	**97**
Run	**91**	83	90	89
Nordic walk	**85**	66	66	70
Walk	50	62	67	**71**
Football	63	55	47	**78**
Bike	52	**74**	67	72
TOTAL	83	60	87	**89**

TABLE III
DETAILED ACTIVITY-RECOGNITION RESULTS OF THE HYBRID MODEL USING BOTH SUPERVISED AND UNSUPERVISED DATA FOR TRAINING AND TESTING OF THE CLASSIFIER (PERCENTAGES)

Annotation	Recognized Activity								
	Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk	Foot-ball	Bike
Lie	97	0	0	3	0	0	0	0	0
Row	0	87	0	13	0	0	0	0	0
ExBike	0	0	18	66	0	0	16	0	0
Sit/ Stand	0	0	0	97	0	0	2	0	0
Run	0	0	0	0	89	7	2	1	0
Nordic walk	0	0	0	1	5	70	24	0	0
Walk	1	0	0	21	0	5	71	1	1
Football	0	0	0	2	7	1	12	78	0
Bike	2	0	1	12	1	1	12	1	72

TABLE IV
DETAILED ACTIVITY-RECOGNITION RESULTS OF THE HYBRID MODEL USING ONLY SUPERVISED DATA FOR TRAINING AND TESTING OF THE CLASSIFIER (PERCENTAGES)

Annotation	Recognized Activity								
	Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk	Foot-ball	Bike
Lie	99	0	0	0	0	0	0	0	1
Row	0	97	3	0	0	0	0	0	0
ExBike	0	0	81	16	0	0	2	0	1
Sit/ Stand	0	1	3	95	0	0	0	0	0
Run	0	0	0	0	90	7	1	2	0
Nordic walk	0	0	0	1	6	78	13	1	0
Walk	0	0	1	0	3	12	81	2	1
Football	0	0	0	0	6	0	6	88	0
Bike	3	0	2	0	1	0	3	1	91

63.2%; 5) running 1.9%; 6) Nordic walking 2.8%; 7) walking 16.8%; 8) football 1.5%; and 9) cycling with a regular bike 3.6%. The classification results are summarized in Table II. The results of the hybrid model are described in Table III. To assess the importance and reliability of supervised and unsupervised data sets, the following results were also calculated.

1) The total classification accuracy was calculated using only supervised data both in training and testing of the model (leave-one-subject-out cross-validation). The test was performed in order to obtain activity-recognition results that are comparable to those of the earlier studies with laboratory data. The total classification accuracy was 90%, increasing by 1% unit compared to the result obtained by using all collected data in training and validation. The results are shown in Table IV.

TABLE V
DETAILED ACTIVITY-RECOGNITION RESULTS OF THE HYBRID MODEL USING SUPERVISED DATA FOR TRAINING AND UNSUPERVISED DATA FOR TESTING OF THE CLASSIFIER (PERCENTAGES)

Annotation	Recognized Activity								
	Lie	Row	Ex-Bike	Sit/ Stand	Run	Nordic walk	Walk	Foot-ball	Bike
Lie	98	0	1	1	0	0	0	0	0
Sit/ Stand	1	2	9	80	0	0	2	1	5
Walk	2	4	29	15	0	3	30	13	4
Bike	1	3	17	6	1	1	5	17	49

2) The supervised data were used for the training, whereas the unsupervised data were used for the testing of the model. The test was performed in order to assess the feasibility of a scenario in which an activity-recognition device would be trained with laboratory data and be used in out-of-laboratory settings. The total classification accuracy was 72% decreasing by 17% unit compared to the result obtained by using all collected data in training and validation. Only four activities were annotated by the subjects during the free period: lying down, sitting and standing, walking, and cycling. The results are shown in Table V.

IV. DISCUSSION

Activity data were collected for 68 h from 12 subjects, out of which the activity was supervised for 21 h and unsupervised for 47 h. Activities were recognized from the data by using 3-D accelerometers on hip and wrist and GPS information. The total accuracy of the activity recognition using both supervised and unsupervised data was 89%. In comparison to our previous study in which only supervised data were used and the total accuracy of 86% was achieved [20], the results obtained here show slightly improved performance.

The aim of the study was to assess the feasibility of activity recognition in out-of-laboratory settings. The 1% unit difference between the classification accuracy obtained using all data and that obtained using only supervised data suggest that activity recognition is also feasible in out-of-laboratory. However, the 17% unit decrease in the classification accuracy when only supervised data were used for training and only unsupervised for validation suggests that in order to obtain an activity-recognition algorithm feasible in out-of-laboratory settings, it must also be trained with annotated real-life data.

The hybrid model classifier proved to provide better results than the reference classifiers. It confirms our hypothesis that combining human *a priori* knowledge and the nonlinear classification process of neural networks may provide a basis for activity recognition with even greater variety of activities. However, with ANNs, an important issue is the noise and inaccuracy in the everyday activity data. For that reason, special care should be taken to obtain an adequate learning rate for the ANNs, as a very big rate can prevent the convergence of the model. As the hybrid model provided the best classification results, mainly its results are discussed in the following.

Cycling with an exercise bike and regular bike introduced difficulties in this study. In our earlier study, we had measured acceleration from the wrist and chest. In that study, cycling with an exercise bike was detected with the accuracy of 75–82% [20]. In this study, an accelerometer placed on the hip could not produce a signal that could discriminate cycling and footsteps as well. This can be observed, for example, in the result summary in Table II. One can observe a clear tradeoff between the accurate detection of the two cycling activities and the rest of the activities. Automatically generated decision tree was the only classifier that could recognize the two cycling activities with nearly acceptable accuracy. However, this resulted in decrease in the detection accuracies of the other activities as well as in the total detection accuracy. Other classifiers concentrated on the other activities, thus leading to a worse detection of the cycling activities. The detection of cycling with a regular bike outdoors has better accuracy, as the GPS signal provides additional information for this task.

Football playing was detected with 88% accuracy from the supervised data that was a surprisingly high accuracy. However, it seems that the unsupervised period has included some movements similar to football, which degraded the recognition accuracy to 78% when all data were considered. Nevertheless, we feel that the accuracy is encouraging for future research in the recognition of more complex sports.

In supervised data, walking was detected with acceptable accuracy of 81%. If Nordic walking and walking had been considered a single class, the recognition accuracy would have been 93%. Including the unsupervised period decreased the accuracy with 10% unit for the hybrid model, which seems acceptable because the exact annotation of walking in different day-to-day situations is difficult as the walking periods may be of short duration.

Lying was detected with 97% and sitting and standing were recognized with 97% accuracy, as well. Seventy-eight percent of the unsupervised data comprised lying, sitting, or standing. This supports the assumption that the recognition of these passive activities is of major importance, as everything else not belonging to these activities can be considered more health enhancing. Thus, as a simple index of subject's overall activity, a percentage showing the amount of time spent in any other activity than these three could be used. For that purpose, the recognition accuracies obtained for these three activities in out-of-laboratory settings are encouraging.

In our previous study, we had recognized that the absence of accelerometers on the lower body (below waistline) was a limitation in the sensor setup. This was especially noticed as the inability to differentiate sitting and standing. For that reason, we repositioned the 3-D accelerometer from the chest level to the hip, as there were indications that such a placement could enable the discrimination of these two activities [23]. However, it became clear that, in our study, this discrimination was not possible regardless of the accelerometer replacement. As the subjects in our study wore sport clothes, the belt with the accelerometers had to be placed on top of the clothes. For that reason, it was not tightly connected to other clothes or the body, and the position of the accelerometers did not stay fixed. It seems that in order to obtain more precise acceleration information on hip, special attention must be paid on the sensor location and attachment. However, also such subject-dependent factors as body shape influence the sensor orientation on the waist. For that reason, the accurate discrimination of sitting and standing using waist-level accelerometry without user-specific training is complicated.

In the previous study, we used accelerometers with the sampling frequency of 200 Hz. For the current study, we dropped the sampling frequency to 20 Hz, which consumes less power. This decision was also backed up by other studies [6], [24], suggesting that such a sampling frequency should be enough. However, this proved to be a wrong decision because the impulses produced, for example, by a foot hitting the ground during running and a pole hitting the ground during Nordic walking diminished notably, and as the signal features used for discriminating these activities were based on the impulses, the activity-recognition accuracy also decreased.

For this study, 2 g accelerometers were replaced by 10 g ones, as we had noticed that the -2 to 2 g range is not sufficient during vigorous exercise. In general, -10 to 10 g scale was enough for the exercises on our protocol. Larger scale resulted in decreased signal resolution, but it seems that the decrease had negligible influence on the signal features.

Our future challenges include adjusting the activity-detection algorithms to real-time performance and for mobile devices. That way, the continuous monitoring of daily activities could be performed unobtrusively, and the changes in the daily durations of different activities could be reported that could motivate the user to prevent chronic diseases associated with physical inactivity.

ACKNOWLEDGMENT

The authors thank A. Ylisaukko-Oja, J. Vilmi, and P. Korpipää for help with the data-collection system. They also thank M. van Gils for help in data analysis and L. Cluitmans for data visualization.

REFERENCES

[1] *WHO World Health Report 2002*. [Online]. Available: http://www.who.int/whr/2002/en/whr02_en.pdf.

[2] T. Kottke, P. Puska, and J. Salonen, "Projected effects of high-risk versus population-based prevention strategies in coronary heart disease," *Amer. J. Epidemiol.*, vol. 121, pp. 697–704, 1985.

[3] D. Ashton, *Exercise, Health Benefits and Risks* (European Occupational Health Series 7), Copenhagen, Denmark: Villadsen & Christensen, 1993.

[4] R. Tonino, "Effect of physical training on the insulin resistance of aging," *Amer. J. Physiol. Endocrinol. Metab.*, vol. 256, pp. E352–E356, 1989.

[5] P. Dargent-Molina, F. Favier, H. Grandjean, C. Baudoin, A. Schott, E. Hausherr, P. Meunier, and G. Breart, "Fall-related factors and risk of hip fracture: The EPIDOS prospective study," *Lancet*, vol. 348, pp. 145–149, 1996.

[6] C. Bouten, A. Sauren, M. Verduin, and J. Janssen, "Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking," *Med. Biol. Eng. Comput.*, vol. 35, pp. 50–56, 1997.

[7] M. Fruin and J. Rankin, "Validity of a multi-sensor armband in estimating rest and exercise energy expenditure," *Med. Sci. Sports Exercise*, vol. 36, pp. 1063–1069, 2004.

[8] C. Bouchard, R. Shephard, and T. Stephens, Eds., *Physical Activity, Fitness, and Health: International Processings and Consensus Statement*. Champaign, IL: Human Kinetics, 1994.

[9] E. Tapia, S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Proc. 2nd Int. Conf. Pervasive Comput.*, 2004, pp. 158–175.

[10] M. Mathie, A. Coster, N. Lovell, and B. Celler, "Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiol. Meas.*, vol. 25, pp. R1–R20, 2004.

[11] K. Aminian and B. Najafi, "Capturing human motion using body-fixed sensors: Outdoor measurement and clinical applications," *Comput. Animation Virtual Worlds*, vol. 15, pp. 79–94, 2004.

[12] M. Mathie, A. Foster, N. Lovell, and B. Celler, "Detection of daily physical activities using a triaxial accelerometers," *Med. Biol. Eng. Comput.*, vol. 41, pp. 296–301, 2003.

[13] K. Aminian, P. Robert, E. Buchser, B. Rutschmann, D. Hayoz, and M. Depairon, "Physical activity monitoring based on accelerometry: Validation and comparison with video observation," *Med. Biol. Eng. Comput.*, vol. 37, pp. 304–308, 1999.

[14] F. Foerster and J. Fahrenberg, "Motion pattern and posture: Correctly assessed by calibrated accelerometers," *Behav. Res. Methods Instrum. Comput.*, vol. 32, no. 3, pp. 450–457, 2000.

[15] J. Ng, A. Sahakian, and S. Swiryn, "Accelerometer-based body-position sensing for ambulatory electrocardiographic monitoring," *Biomed. Instrum. Technol.*, vol. 37, pp. 338–346, 2003.

[16] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring." *Comput. Human Behav.*, vol. 15, pp. 571–583, 1999.

[17] D. Karantonis, M. Narayanan, M. Mathie, N. Lovell, and B. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 156–167, Jan. 2006.

[18] U. Maurer, A Smailagic, D. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," presented at the Int. Workshop Wearable Implantable Body Sensor Netw. Cambridge, MA, 2006.

[19] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. 2nd Int. Conf. Pervasive Comput.*, 2004, pp. 1–17.

[20] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006.

[21] J. Pärkkä, L. Cluitmans, and P. Korpipää, "Palantir context data collection, annotation and PSV file format," in *Proc. Pervasive 2004 Workshop*, pp. 9–16.

[22] R. Johnson and J. Shore, "Which is the better entropy expression for speech processing: −S log S or log S?," *IEEE Trans. Acoust.*, vol. 32, no. 1, pp. 129–137, Feb. 1984.

[23] S.-W. Lee and K. Mase, "Activity and location recognition using wearable sensors," *Pervasive Comput.*, vol. 1, no. 3, pp. 24–32, 2002.

[24] M. Sun and J. Hill, "A method for measuring mechanical work and work efficiency during human activities," *J. Biomech.*, vol. 26, pp. 229–242, 1993.

Miikka Ermes received the M.Sc. (Tech.) degree in digital signal processing from Tampere University of Technology, Tampere, Finland, in 2005.

He is currently working as a Research Scientist at the VTT Technical Research Centre of Finland, Tampere. His current research interests include measuring physical activity with wearable sensors and EEG signal processing.

Juha Pärkkä (S'01–M'07) received the M.Sc. (Tech.) degree in information technology (digital signal processing) from Tampere University of Technology, Tampere, Finland, in 1997.

He is currently working as a Senior Research Scientist at the VTT Technical Research Centre of Finland, Tampere. His current research interests include biomedical signal processing and ubiquitous computing.

Jani Mäntyjärvi received the M.Sc. degree in biophysics and the Ph.D. degree in information processing from the University of Oulu, Oulu, Finland, in 1999 and 2004, respectively.

He is currently a Senior Research Scientist at the VTT Technical Research Centre of Finland, Oulu. His current research interests include pervasive and context-aware computing for handheld devices, wearable sensing, and technologies for adaptive interaction.

Ilkka Korhonen (M'98) received the M.Sc. and D.Tech. degrees in digital signal processing from Tampere University of Technology, Tampere, Finland, in 1991 and 1998, respectively.

He is currently leading a team in Pervasive Health Technologies and is a member of the Strategic Steering Board for Information and Communication Technologies at the VTT Technical Research Centre, Oulu, Finland. He is a docent in medical informatics (with specialty on biosignal processing) in Ragnar Granit Institute, Tampere University of Technology. His current research interests include use of information and communication technology (ICT) for health and wellness, biosignal interpretation methods, and pervasive health care technologies, and especially, their application in critical care patient monitoring, wearable biomedical monitoring, home health monitoring, and eHealth/mHealth. He is the author or coauthor of more than 100 papers published in international scientific journals and conference proceedings.

Mr. Korhonen is a member of the IEEE Engineering in Medicine and Biology Society (EMBS) Technical Committee on Wearable Biomedical Sensors and Systems.

PUBLICATION **P3**

Estimating Intensity of Physical Activity

A Comparison of Wearable Accelerometer and Gyro Sensors and 3 Sensor Locations

In: Proceedings of the 29th Annual International Conference of the EMBS, Lyon, France, 23–26 August 2007. Pp. 1511–1514.

Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Estimating Intensity of Physical Activity: A Comparison of Wearable Accelerometer and Gyro Sensors and 3 Sensor Locations

J.Pärkkä, *Member, IEEE,* M.Ermes, K.Antila, M.van Gils, *Member, IEEE*, A.Mänttäri, H.Nieminen

***Abstract*—Automatic estimation of physical activity using wearable sensors can be used for promotion of a healthier lifestyle. In this study, accelerometers and gyroscopes attached to ankle, wrist and hip were used to estimate intensity of physical activity. The estimates are compared to metabolic equivalent (MET) obtained from a portable cardiopulmonary exercise testing system. Data from common everyday tasks and exercise were collected with 11 subjects. The tasks include, e.g., ironing, vacuuming, walking, running and cycling on exercise bicycle (ergometer). The strongest linear correlation with metabolic equivalent was obtained with the tri-axial accelerometer attached to the ankle (r=0.86).**

I. Introduction

According to World Health Organization (WHO), 60% of the world's population fail to follow the minimum recommendation of 30 minutes moderate intensity physical activity daily [1]. Physical inactivity is a severe health risk in modern societies. It is known to contribute to many chronic diseases such as cardiovascular disease, type 2 diabetes, cancers and osteoporosis [2],[3]. Thus all measures promoting a more active lifestyle are welcome. In aging societies, promotion of a more active lifestyle means better quality life for a great number of people and reduced health care costs.

One way to promote a healthier lifestyle is to develop methods that can automatically estimate physical activity or energy expenditure and show, e.g., a daily summary to the user. Intensity of physical activity and energy expenditure can be estimated objectively in many ways, e.g., with direct and indirect calorimetry, doubly labeled water, heart rate, temperature, ventilation, movement sensors, questionnaires and diaries [4]. Many of these methods are not applicable to long-term measurements in free-living conditions. Accelerometers have been used in many studies to estimate physical activity and energy expenditure in different tasks [4]-[8]. They have also been used for automatic recognition of daily activities [9]-[11]. However, use of gyro sensors (angular rate sensors) on estimation of physical activity or energy expenditure has not been studied extensively. In this study, we compare the performance of accelerometer and gyro sensors attached to ankle, hip and wrist in estimating intensity of physical activities. These sensors can be used unobtrusively also in free-living conditions in long term.

Manuscript received April 16, 2007. This work is funded by Nokia.

J. Pärkkä is with Technical Research Centre of Finland, P.O.Box 1300, 33101 Tampere, Finland (phone: +358 02 722 3346, fax: +358 02 722 3380, email: juha.parkka@vtt.fi.).

M. Ermes, K. Antila, M. van Gils are with Technical Research Centre of Finland, P.O.Box 1300, 33101 Tampere, Finland. ({miikka.ermes; kari.antila; mark.vangils}@vtt.fi).

A. Mänttäri is with UKK Institute, P.O.Box 30, 33501 Tampere, Finland (ari.manttari@uta.fi).

H. Nieminen is with Nokia, Itämerenkatu 11-13, 00180 Helsinki, Finland (heikki.v.nieminen@nokia.com).

II. Methods

A. Data Collection

Data were collected during common everyday tasks like household activities, walking, running and bicycling. The detailed list of tasks is given in Table I. The measurements were organized indoors in February 2007.

TABLE I
Tasks performed with 3 data loggers and a breathing gas analyzer

Task	Duration [min]
1. Hanging laundry	2
2. Ironing laundry	2
3. Folding laundry	2
4. Putting away laundry (on shelter)	2
5. Vacuuming	5
6. Walking stairs up	1
7. Walking stairs down	1
8. Walking and pushing shopping cart	5
9. Walking and carrying bags (3 laps, bags 2.7 kg each)	5
10. Walking (free pace)	3
11. Running (free pace)	2
12. Cycling on bike ergometer (65% of max performance)	5
BREAK	20
13. Walking on treadmill (35 % of maximal performance)	4
14. Walking on treadmill (45 % of maximal performance)	4
15. Walking on treadmill (55 % of maximal performance)	4
16. Running on treadmill (65 % of maximal performance)	4
17. Running on treadmill (75 % of maximal performance)	4
18. Running on treadmill (85 % of maximal performance)	4

The treadmill speeds and bike ergometer resistance were different for four user groups: male 25-35 yrs, female 25-35, male 50-60, female 50-60. The goal was to adjust the speeds so that they would be approximately equally strenous for these user groups. Estimate of a group's maximal performance was taken from a table of normative values of VO_{2max} with specific reference to age and sex [12]. The estimates were done by a sports testing professional.

Three stand-alone, battery-operated data loggers with accelerometer (Kionix KXPA4, Kionix Inc., Ithaca, NY, USA) and gyro (XV-3500, Epson Toyocom Corp., Tokyo Japan) sensors were used to collect data. The data loggers were attached firmly with special straps to wrist (non-dominating hand) and ankle (same side as wrist sensor) and with belt on hip. These sensor locations were chosen, because they were seen realistic in everyday use. Sampling rate of 50Hz was used. The accelerometer has +/- 18g and the gyro sensor +/- 100deg/s dynamical range. The three data loggers were synchronized offline with markers on data. The markers were generated by piling the data loggers on each other and knocking them simultaneously on table.

Reference data were collected with the portable cardiopulmonary exercise testing system (Metamax 3B, Cortex Biophysik GmbH, Leipzig, Germany). The Metamax data collection system is designed for exercise testing, thus the masks and equipment are light-weight, easy to carry and do not restrict movements. Clock of Metamax PC was synchronized in the beginning of each measurement session. The accuracy obtained was in the order of +/- 3 sec.

11 subjects took part in the data collection. The subjects were students and employees of participating organizations. The study has been approved by local ethical committee. The subjects' mean age was 38.6 (std 13.1), mean length 170.3 cm (8.5), mean weight 67.5 kg (10.7) and mean BMI 23.2 kg/m^2 (2.6). Before a subject was allowed to participate in the study, he filled in a health questionnaire and went through a health check. Each measurement session was supervised by professional nurse.

While supervising the measurement session, the nurse also annotated each task using an annotation application running on a PDA [13]. The application lists all task names with one radio button for each task plus a button for transition periods. Annotation of each task was turned on a few seconds after task start and turned off a few seconds before task end to avoid annotation of transition periods as tasks of interest. The PDA clock was synchronized with other clocks in the beginning of the measurement by tapping the screen simultaneously when the data loggers were knocked on table.

The total amount of data collected with 11 subjects was 10 h 7 min (in average about 55 min per case). The time spent in transition periods is not included in this figure.

B. *Signal processing*

The 50Hz signals from tri-axial accelerometers and gyro sensors were the starting points for estimation of metabolic equivalent. A modified version of the integral method [7] was used to compute the estimate for metabolic equivalent. The integral method takes absolutes of the 3D signals, integrates over the given period and sums the 3D signals into one signal (1)

$$IMA_{tot} = \int_{t=t_0}^{t_0+T} |a_x| dt + \int_{t=t_0}^{t_0+T} |a_y| dt + \int_{t=t_0}^{t_0+T} |a_z| dt \qquad (1)$$

,where a_x, a_y and a_z are the tri-axial sensor signals. We summed the 3 absolute values at each sample into one signal before integration. Before integration, the signal was also band-pass filtered (0.5 – 11 Hz) to highlight accelerations caused by human movements. The estimates were obtained by fitting a line on the data set (integral value vs. measured metabolic equivalent). The integral method was introduced in [7] and has been widely used since then in estimation of energy expenditure from accelerometer data.

Unit of the metabolic equivalent is MET. One MET is by definition obtained at complete rest. The moderate intensity activities are those, whose intensity is 3-6 METs, thus consuming 3-6 times the energy consumed in rest. Such activities are typically leisure time walking and cycling. In this study, metabolic equivalent from Metamax was used as target value. One MET equals 3.5 ml of consumed oxygen per kg per minute (1 MET = 3.5 ml/kg/min VO2). Thus, the value is more convenient than VO2 (oxygen consumption, l/min), because it is relative to subject's own resting metabolic rate and takes into account the subject's body weight. This makes it easier to pool together data from different subjects. As metabolic equivalent is obtained from a breathing gas analyzer, it is a breath-by-breath value, thus no regular sampling is available.

To find a period of each task, where the oxygen uptake represents well the true intensity of physical activity without the previous activities affecting too much the value, a 30-second "steady state" period of each task was selected for comparison with metabolic equivalent. The period used in comparison is the 30 last seconds of each task as taken from the annotation. The median of metabolic equivalent in this 30-sec period is used as reference.

III. Results

Pearson linear correlation was computed between the estimates and measured metabolic equivalent. The MET estimate was calculated using the integral method separately for ankle, hip and wrist sensors. Table II summarizes the results. The correlations are also depicted in Fig 1 for each sensor type and sensor location.

TABLE II
NUMBER OF TASKS, PEARSON CORRELATION AND RMS ERROR BETWEEN MEASURED MET AND MET ESTIMATES

	Acc Ankle	Acc Hip	Acc Wrist	Gyro Ankle	Gyro Hip	Gyro Wrist
N	163	177	178	163	177	178
r	0,86	0,80	0,81	0,84	0,69	0,48
RMSE [MET]	1,21	1,42	1,40	1,32	1,71	2,09

The best correlation, r=0.86, was obtained using the ankle acceleration sensor. This comparison comprised 156 of max 198 tasks (11 subjects, 18 tasks each). The missing tasks were due to hardware problems. Also visually assessed, the data seem to have rather a nice linear relationship. The RMS error between linear fit and data was 1.17 MET.

The measured MET and estimated METs from different sensors and sensor locations are plotted task-wise in Fig 2. Measured METs are given task-wise in Table III for different age and gender groups.

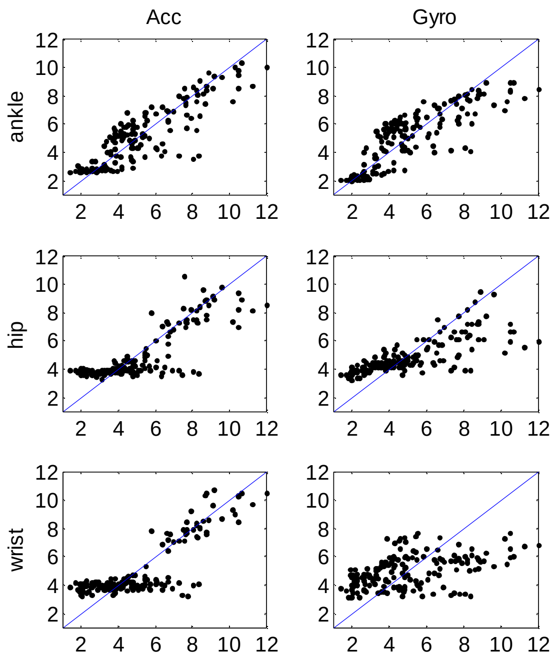


Fig. 1 Measured MET (x) vs. estimated MET (y) from accelerometer and gyro sensors. Sensor locations are ankle, hip and wrist. The lines across the plots show the output of an ideal estimator.

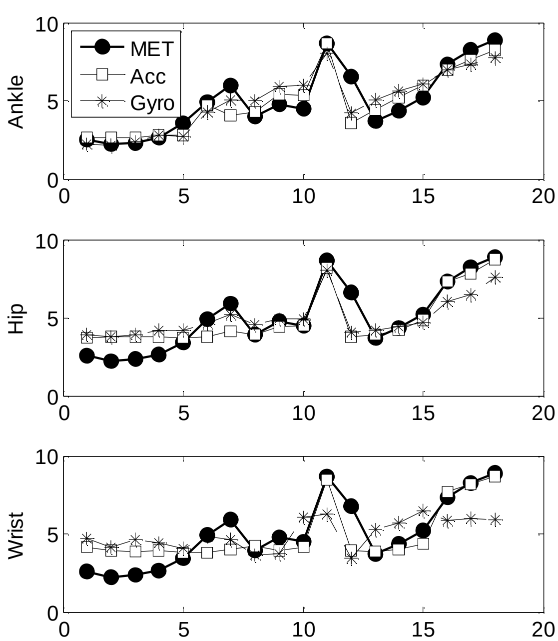


Fig. 2 Task-wise mean of measured METs and estimates (filled circle: measured MET, open square: estimate from accelerometer, asterisk: estimate from gyro sensor. Sensor locations are ankle (topmost), hip (middle) and wrist (bottom). Names for task numbers can be found in Tables I and III.

TABLE III
MEASURED METABOLIC EQUIVALENTS (MET) FOR MALES IN AGE GROUP 25-35, FEMALES 25-35, MALES 50-60 AND FEMALES 50-60 (N=4,3,2,2, RESPECTIVELY)

Task	M 25-35	F 25-35	M 50-60	F 50-60
1. Hanging laundry	2,5	2,4	3,1	2,9
2. Ironing laundry	2,1	2,2	2,4	2,4
3. Folding laundry	2,3	2,0	2,6	2,7
4. Putting away laundry (on shelf)	2,6	2,4	2,8	3,2
5. Vacuuming	3,4	3,3	3,0	4,2
6. Walking stairs up	5,0	4,0	4,6	5,5
7. Walking stairs down	5,6	6,4	6,2	5,2
8. Walking and pushing shopping cart	4,0	4,5	2,9	4,0
9. Walking and carrying bags	4,5	5,5	4,1	5,0
10. Walking	4,3	5,4	3,9	4,8
11. Running	9,5	7,6	8,7	8,3
12. Cycling on bike ergometer (65%)	8,1	7,2	5,9	4,8
BREAK				
13. Walking on treadmill (35 %)	4,2	3,4	3,6	3,5
14. Walking on treadmill (45 %)	4,9	4,1	3,7	4,0
15. Walking on treadmill (55 %)	6,0	4,8	4,7	4,2
16. Running on treadmill (65 %)	8,9	6,6	7,1	5,4
17. Running on treadmill (75 %)	9,9	7,5	7,5	6,4
18. Running on treadmill (85 %)	10,6	8,2	8,2	7,8

IV. DISCUSSION

The best sensor location for estimating MET was ankle (Table II, Fig 1, Fig 2). Wrist and hip gave smaller linear correlations. This was a clear, but slightly surprising result, since many studies have concluded that accelerometers attached to human trunk instead of extremities should give best estimates of energy expenditure. The reason for the different result might lie in task set used. All tasks in our data set are performed in standing posture and in most tasks feet also do most of the work. Our data set does not include activities that heavily involve hand movements. Also carrying bags involves more foot movements than hand movements although hands do static work in that task.

Crouter et al. [5] used a hip actigraph to estimate energy expenditure. They suggest use of two different regression equations for different activities. They divided activities into low-variability (e.g. running) and high variability signals (e.g. household activities). With our data set, one linear fit gave a good estimate for ankle signal (Fig 1). In our study, a constant value would represent low activity (high variability) task estimates very well, when hip or wrist accelerometers are used. Gyro sensors on the other hand are more sensitive to small intensity changes in low activity tasks.

Task-wise analysis (Fig 2.) of accelerometer estimates reveals that the estimates obtained with ankle accelerometer perform rather well in all other tasks except walking stairs down (7) and cycling (12). This is probably due to lack of steps in cycling (spinning does not produce similar accelerations), and effect of walking stairs up before walking down (duration of these activities was too short, previous activity affects measured MET value). Wrist and hip accelerometers, on the other hand, react only to tasks involving running. Other tasks are given more or less a constant estimate of MET with wrist and hip accelerometers. Summary of task-wise analysis is provided in Table IV.

Task-wise analysis of gyro sensor estimates (Fig 2.) shows the sensor attached on ankle gives second best estimate (r=0.84) of MET after ankle accelerometer. Ankle gyro signal integral tends to overestimate intensity of walking (tasks 8-10, 13-15) and underestimate intensity of running (tasks 11, 16-18) more than ankle accelerometer. Intensity of cycling is heavily underestimated also with ankle gyro.

Hip gyro gives an accurate estimate of all tasks involving walking (tasks 5-10, 13-15). It also gives accurate intensity estimate of running with free pace (task 11). However, it underestimates intensity of running on treadmill and cycling. It also overestimates intensity of household activities.

Wrist gyro has a rather distinct profile from all other sensor-location combinations. It gives accurate intensity estimates for vacuuming, walking stairs up, walking and pushing shopping cart. It overestimates intensity of household activities (tasks 1-4) and walking activities (tasks 10, 13-15). It underestimates intensity of running (tasks 11, 16-18), cycling, walking stairs down and walking and carrying bags. It clearly reacts on all periodic hand movements (e.g., tasks 10-11, 13-18).

Measured metabolic equivalents are listed in Table III for males and females and for age groups 25-35 and 50-60. Household activities (tasks 1-5) generally do not reach the level of moderate intensity (3-6 MET) physical activity, except vacuuming. The laundry activities done were considered rather easy compared to "real" laundry work. These tasks were also short in duration. Their intensity in free living conditions is probably higher than what was measured in this study. Walking in different forms (free pace, with bags, with shopping cart, stairs) are typical moderate intensity activities. Running in general exceeds the intensity defined as moderate intensity. It must be kept in mind that to gain the positive health effects, the 30-min-moderate-intensity activity should be done at least in periods of 10 minutes continuous activity. When exercising longer or harder than moderate intensity, more health benefits can be obtained. When comparing age and gender groups, it is notable that the treadmill speeds in female groups and especially in the female 50-60 group were not as demanding as the speeds in male groups.

TABLE IV SUMMARY OF TASK-WISE ANALYSIS. SYMBOL ↗ MEANS THE ESTIMATE OVERESTIMATES TRUE MET (↘ : UNDERESTIMATES)

	Laundry	Vacuum	Stairs	Walk	Run	Cycle
Ankle Acc		↘	↘			↘
Hip Acc	↗		↘			↘
Wrist Acc	↗		↘			↘
Ankle Gyro		↘		↗	↘	↘
Hip Gyro	↗				↘	↘
Wrist Gyro	↗			↗	↘	↘

CONCLUSION

This study gave us valuable information on correlation of wearable sensor signals and intensity of different physical activities. Based on the results of this study, a larger study with more subjects is started.

ACKNOWLEDGMENT

The authors want to express their gratitude to Kirsi Mansikkamäki and Ulla Hakala for their work during data collection.

REFERENCES

[1] P. Puska, H. Benaziza, D. Porter, Physical Activity, WHO Information Sheet on Physical Activity, WHO 2003. Available: http://www.who.int/dietphysicalactivity/media/en/gsfs_pa.pdf

[2] H. Hagendoorn, I. Vuori, P. Oja, Guidelines for the development of national policies and strategies for promoting health through physical activity, The European Network for the Promotion of Health-Enhancing Physical Activity, 2002, Available: http://www.who.int/gb/ebwha/pdf_files/WHA57/A57_R17-en.pdf

[3] R. Pate, M. Pratt, S. Blair, W. Haskell, C. Macera, C. Bouchard, D. Buchner, W, Ettinger, G. Heath, C. King, et al., " Physical Activity and Public Health: a Recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine", *JAMA*, vol. 273, no. 5, pp. 402-407, Feb 1, 1995.

[4] M.J. Lamonte, B.E.Ainsworth "Quantifying energy expenditure and physical activity in the context of dose response", Medicine & Science in Sports & Exercise, vol 33(6), suppl., pp. 370-378, 2001.

[5] S.E.Crouter, K.G.Clowers, D.R.Bassett, "A novel method for using accelerometer data to predict energy expenditure", *J Appl Physiol*, vol 100, pp. 1324-1331, 2006.

[6] A. Wixted, D. Thiel, D. James, A. Hahn, C. Gore, D. Pyne, "Signal processing for estimating energy expenditure of elite athletes using triaxial accelerometers, *IEEE Sensors*, vol. 7, issue 4, pp.798-801, 2007.

[7] C. Bouten, K. Koekkoek, M. Verduin, R. Kodde, J. Janssen, "A Triaxial Accelerometer and Portable Data Processing Unit for the Assessment of Daily Physical Activity", *IEEE Transactions on Biomedical Engineering*, vol. 44, pp. 136-147, 1997.

[8] M. Mathie, A.C.F. Coster, N.H.Lovell, B.G.Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement", *Physiological Measurement*, vol. 25, pp. R1-R20, 2004.

[9] L. Bao, S. Intille, "Activity Recognition from User-Annotated Acceleration Data", In *Proc. of 2nd International Conference on Pervasive Computing*. pp. 1-17, 2004.

[10] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, I. Korhonen "Activity Classification Using Realistic Data from Wearable Sensors", *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, pp.119-128, 2006.

[11] M. Ermes, J. Pärkkä, J. Mäntyjärvi, I. Korhonen "Detection of Daily Activities and Sports with Wearable Sensors in Controlled and Uncontrolled conditions", *IEEE Transactions on Information Technology in Biomedicine*,(accepted for publication).

[12] B.A.Franklin, M.H.Whaley, E.T.Howley, G.J.Balady, *ACSM's guidelines for exercise testing and prescription*. Philadelphia, Lippincott Williams & Wilkins, 2000.

[13] J. Pärkkä, L. Cluitmans, P. Korpipää, "Palantir Context Data Collection, Annotation and PSV File Format", in *Pervasive 2004 Workshop Proceedings: Benchmarks and a database for context recognition*, pp 9-16, 2004.

PUBLICATION **P4**

Personalization Algorithm for Real-Time Activity Recognition using PDA, Wireless Motion Bands and Binary Decision Tree

In: IEEE Transactions on Information Technology in Biomedicine 2010. Vol. 14, No. 5, pp. 1211–1215
Reprinted with permission from the publisher.

Personalization Algorithm for Real-Time Activity Recognition Using PDA, Wireless Motion Bands, and Binary Decision Tree

Juha Pärkkä, Luc Cluitmans, and Miikka Ermes

***Abstract*—Inactive and sedentary lifestyle is a major problem in many industrialized countries today. Automatic recognition of type of physical activity can be used to show the user the distribution of his daily activities and to motivate him into more active lifestyle. In this study, an automatic activity-recognition system consisting of wireless motion bands and a PDA is evaluated. The system classifies raw sensor data into activity types online. It uses a decision tree classifier, which has low computational cost and low battery consumption. The classifier parameters can be personalized online by performing a short bout of an activity and by telling the system which activity is being performed. Data were collected with seven volunteers during five everyday activities: lying, sitting/standing, walking, running, and cycling. The online system can detect these activities with overall 86.6% accuracy and with 94.0% accuracy after classifier personalization.**

***Index Terms*—Binary decision tree, classifier personalization, real-time activity recognition.**

I. Introduction

REGULAR physical activity is known to produce long-term health benefits. In 1995, the World Health Organization (WHO) reported that 60% of the world's population fails to achieve the minimum recommendation of 30 min moderate intensity physical activity daily [1]. In a recent study [2], compliance with physical activity recommendations was assessed with 78 postal workers in the U.K. Only 10% of the participants succeeded in complying with the 30-min daily recommendation. Today's sedentary lifestyle and parents' habits also affect the children. In a recent study [3], 17% of the 7-year-old Finnish school children were overweight and 5% were obese. Thus, the benefits of a more active lifestyle should be emphasized not only among the working-age population, but also among children and elderly. Physical activity guidelines for Americans [4] were recently published to promote physical activity. The new guidelines recommend 60 min of physical activity daily. The recommendation also emphasizes that the daily physical activity should include 1) aerobic exercise; 2) muscle-strengthening exercise; and 3) bone-strengthening exercise. The guidelines are important in educating people about the level of physical activity required to achieve health benefits. However, still there is a clear need to motivate people to start exercising and to continue exercising regularly.

Methods for objectively measuring the level of physical activity have focused traditionally on the indirect measurement or assessment of energy expenditure. Several methods have been developed to assess the energy expenditure with unobtrusive sensors as part of everyday life (e.g., step counters and heart rate monitors). Accelerometers have been shown to be especially suitable for estimating the human movement [5]. However, more research is needed to objectively measure the type of activity (e.g., aerobic, muscle-strengthening, and bone-strengthening). Previous studies have shown that it is possible to infer the daily distribution of activities into different activity types with good accuracy using offline [6]–[8] and online [9] methods based on accelerometer measurements.

Methods commonly used for activity classification were recently reviewed in [10]. Methods like artificial neural networks, support vector machines, *K*-nearest neighbor, decision trees, Bayesian classifiers, etc., are commonly used, but no single classification scheme has proven to be superior in this task. When porting the activity-recognition algorithms from PC into mobile devices, complexity of the algorithm becomes an issue. A complex algorithm consumes more time and more processing power. In our recent study [11], we compared different classification algorithms running on a mobile phone. The results show that very simple classification algorithms outperform more complex algorithms on a battery-powered device. The battery lifetime of an activity-recognition application running with a simple algorithm is significantly longer than with an algorithm requiring intensive computation. However, when using a simple classification algorithm, the features used as inputs to the classifier have to be carefully selected to maintain good classification accuracy.

In this study, we report results of an activity-recognition algorithm based on a decision tree classifier to automatically recognize physical activities on a portable device, online. We also report results of a personalization algorithm used to update the decision tree threshold values with user's own data. The personalization represents a situation, where a customer buys a device, whose algorithm uses default parameters that have been trained to give optimum performance for average users. After buying the device, the user might be interested in personalizing the activity-recognition algorithm to achieve better recognition accuracy. This procedure is necessary, especially, for people

Manuscript received September 8, 2009; revised March 20, 2010; accepted June 14, 2010. Date of publication June 28, 2010; date of current version September 3, 2010. This work was supported in part by TEKES (The Finnish Funding Agency for Technology and Innovation) under the project "Ramose—Data Processing and Decision Making Methods for Mobile, Distributed Computing Environment."

The authors are with the VTT Technical Research Centre of Finland, P.O. Box 1300, Tampere 33101, Finland (e-mail: juha.parkka@vtt.fi; luc.cluitmans@vtt.fi; miikka.ermes@vtt.fi).

Digital Object Identifier 10.1109/TITB.2010.2055060

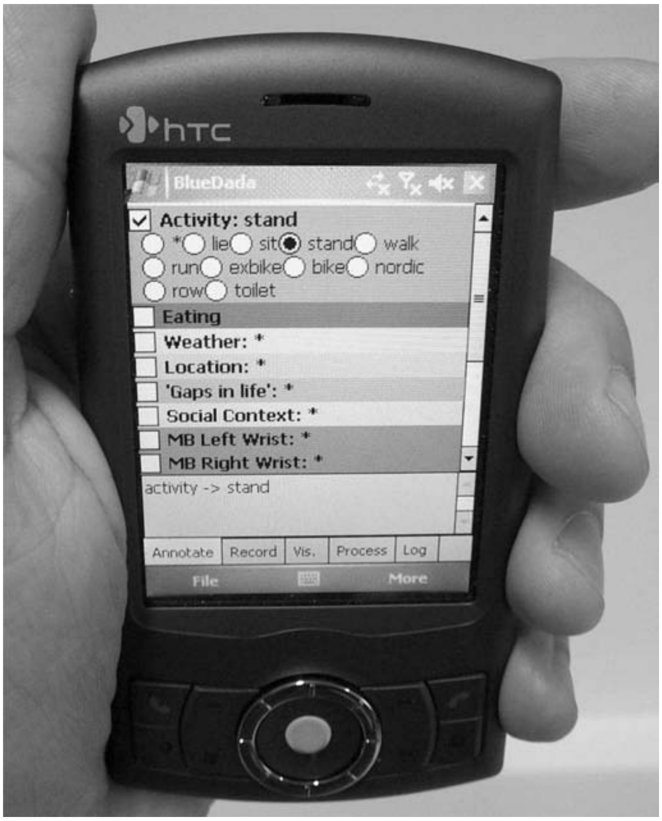


Fig. 1. PDA with annotation screen of activity recognition software.

performing the activities in a different way (e.g., different pace or intensity) than the default users.

II. Methods

A. Data Collection

The goal of our data collection was to evaluate the activity-recognition system running on a portable device and the usefulness of the personalization feature in improving the classification accuracy. The central device in data collection and activity recognition is a personal digital assistant (PDA) (HTC P3300, HTC, Taiwan). The PDA is running an application that is based on an annotation application used in earlier studies [12] (see Fig. 1). In addition to annotation functions, the current application receives data over Bluetooth connection, computes feature signals from raw data online, classifies the data in second-by-second basis online, and stores the data on a memory card. When the personalization feature is used, the user can tell the system which activity he is performing, by annotating his activity type with start and end times using the pen stylus.

The human movements were quantified with Nokia wireless motion bands [13] using the 3-D accelerometer signal and Bluetooth connection for data transfer to the PDA (see Fig. 2). Fifty Hertz of sampling rate was used with accelerometers. Data were collected with wireless motion bands attached to volunteers' both ankles and wrists. The wireless motion bands were attached using velcro straps. The velcro straps were adjustable so that firm attachment was possible for all volunteers.

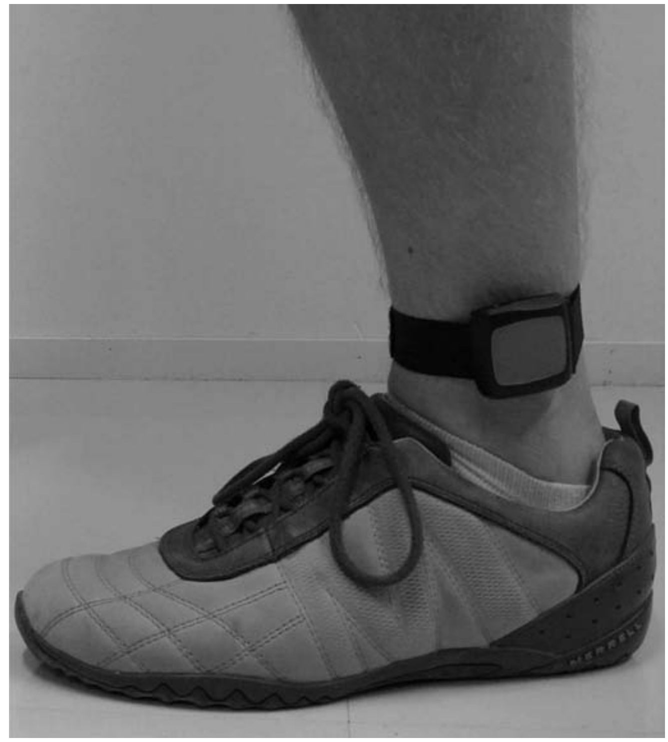


Fig. 2. Wireless motion band attached on ankle.

TABLE I
Activities During Evaluation

Task	Duration
Lying	5 min
Sitting	5 min
Standing	5 min
Walking	5 min
Bicycling	5 min
Running	5 min

The activity-recognition system was evaluated with seven volunteers. Median volunteer age was 27 yrs (range 4–37 yrs), and median length was 180 cm (range 92–187 cm) (see Table V). The volunteers performed six activities according to the activity plan (see Table I) and annotated these using the PDA.

B. Feature Extraction and Classification

Four features were computed from the raw sensor data. These were 1) intensity of highest peak in power spectral density (PSD); 2) signal average; 3) signal spectral entropy; and 4) signal variance. The time-domain features were computed from the most recent 255 samples (5 s), and the frequency-domain features were computed from the same 255 samples and one added zero for efficient fast Fourier transform (FFT) implementation with 256 samples. In our earlier study, we found that features computed from an ankle sensor signal represent the activity type very well, better than those attached to wrist or hip [14]. Combination of the ankle sensor data with other sensor data did not improve classification accuracy. Thus, only ankle sensor data are used for computing the feature signals. Furthermore, only the vertical direction of the 3-D sensor is used. This has the

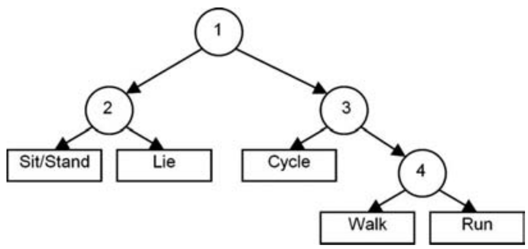


Fig. 3. Binary decision tree used for activity recognition on PDA.

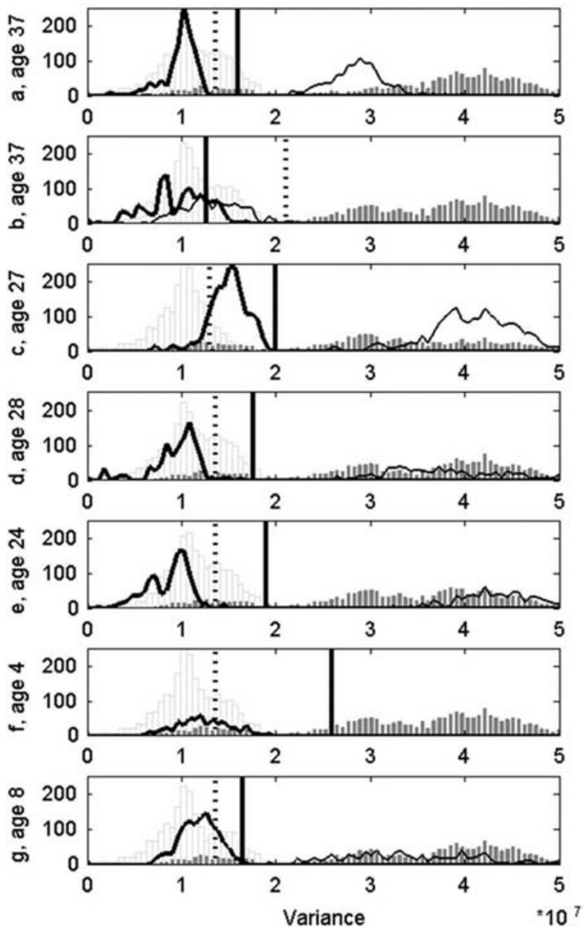


Fig. 4. Effect of personalization. Variance of volunteers' data in node 4, which separates walking and running. Bars represent training data, and line plots represent the individual data (leave-one-out cross-validation). Line plots have been multiplied with 3 to make them better visible. Data during walking is seen on the left as open bars and thick line. Running is represented on the right as filled bars and thin line. Cases (a)–(e) are adult data, cases (f) and (g) are child data. The ages of cases (a)–(g) are 37, 37, 27, 28, 24, 4, and 8 years, respectively. Vertical lines show node 4 threshold, dotted vertical line is the original threshold obtained with training data, and the solid line is that after personalization.

advantage that the activity-recognition system is simple and can be implemented with small amount of rather simple sensors.

Structure of the binary decision tree used for automatic activity recognition with PDA includes four nodes (see Fig. 3). The tree is structured so that there is one threshold value in each node. Feature signal samples with values larger than the threshold fall into the right branch, while those smaller than the threshold fall into left branch. Node 1 utilizes intensity of the highest peak in PSD to discriminate activities containing regular movements from static activities. Node 2 utilizes signal average to discriminate upright positions from positions in horizontal direction. Node 3 discriminates cycling from walking and running with the help of spectral entropy. Cycling produces one peak in PSD without harmonics, while walking and running produce a peak with multiple harmonics. The harmonics increase signal entropy in walking and running. Node 4 utilizes signal variance to discriminate walking from running. Sitting and standing were combined into one activity, because they both represent static activity with low energy consumption.

TABLE II
ONLINE, SUPERVISED PERSONALIZATION ALGORITHM FOR UPDATING DECISION TREE THRESHOLDS

1) discard 30 sec of new data from start and end to avoid possibly false annotations in start and end of activity
2) compute the features needed in each node
3) for each feature, compute 10^{th} and 90^{th} percentiles
4) for the new activity, the new representative 10^{th} percentile is a weighted mean of training data 10^{th} percentile and 10^{th} percentile of the new, personal data (personal data is given 75% weight)
5) similarly to 4, update representative 90^{th} percentiles for each activity
6) for each node, see which activities fall to left and which to right branch
i) select maximum of the 90^{th} percentiles of the data falling left
ii) select minimum of the 10^{th} percentiles of the data falling right
iii) personalize the threshold by computing mean of the two values mentioned above

A personalization algorithm was developed and used to improve classifier accuracy. The algorithm searches for optimum decision boundary between the activities falling left and right in each node (see Fig. 4). When using the device for automatic activity recognition, the user can annotate his activities and thus personalize his algorithm online. The personalization algorithm takes 3–10 min of new data with annotation and uses that for updating the thresholds in each node. The updating rules are given in Table II.

III. RESULTS

All data from the seven volunteers were used in evaluation. Leave-one-subject-out cross-validation was used in training and testing the classifier. Thus, data of one volunteer was left out and threshold values were defined based on the data from six volunteers. The obtained classifier was then tested on the subject's data that were left out earlier. The same procedure was repeated for all volunteers. Confusion matrix was computed by summing the seven confusion matrices together.

The original confusion matrix as obtained without personalization shows a rather good overall performance to other activities, except walking (see Table III). The original accuracy is 86.6%. This is the normalized accuracy, meaning that each activity has equal weight even if they would have different amount of samples.

The confusion matrix after personalization shows an improved accuracy (see Table IV). The accuracy after personalization is 94%.

Individual classification accuracies for each volunteer in original stage and after introducing data for personalization were computed (see Table V). Data for personalization are presented

TABLE III
CONFUSION MATRIX OF ORIGINAL DECISION TREES [IN PERCENTAGE]

Annotation	Recognized Activity				
	Sit/stand	Lie	Cycle	Walk	Run
Sit/stand	100	0	0	0	0
Lie	0	100	0	0	0
Cycle	4	0	94	1	1
Walk	17	0	31	48	4
Run	2	0	1	13	84

TABLE IV
CONFUSION MATRIX OF PERSONALIZED DECISION TREES [IN PERCENTAGE]

Annotation	Recognized Activity				
	Sit/stand	Lie	Cycle	Walk	Run
Sit/stand	100	0	0	0	0
Lie	0	100	0	0	0
Cycle	5	0	94	0	1
Walk	17	0	1	79	4
Run	2	0	0	5	93

TABLE V
INDIVIDUAL CLASSIFICATION ACCURACIES IN ORIGINAL STAGE AND AFTER INTRODUCTION OF EACH NEW SET OF PERSONAL TRAINING DATA [IN PERCENTAGE]

ID(sex,age,length)	Original	Lie	Sit	Stand	Walk	Bike	Run
Case 1 (M,37,180)	87	87	87	87	80	96	99
Case 2 (F,37,156)	80	80	80	80	73	80	89
Case 3 (M,27,180)	79	79	79	79	80	81	99
Case 4 (M,28,186)	90	90	90	90	87	87	90
Case 5 (M,24,187)	88	88	88	88	84	88	88
Case 6 (M,4,92)	74	74	74	74	70	71	77
Case 7 (M,8,128)	95	95	95	95	88	89	99

TABLE VI
MISSING SAMPLES PER MINUTE

ID	Missing samples (%)
Case 1	0.22
Case 2	0.01
Case 3	0.03
Case 4	0.86
Case 5	0.41
Case 6	2.02
Case 7	0.16
TOTAL	0.43

to the algorithm one activity at a time (about 5 min of annotated data) and accuracies are computed after introduction of each new activity.

Some packets were lost due to the wireless data transfer. The number of missing samples per case ranges from 0.01% to 2.02% (see Table VI). Overall, 0.43% of samples are missing on each minute of the data received over the wireless Bluetooth connection.

IV. DISCUSSION

A simple and effective online classifier for activity recognition using wireless sensors and a PDA was presented. The classifier is accompanied with a personalization algorithm. The selected classifier, a binary decision tree, is an effective algorithm, requiring only a few comparisons and thus consuming very little battery power compared to more complex classifiers.

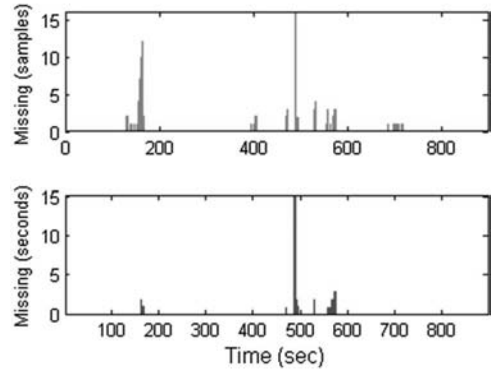


Fig. 5. Missing samples in one recording. In each second, 50 new samples per channel are expected to arrive over the wireless Bluetooth connection. The figure shows two layers of missing data. (Top panel) Number of samples missing, when some packets are received (short breaks). (Bottom panel) Number of seconds with no incoming packets (longer breaks). In the bottom panel, bars are placed on the time, when the next packets after a break are received.

Despite of its low computational cost, it can produce decent classification accuracy with carefully selected features.

The binary decision tree algorithm has been found very efficient, requiring only simple comparisons and doing classification rapidly [15]. A decision tree classifier can classify many orders of magnitude faster than most classifiers that depend on distance calculation between input pattern and stored exemplars [16]. During classification, a decision tree consumes very little memory.

In this study, the original classification accuracies for static activities are almost perfect. Also cycling can be detected almost perfectly. Walking and running however are being mixed. One reason for this is that some people ran with very low speed and with very low step, so that the normal acceleration impact when the heel hits the ground in running does not appear in the data. In a way, it is thus justified to say that the activity performed is walking instead of running. Some part of activity annotated as walking is also recognized as sitting or standing. In this study, this is only true in cases when the volunteer had to stop, e.g., because of traffic lights.

The personalization algorithm improves accuracy on average with 7.4%. Thus, there is a trend that personalization improves classification accuracy; however, the dataset is too limited to prove statistical significance. Interestingly, with all except case 3, introduction of personal walking data first decreases classification accuracy. The decrease is corrected by introducing more activities (cycling, running) later. However, it gives an indication that it is advisable to introduce personal data with multiple activities instead of just one.

The poor classification accuracy of case six data is due to lots of short connection breaks (see Tables V and VI). The personalization algorithm finds the optimal threshold values, but cannot compute features when data are missing. Wireless transmission of data to PDA is not as reliable as data transfer with wired connection. The sensor indexed measured samples with a running index from 0 to 254, thus in blocks of approximately 5 s. We experienced two types of breaks: short breaks of less than 5 s and longer breaks leading to no incoming packets from the sensor (see Fig. 5). When data are collected with multiple sensors simultaneously, the breaks seem more common than when using only a couple of sensors.

Currently, the bottleneck in wide utilization of such activity-recognition systems is the battery consumption. An effective classifier limits battery consumption of the mobile device, but one problem remains. Wireless transmission of data from sensor unit to the mobile device consumes too much power. Currently, the battery in wireless motion bands drains in 0.5–1 h. It is not yet enough for getting the whole-day distribution of activities.

V. Conclusion

A simple and effective decision tree classifier and a personalization algorithm were implemented on PDA. The online system can detect the most common daily activities with overall 86.6% normalized accuracy and with 94.0% normalized accuracy after classifier personalization.

Acknowledgment

The authors would like to thank volunteers of this study for performing the activities and for collecting the data. They would also like to thank Nokia for providing the wireless motion bands for this study.

References

[1] R. R. Pate, M. Pratt, S. N. Blair, W. L. Haskell, C. A. Macera, C. Bouchard, D. Buchner, W. Ettinger, G. W. Health, A. C. King, A. Kriska, A. S. Leon, B. H. Marcus, J. Morris, R. S. Paffenbarger, K. Patrick, M. L. Pollock, J. M. Rippe, J. Sallis, and J. H. Wilmore, "Physical activity and public health: A recommendation from the centers for Disease Control and Prevention and the American College of Sports Medicine," *JAMA*, vol. 273, no. 5, pp. 402–407, Feb. 1995.

[2] S. F. M. Chastin, P. M. Dall, W. W. Tigbe, M. P. Grant, C. G. Ryan, D. Rafferty, and M. H. Granat, "Compliance with physical activity guidelines in a group of UK-based postal workers using an objective monitoring technique," *Eur. J. Appl. Physiol., 106*, pp. 893–899. [Online]. Available: http://www.springerlink.com/content/30n21232w13155t2/

[3] M. Vanhala, R. Korpelainen, P. Tapanainen, K. Kaikkonen, H. Kaikkonen, T. Saukkonen, and S. Keinänen-Kiukaanniemi, "Lifestyle risk factors for obesity in 7-year-old children," *Obes. Res. Clin. Pract.*, vol. 3, pp. 99–107, 2009.

[4] M. O. Leavitt, Ed., *2008 Physical Activity Guidelines for Americans* [Online]. Washington DC, U.S: Department of Health and Human Services, Oct. 2008. Available: www.health.gov/paguidelines.

[5] M. Mathie, A. Coster, N. Lovell, and B. Celler, "Accelerometry: Providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiol. Meas.*, vol. 25, pp. R1–R20, 2004.

[6] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. 2nd Int. Conf. Pervasive Comput.*, 2004, pp. 1–17.

[7] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, Jan. 2006.

[8] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Trans. Inf. Technol. Biomed*, vol. 12, no. 1, pp. 20–26, Jan. 2008.

[9] M. Ermes, J. Pärkkä, and L. Cluitmans, "Advancing from offline to online activity recognition with wearable sensors," in *Proc. 30th EMBS Conf.*, 2008, pp. 4451–4454.

[10] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, and R. Crompton, "Activity identification using body-mounted sensors—A review of classification techniques," *Physiol. Meas.*, vol. 30, pp. R1–R33, 2009.

[11] V. Könönen, J. Mäntyjärvi, H. Similä, J. Pärkkä, and M. Ermes, "Automatic feature selection for context recognition in mobile phones," *Pervasive Mobile Comput.*, vol. 6, pp. 181–197, 2010.

[12] J. Pärkkä, L. Cluitmans, and P. Korpipää, "Palantir context data collection, annotation and PSV file format," in *Proc. Pervasive 2004 Workshop Proc.: Benchmarks Database Context Recognit.*, 2004, pp. 9–16.

[13] K. Laurila, T. Pylvänäinen, S. Silanto, and A. Virolainen, "Wireless motion bands," presented at the UbiComp 2005 Workshop "Ubiquitous computing to support monitoring, measuring and motivating exercise," Tokyo, Japan, Sep. 11–14, 2005.

[14] J. Pärkkä, M. Ermes, K. Antila, M. van Gils, A. Mänttäri, and V. Nieminen, "Estimating intensity of physical activity: A comparison of wearable accelerometer and gyro sensors and 3 sensor locations," in *Proc. 29th EMBS Conf.*, 2007, pp. 3056–3059.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[16] Y. Lee and R. P. Lippmann, "Practical characteristics of neural network and conventional pattern classifiers on artificial and speech problems," in *Advances in Neural Information Processing Systems 2*, D. S. Touretzky, Ed. San Francisco, CA: Morgan Kaufmann, 1990, pp. 168–177.

Juha Pärkkä received the M.Sc. (Tech.) in information technology (digital signal processing) from Tampere University of Technology, Tampere, Finland, in 1997.

He is currently a Senior Research Scientist at VTT Technical Research Centre of Finland, Tampere. His research interests include biomedical signal processing, ubiquitous computing, and personal health systems.

Luc Cluitmans received the M.Sc. and Ph.D. degrees in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 1992 and 1999, respectively.

He is currently a Senior Research Scientist at VTT Technical Research Centre of Finland, Tampere, Finland. His research interests include biomedical signal acquisition and processing.

Miikka Ermes received the M.Sc. (Tech.) and D.Sc. (Tech.) degrees in digital signal processing from Tampere University of Technology, Tampere, Finland, in 2005 and 2009, respectively.

He is currently a Research Scientist in the Personal Health Systems Team, VTT Technical Research Centre of Finland, Tampere. His research interests include context recognition with wearable sensors and electroencephalogram signal processing.

PUBLICATION **P5**

Automatic Feature Selection and Classification of Physical and Mental Load using Data from Wearable Sensors

In: Proceedings of the 10th International Conference on Information Technology and Applications on Biomedicine. ITAB 2010, Corfu, Greece, 2.–5.11.2010. Pp. 1–5.

Proceedings of the 10th IEEE International ITAB Conference
Corfu, Greece, 2-5 Nov 2010

Automatic Feature Selection and Classification of Physical and Mental Load using Data from Wearable Sensors

Juha Pärkkä, Miikka Ermes, Mark van Gils

Abstract—**Long-term monitoring of health is essential in many chronic conditions, but automatic monitoring is not yet utilized routinely with mental stress. Accelerometers, magnetometers, ECG, respiratory effort, skin temperature and pulse oximetry were used with 12 health volunteers in this study for monitoring 1) heavy mental load, 2) normal mental load, 3) walking, 4) running and 5) lying. Heavy mental load consisted of a 20-min IQ test and normal mental load was represented by reading a comic book. Automatic feature selection using sequential forward search was used to select the best features for classification of the five activities. Normalized heart rate, utilizing activity context information was found to be the most powerful feature for discriminating heavy mental load from normal. Classification accuracy for all 5 activities was 89% with a custom decision tree and with a k-nearest neighbor classifier and 85% with an artificial neural network.**

I. Introduction

LONG-TERM MONITORING OF HEALTH is essential in chronic conditions such as hypertension [1] and diabetes [2]. Monitoring of blood pressure helps people with hypertension to administer medication better. Monitoring of blood glucose helps people with diabetes to manage their health, e.g., with medication, diet and exercise better. However, the wellbeing of an individual comprises not only physiological, but also psychological and social factors.

Stress and burnout are major public health problems in many industrialized countries. In Finland, in the year 2008 mental disorder was the most common reason for disability pension [3]. OECD estimated that on average one third of disability pensions are caused by mental health problems in 13 OECD countries [4].

Burnout is a psychological consequence of prolonged work stress and it has been reported to co-exist with physical and mental disorders. It has been found to predict disability pensions [5]. Recently, burnout was also found to be related with all-cause mortality in people under 45 years of age [6]. The key work factors affecting psychological ill health are long hours worked, work overload and pressure and the effects of these on personal lives, lack of participation in decision making, poor social support, unclear management and work role [7]. Also, the work environment has also become mentally more burdening, e.g., work intensity and number of complex tasks in work have increased [4]. In Finland, political decisions made to fight these trends are currently to 1) improve employee working capability and 2) support employees to continue longer at work [8].

Stress causes sympathetic responses like smaller heart rate variability (HRV) and higher blood pressure that help in coping with the difficult situation that requires fast reaction or high concentration. The sympathetic responses decrease as the stressful situation passes, e.g., during sleeping or holidays. Prolonged stress and insufficient recovery may lead to burnout and physical illnesses [9].

Currently stress is assessed using questionnaires like the Maslach Burnout Inventory [10]. Automatic monitoring of stress over a long term would help in early identification of stress and in earlier intervention. Automatic monitoring could help people to self-manage their stress better and in stress monitoring setting to reduce the burden caused by manual questionnaires. Comprehensive monitoring of stress requires monitoring of mental and physical load as well as recovery.

Several studies have presented good accuracy detection of physical activities using wearable sensors [11][12][13]. Central methods for analysis have been reviewed in [14],[15]. Studies dealing with automatic detection of mental load using wearable sensors typically report signal feature characteristics during different mental loads [16],[17],[18], but do not yet proceed to automatic classification. Only a few studies show classification rates for identification of different mental loads using wearable sensors [19],[20]. Originating from affective computing and emotion research, somewhat similar studies have used electrodermal activity [21] and electromyogram, electrocardiogram, skin conductivity and respiration signals [22] for recognition of psychological and emotional conditions.

In our earlier studies, we have studied methods for automatic identification of physical activities [12][13][23], mental load and recovery [18] in out-of-lab conditions. In this study, we focus on identification of mental load using data from wearable sensors and show classification accuracies for five activities, including two activities involving mental load. Our hypothesis is that by combining measurements from many wearable sensors, it is possible to automatically recognize, not only physical load, but also mental load. By utilizing the information from automatic activity recognition (activity context) it is possible to improve automatic recognition rates for mental load. Using only, e.g., heart rate (HR) features for classification of mental load is error-prone, because also physical activity causes sympathetic response and, e.g., increases HR.

Manuscript received July 9, 2010. The study was funded in part by Tekes (Finnish Funding Agency for Technology and Innovation), VTT, Nokia, Suunto, European Commission under HeartCycle project (grant IST/FP7-216695) and by Signe and Ane Gyllenberg foundation.

Juha Pärkkä is with Technical Research Centre of Finland, P.O.Box 1300, 33101 Tampere, Finland (phone: +358 02 722 3346, fax: +358 02 722 3380, email: juha.parkka@ vtt.fi).

Miikka Ermes and Mark van Gils are with Technical Research Centre of Finland, P.O.Box 1300, 33101 Tampere, Finland (email: {miikka.ermes; mark.vangils} @vtt.fi)

978-1-4244-6561-3/10/$26.00 ©2010 IEEE

II. METHODS

A. Data Collection

The purpose of the data collection was to use different wearable sensors during mental and physical load and rest and to identify the most useful sensors, feature extraction methods and classification methods for automatic recognition of those activities. The data were collected as part of the study described in [12] and [13]. This study focuses on the analysis of data measured during mental load since that had not been covered in previous analyses.

The data includes (Fig.1) 3D accelerations (Analog Devices ADXL 202E, Fs=20Hz) from hip and wrist, 3D compass data from hip (Honeywell HMC1023, Fs=20Hz), 1-lead ECG (Embla A10, Fs=200Hz), respiratory effort using respiratory inductive plethysmography (RIP) sensors (Embla XactTrace, Fs=200Hz), skin temperature from armpit (YSI 409B, Fs=1Hz) and photoplethysmography (PPG) using a finger pulse oximeter (Embla XN oximeter, Fs=1Hz). Volunteer activities were annotated using an annotation application running on a PDA [12]. Data were stored on a solid-state-memory recorder (Embla A10). The recorder was carried in a rucksack.

Twelve healthy volunteers (10 males, 2 females) were recruited with ads at the local university. Mean volunteer age was 27.1±9.2 yrs, range 19…49 yrs. Mean length was 179.2±6.2 cm, range 167…190 cm. Mean weight was 76.6±7.6 kg, range 60…85 kg. Mean body mass index (BMI) was 23.8±1.9 kg/m^2, range 21.5….26.4 kg/m^2. A written consent was obtained from each volunteer.

Table I describes the activities done during the measurement sessions. The measurement session started with a 3-min rest in lying position (Fig. 2). Heavy mental load was caused with a 20-min IQ-test (intelligent quotient) consisting of completion of geometrical series.

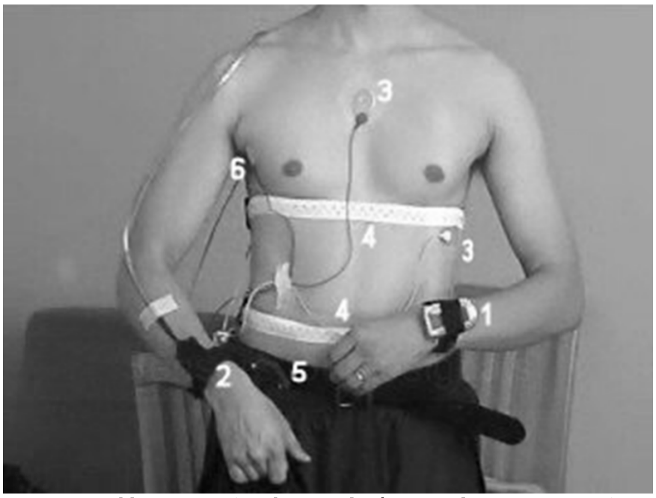


Fig. 1 Wearable sensors: 1) audio recorder for secondary annotation, 2) 3D acceleration sensor box on wrist, 3) two ECG electrodes, 4) two RIP belts, 5) sensor box with 3D accelerometer and 3D magnetometer on hip, 6) skin temperature sensor in armpit.

TABLE I
TARGET ACTIVITIES

Activity	Duration [min]
Indoor activities	
Lying	3
Heavy mental load: sitting and doing IQ test on computer	20
Lying	3
Normal mental load: sitting and reading comics	5
Lying	3
Outdoor activities	
Walking in a park	5
Running in a park	5

Fig. 2 Photos of activities performed by volunteers: a) resting: lying, b) heavy mental load: doing IQ test on computer, c) normal mental load: reading comics, d) physical activities: walking and running in park. In b, oximeter can be seen on index finger.

The IQ-test was performed on a computer in a sitting position. The participants were asked to solve as many of the tasks as possible during the 20-min period to get an IQ score. They were not told the score would not be needed for the study. After the IQ test the volunteers rested in a lying position for 3 minutes. Next, they continued with a task requiring normal mental load: reading comics in a sitting position at a desk. This was followed by a 3-min resting period. Outdoor activities were performed in a park, minimum 5 min of walking and minimum 5 min of running. Each user was given the freedom to exercise at a comfortable pace.

B. Feature extraction and feature selection

During the tasks, data were stored for offline analysis. Feature signals (Fs=1Hz) were computed from the collected raw data using Matlab (Matlab 2009b, Mathworks Inc., Natick, MA, USA). The computed features include time and frequency domain features, e.g., HRV, acceleration (e.g., min, max, mean, variance, peak frequency, peak power, power spectral density (PSD) entropy, estimated energy expenditure), compass bearings (e.g., inclination, declination) and respiratory effort (e.g., amplitude, std, entropy, mean freq., peak req., range). Sequential Forward Search (SFS) was used together with an artificial neural network (ANN) and K-nearest neighbor (KNN) classifiers to find the best features for classification.

For automatic recognition of mental load, HR and HRV features were computed [24]. Four normalized heart rate features were computed: $HRnorm_{ht}$, $HRnorm_{ht\_sit}$, *HRnorm*$_{sit,walk}$, *HRnorm*$_{sit,run.}$ The first two utilize the histogram transformation method [25], that computes HR distribution for a group of people (estimate of whole population) and for the individual and combines these by weighted sum. A cumulative sum is then computed from the combined histogram. The individual HR histogram was weighted with 0.7 and the group HR distribution with 0.3. The cumulative distribution function is used to normalize the individual HR data to range 0…100%. HR data with all HR data in all activities was used for *HRnorm*$_{ht}$. HR distribution of activities done in sitting position was used for *HRnorm*$_{ht\_sit}$. Histogram transformation was applied in leave-one-case-out manner, using data from 11 cases for group data and 1 for individual data.

Activity context was used for HR scaling by computing the median HR during activities. In computation of *HRnorm*$_{sit,walk}$, median HRs during sitting and walking were used for linear scaling. Median HR during sitting received value 0 and median HR during walking the value 100. In computation of the *HRnorm*$_{sit,run.}$, the HR data was similarly scaled using median HR during sitting and running (1).

$$HRnorm_{sit,run} = 100 * \frac{HR - HR_{median,sit}}{HR_{median,run} - HR_{median,sit}} \quad (1)$$

, where HR = stored HR data vector, $HR_{median,sit}$ = median HR during sitting activity, $HR_{median,run}$ = median HR during running activity and $HRnorm_{sit,run}$ = normalized HR vector.

C. Classification

Three classifiers were used for automatic classification. ANN and KNN were used with SFS feature selection. The ANN was a multilayer perceptron network with the resilient propagation learning algorithm and with network size 5:7:5. The KNN classifier was used with five inputs and with parameter k=5. A custom binary decision tree was built for reference. No SFS was used with the custom decision tree. The classification results were computed using leave-one-person-out cross-validation (CV) for KNN and ANN. Custom decision tree is a fixed structure and thus CV was not used with it. Classification was done with 1-s time resolution, thus each second of the data was classified and compared with annotation.

The structure of binary decision tree includes four nodes (Fig 3). Feature signal samples with values larger than the node threshold fall into the right branch, while those smaller than the threshold fall into left branch. The node decisions can be described as 1) footsteps?, 2) lying?, 3) mental stress? and 4) running?. Node 1 utilizes intensity of the highest peak in PSD of vertical hip acceleration computed with 10-sec window to discriminate activities containing regular movements from static activities. Node 2 utilizes mean vertical hip acceleration to discriminate upright positions from positions in horizontal direction. Node 3 discriminates heavy and normal mental load found in sitting position using *HRnorm*$_{sit,run}$. Node 4 utilizes vertical hip acceleration range to discriminate walking from running. Custom decision tree features were selected with visual inspection of feature signals.

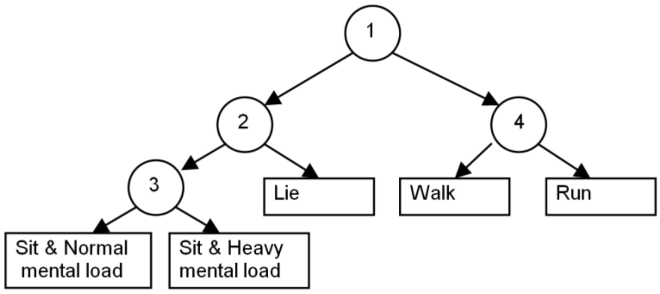


Fig. 3 Custom decision tree structure

III. Results

Data from all 12 cases were used in the analysis. Fig.4 depicts the HR during normal mental load (read) and heavy mental load (IQ) in each case. Table II summarizes HRV features and respiratory effort SD during different activities and shows, which features are significantly different between normal and heavy mental load. Wilcoxon rank sum test was used with $p<0.05$ to test the null hypothesis that the feature distributions of heavy and normal mental load have equal median. Table III shows correct classification accuracies per activity for all three classifiers. Table IV shows the confusion matrix for the custom decision tree.

The features selected by SFS for KNN were: 1) Maximum hip vertical acceleration, 2) HR normalized using sitting and running heart rates, 3) standard deviation of respiratory effort signal, 4) minimum hip vertical acceleration and 5) declination angle of hip magnetometer. The features selected for ANN were: 1) maximum hip horizontal acceleration, 2) estimate of energy expenditure using all 3 dimensions of hip accelerations, 3) minimum hip vertical acceleration, 4) HR normalized using histogram transformation and all sitting HR data, 5) peak power of horizontal hip accelerations computed with 10-sec window.

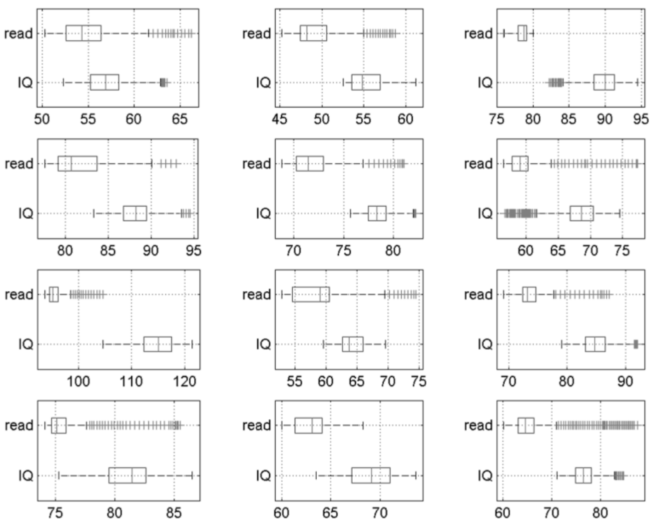


Fig. 4 HR distribution during normal (read) and heavy (IQ) mental loads in each case.

TABLE II

MEAN (AND STD) OF HRV AND RESPIRATOTY EFFORT FEATURES DURING DIFFERENT ACTIVITIES (* INDICATES SIGNIFICANT DIFFERENCE BETWEEN NORMAL AND HEAVY MENTAL LOAD WITH P<0.05 AND USING WILCOXON RANK SUM TEST)

Feature	Lie	Sit & norm load	Sit & heavy load	Walk	Run
Resp. Effort SD	1.4 (0.6)	1.3 * (0.3)	0.7 * (0.2)	2.1 (0.3)	16.3 (5.8)
HR	67.5 (2.3)	68.4 (1.5)	77.3 (1.5)	102.7 (3.9)	160.8 (4.7)
SD2	100.0 (24.7)	100.5 (25.0)	72.7 (17.6)	57.8 (18.8)	18.5 (16.9)
SD1	46.1 (7. 6)	41.9 (7.0)	27.3 (5.2)	30.5 (8.7)	11.0 (9.4)
SD1/SD2	0.4 (0.1)	0.4 (0.1)	0.38 (0.1)	0.4 (0.1)	0.6 (0.3)
RMSSD	65.2 (10.7)	59.2 (10.0)	38.7 (7.4)	43.1 (12.3)	15.6 (13.3)
HR SD	78.6 (17.2)	77.7 (17.2)	55.1 (12.4)	47.9 (14.4)	15.6 (13.7)
$HRnorm_{sit,walk}$	-8.1 (5.9)	-1.3 * (4.1)	24.5 * (4.2)	84.6 (10.4)	254.1 (15.1)
$HRnorm_{sit,run}$	-1.5 (2.6)	-0.5 * (1.7)	10.0 * (1.6)	37.2 (4.4)	101.6 (5.3)
$HRnorm_{ht\_sit}$	19.0 (6.8)	16.6 * (4.3)	51.8 * (12.5)	90.3 (4.4)	100.0 (0.0)
$HRnorm_{ht}$	12.3 (3.4)	12.1 * (2.1)	27.4 * (3.0)	68.8 (5.8)	98.5 (0.7)

TABLE III

CLASSIFIER RESULTS [%]

	Custom Decision Tree	K-nearest neighbor	Artificial Neural Network
Lie	**98**	**98**	91
Normal mental	78	68	**94**
Heavy mental	**93**	84	59
Walk	84	**95**	85
Run	91	**100**	99
TOTAL	**89**	**89**	85

TABLE IV

CONFUSION MATRIX OF CUSTOM BINARY DECISION TREE [%]

Annotation	Recognized Activity				
	Lie	Medium	Heavy	Walk	Run
Lie	**98**	1	1	0	0
Normal mental	0	**78**	22	0	0
Heavy mental	0	7	**93**	0	0
Walk	3	0	13	**84**	0
Run	0	0	0	9	**91**

The inter-individual variability in HR was measured by computing means (and SD) of HR data during activities: lying 67.5 (14.0) beats per minute (bpm), normal mental load 68.4 (13.3) bpm, heavy mental load 77.3 (16.5) bpm, walking 102.7 (21.5) bpm and running 160.8 (10.0) bpm. These figures were computed by first computing the individual mean HR during each activity and then computing the mean and SD over all cases.

IV. DISCUSSION

Activities were classified using wearable sensor data into both mental and physical activities. The results show good accuracy for the classification of both physical and mental activities. For identification of physical activities, accelerometer data and features computed from them are the most useful ones. For identification of mental load, accelerometer, HR and respiratory effort features are useful.

When looking at the HR data during normal and heavy mental loads, an elevated HR can be found in all 12 cases during heavy mental load (Fig.4, Table II). Even though the result is as expected, the consistency of the difference in all cases is remarkable. This suggests that automatic monitoring of mental load is feasible. Table II shows how well HRV and respiratory effort features can be used for classification. The normalized HRs and respiratory effort SD are the best indicators for differentiating heavy and normal mental load. None of the not-normalized HR features are significantly different between heavy and normal mental loads.

In order to efficiently utilize the HR data in classification, we should get rid of the great inter-individual variability. Two methods: histogram transformation and scaling with information from activity classification were utilized in this study. As can be seen from the list of the features selected by SFS with KNN and ANN, HR normalization with data from sitting and running activities was selected for KNN and HR normalization with histogram normalization over sitting activities was selected for ANN. In case of custom decision tree, the thresholding was most successful with HR normalized using data from sit and run activities.

In addition to features computed from accelerometer and IR data, the standard deviation of respiratory effort signal vas selected for KNN classifier. This feature has in most ases the smallest values during heavy mental load and argest values during running (Table II). Even normal nental load and lying have larger respiratory effort SD han what heavy mental load has. This is probably due to he fact that during the IQ test (as often during heavy nental load), people concentrate intensively and do not speak or move. In general, speaking and movement hange the dynamics of respiratory effort signal. For KNN, SFS selects also the hip declination feature omputed using the magnetometer data. This feature tells he angle between horizontal axis parallel to direction of he volunteer’s body and the direction of the magnetic ield. Thus, in this case it may tell differences in sitting osition (at a desk / leaning back).

Also PPG signal was measured with a finger oximeter, but quality of that signal was not good enough for analysis, because of clipping. Similarly, skin temperature was measured, but although the sensor was placed in volunteer armpit, it reflects mostly temperature of the surroundings (e.g., indoor vs. outdoor).

SFS feature selection allows rapid selection of features that are potentially useful in classification. However, the resulting feature set should always be checked against domain knowledge to allow a sensible set of final features.

Tables III and IV show the classification accuracies for each classifier and the confusion matrix for custom decision tree. The heavy mental load can be detected using a custom decision tree with 93% accuracy and the normal mental load with 78% accuracy. Lying, walking and running can be detected with 98%, 84% and 91% accuracy, respectively. The best correct classification accuracies are obtained with custom decision tree and KNN. Custom decision tree is computationally very efficient, while KNN is very demanding. ANN correct classification accuracy is 85%.

In order to use this method in real life, baseline HR without stress in sitting position should be available for normalization of the individual HR distribution.

V. Conclusion

Both mental and physical load and activities were automatically recognized from wearable sensor data with good accuracy. Activity context can be used for normalizing the individual HR data and for improving the correct recognition rate of mental load.

References

[1] W.J. Verberk, A.A. Kroon, J.W.M. Lenders, A.G.H. Kessels, G.A. van Montfrans, A.J. Smit, P.-H.M. van der Kuy, P.J. Nelemans, R.J.M.W. Rennenberg, D.E. Grobbee, F.W. Beltman, M.A. Joore, D.E.M. Brunenberg, C. Dirksen, T. Thien, P.W. de Leeuw, "Self-measurement of blood pressure at Home Reduces the Need for Antihypertensive Drugs – A Randomized, Controlled Trial", Hypertension, vol. 50, pp. 1019-1025, 2007.

[2] J.M. Evans, R.W. Newton, D.A. Ruta, T.M. MacDonald, R.J. Stevenson, A.D. Morris, "Frequency of blood glucose monitoring in relation to glycaemic control: observational study with diabetes database", *BMJ*, 319(7202):83-6, 1999.

[3] M. Hiltunen, K. Käkönen, J. Kannisto, M. Pellinen, K. Lybäck, R. Goebel, *Pensioners and insured in Finland 2008,* Joint publication of Finnish Centre for Pensions, The Local Government Pensions Institution and State Treasury. Available: http://www.etk.fi/Binary.aspx?Section=42845&Item=64651

[4] OECD, *Sickness, Disability and Work: Keeping on Track in the Economic Downturn – Background Paper*, OECD, 2009. Available: http://www.oecd.org/dataoecd/42/15/42699911.pdf

[5] K. Ahola, R. Gould, M. Virtanen, T. Honkonen, A. Aromaa, J. Lönnqvist, "Occupational burnout as a predictor of disability pension: A population-based cohort study", *Occupational and Environmental Medicine*, 66(5), 284-290, 2009.

[6] K. Ahola, A. Väänänen, A. Koskinen, A. Kouvonen, A. Shirom, "Burnout as a predictor of all-cause mortality among industrial employees: A 10-year prospective register-linkage study", *Journal of Psychosomatic Research*, vol. 69, pp 51-57, 2010.

[7] S. Mitchie, S. Williams, "Reducing work related psychological ill health and sickness absence: a systematic literature review", *Occupational and Environmental Medicine*, vol,60, pp. 3-9, 2003.

[8] R. Gould, H. Nyman, H. Lampi, "Masennukseen perustuvat työkyvyttömyyseläkkeet meillä ja muualla" in H. Uusitalo, M. Kautto, C. Lindell (eds), *Myöhemmin eläkkeelle – selvityksiä ja laskelmia*, (in Finnish), publication of Finnish Centre for Pensions, 2010. Available: http://www.etk.fi/Binary.aspx?Section=64145&Item=64589

[9] T. Honkonen, K. Ahola, M. Pertovaara, E. Isometsä, R. Kalimo, E. Nykyri, A. Aromaa, J. Lönnqvist, "The association between burnout and physical illness in the general population – results from the Finnish Health 2000 Study", Journal of Psychosomatic Research, vol 61, pp 59-66, 2006.

[10] C. Maslach, S.E. Jackson, "The measurement of experienced burnout", Journal of Occupational Behaviour, vol. 2, pp. 99-113, 1981.

[11] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in Proc. 2nd Int. Conf. Pervasive Computing, 2004, pp. 1–17.

[12] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, I. Korhonen, "Activity recognition using realistic data from wearable sensors", *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, January 2006.

[13] M. Ermes, J. Pärkkä, J. Mäntyjärvi, I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, no: 1, 20 – 26, 2008.

[14] S.J. Preece, J.Y. Goulermas, L.P.J. Kenney, D. Howard, K. Meijer, R. Crompton, "Activity identification using body-mounted sensors – a review of classification techniques", Physiol. Meas., vol 30, pp. R1-R33, 2009.

[15] M. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement," Physiol. Measure., vol. 25, no. 2, pp. R1–R20, Apr. 2004.

[16] M. Kusserow, O. Amft, G. Tröster, "Analysis of Heart Stress Response for a Public Talk Assistant System" in E. Aarts et al. (Eds.): Conf. Proc. of AmI 2008, LNCS 5355, pp. 326–342, 2008.

[17] J. Taelman, S. Vandeput, A. Spaepen, S. van Huffel, "Influence of Mental Stress on Heart Rate and Heart Rate Variability", in IFMBE Proc. of 4th ECIFMBE, Antwerp, Belgium, 23–27 Nov, 2008.

[18] J. Pärkkä, J. Merilahti, E.M. Mattila, E. Malm, K. Antila, M. Tuomisto, A. Saarinen, M. van Gils, I. Korhonen, "Relationship of Psychological and Physiological Variables in Long-term Self-monitored Data during Work Ability Rehabilitation Program", IEEE Transactions on Information Technology in Biomedicine, Vol. 13, No: 2, 141 – 151, 2009.

[19] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, U. Ehlert, "Discriminating Stress From Cognitive Load Using a Wearable EDA Device", *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp 410-417, 2010.

[20] B. Arnrich, C. Setz, R. La Marca, G. Tröster, U. Ehlert, "What does your chair know about your stress level?", *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp 207-214, 2010.

[21] M.-Z. Poh, N.C. Swenson, R.W. Picard, "A wearable sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity", *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1243-52, 2010.

[22] J. Kim, E. André, "Emotion Recognition Based on Physiological Changes in Music Listening", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067-83, 2008.

[23] J. Pärkkä, M. Ermes, K. Antila, M. van Gils, A. Mänttäri, H. Nieminen, "Estimating intensity of physical activity: a comparison of wearable accelerometer and gyro sensors and 3 sensor locations", *Conf Proc of 29th IEEE EMBC*, Lyon, France, 23-26 August 2007.

[24] Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology, "Heart rate variability - Standards of measurement, physiological interpretation, and clinical use", *European Heart Journal*, vol. 17, 354-381, 1996.

[25] M. Huiku, K. Uutela, M. van Gils, I. Korhonen, M. Kymäläinen, P. Meriläinen, M. Paloheimo, M. Rantanen, P. Takala, H. Viertiö-Oja and A. Yli-Hankala, "Assessment of surgical stress during general anaesthesia", *British Journal of Anaesthesia*, vol. 98, pp. 447-55, 2007.

PUBLICATION **P6**

Relationship of Psychological and Physiological Variables in Long-term Self-monitored Data during Work Ability Rehabilitation Program

In: IEEE Transactions on Information Technology in Biomedicine 2009. Vol. 13, No. 2, pp. 141–151.
Reprinted with permission from the publisher.

Relationship of Psychological and Physiological Variables in Long-Term Self-Monitored Data During Work Ability Rehabilitation Program

Juha Pärkkä, *Member, IEEE*, Juho Merilahti, Elina M. Mattila, Esko Malm, Kari Antila, Martti T. Tuomisto, Ari Viljam Saarinen, Mark van Gils, *Member, IEEE*, and Ilkka Korhonen, *Member, IEEE*

***Abstract*—Individual wellness comprises both psychological and physiological wellbeing, which are interrelated. In long-term monitoring of wellness, both components should be included. Work-related stress and burnout are persistent problems in industrial countries. Early identification of work-related stress symptoms and early intervention could reduce individual suffering and improve the working productivity and creativity. The goal of this study was to explore the relationship between physiological and psychological variables measured at home by the users themselves or automatically. In all, 17 (3 males and 14 females, age 40–62) people participating in a work ability rehabilitation program (due to work overload) were monitored for three months. Physiological and behavioral variables (activity, bed occupancy, heart rate (HR) and respiration during night, HR during day, blood pressure, steps, weight, room illumination, and temperature) were measured with different unobtrusive wireless sensors. Daily self-assessment of stress, mood, and behaviors (exercise, sleep) were collected using a mobile phone diary. The daily self-assessment of stress and the Derogatis stress profile questionnaire were used as reference for stress status. Results show modest, but significant pooled overall correlations between self-assessed stress level, and physiological and behavioral variables (e.g., sleep length measured with wrist-worn activity monitor: $\rho = -0.22$, $p < 0.001$, and variance of nightly bedroom illumination: $\rho = 0.13$, $p < 0.001$). Strong, but sometimes conflicting correlations can be found at individual level, suggesting individual reactions to stress in daily life.**

***Index Terms*—Actigraph, behavior, heart rate (HR), psychological and physiological variables, sleep, stress, wellness monitoring.**

Manuscript received January 25, 2008. First published October 31, 2008; current version published March 3, 2009. This work was supported by Tekes and Valtion Teknillinen Tutkimuskeskus (VTT) under Project wSense—Long term wellness monitoring by wireless low power sensors at real life settings.

J. Pärkkä, J. Merilahti, E. M. Mattila, I. Korhonen, and M. van Gils are with Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, 33101 Tampere, Finland (e-mail: juha.parkka@vtt.fi; juho.merilahti@vtt.fi; elina.m.mattila@vtt.fi; ilkka.korhonen@vtt.fi; mark.vangils@vtt.fi).

E. Malm was with Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, 33101 Tampere, Finland. He is now with Finwe Ltd., 90570 Oulu, Finland (e-mail: esko.malm@finwe.fi).

K. Antila was with Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, 33101 Tampere, Finland. He is now with the Laboratory of Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115 USA (e-mail: kari.antila@vtt.fi).

M. T. Tuomisto is with the Institute for Extension Studies and Department of Psychology, University of Tampere, 33014 Tampere, Finland (e-mail: klmatu@uta.fi).

A. V. Saarinen was with Rokuan Kuntokeskus, 91670 Rokua, Finland. He is now with Kainuu Central Hospital, 87140 Kajaani, Finland (e-mail: ari.saarinen@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITB.2008.2007078

I. Introduction

Long-term monitoring of health and wellness as a part of everyday life is seen as a central element of both chronic disease management and health or wellness management [1], [2]. Monitoring of physiological variables such as heart rate (HR), blood pressure, or blood glucose has been widely applied for this purpose. However, the wellbeing of the individual includes physiological, psychological, and social factors, all of which are interacting as determinants of health. In fact, behavioral and social factors explain more than 50% of health outcomes [3]. In addition, interpretation of variations in physiological variables measured in uncontrolled conditions, such as at home, requires knowledge of the other contributing factors. Hence, there is a need for comprehensive health monitoring approaches including both psychophysiological and behavioral components. However, only relatively few studies have been published that deal with mutual relationships between these variables in long-term real-life settings [4]–[6]. Some equipment has been developed that is suitable for such studies [7], [8].

An important psychophysiological phenomenon is stress, as work-related stress and burnout are major public health problems. In Finland, about 7% of employees suffer from severe work-related burnout [9]. Another study concludes that 2.5% of employees in Finland suffer from severe burnout and about 24% from mild burnout [10]. The prevalence of severe burnout among employees in Finland, Sweden (7.4%), and the Netherlands (4%–7%) probably gives a good indication of the situation in other industrialized countries also [11]. Recently, the World Health Organization (WHO) started a promotion program on mental health [12]. WHO states that historically, mental health has often been misunderstood and access to mental health care has been too difficult. With the promotion program, WHO seeks to highlight the importance of mental health and attempts to lower barriers preventing access to mental healthcare.

Stress causes sympathetic responses (such as higher cortisol level, smaller HR variability (HRV), and higher blood pressure) that first help us tackling situations requiring fast reaction or high concentration. This is a healthy and normal reaction, which disappears as the stressful situation passes by, e.g., during sleep and holidays. For stress management, a central issue is the successful recovery after a stressful situation [13]. If this recovery fails, an allostatic load is cumulated, which may lead to health consequences such as burnout [13]. Burnout has been defined to comprise emotional exhaustion that does not disappear during free time, depersonalization or cynicism, and reduced sense of personal accomplishment [14]. Prolonged stress has been suspected to cause physical illnesses [15], and even potentially transfer to other people [16].

Traditionally, stress, work exhaustion, and burnout have been assessed using questionnaires, e.g., Bergen burnout indicator (BBI) [17], Derogatis stress profile (DSP) [18], [19], and Maslach burnout inventory [14]. Automatic and more comprehensive methods for monitoring stress status at home over long term could help to obtain early identification of stress, and may lead to earlier intervention when necessary. In such approaches, physiological, psychological, and behavioral monitoring could be applied. Several physiological variables are known to be related to sympathetic stress reactions, e.g., HRV and blood pressure [20]. In addition, behavioral patterns such as daily activity patterns and sleep patterns may be relevant [21], [22]. Self-reported stress and other behavioral events have been recommended to be included together with physiological variables when the stress state of a participant is to be assessed [23]. However, little is known about the mutual correlations of these variables in long-term settings or their relationship to changes in stress status.

Our objective was to study how different physiological and behavioral variables measured in long term at home in uncontrolled conditions by participants themselves, or automatically by wireless sensors, are related to psychological self-assessments or data acquired by standard validated questionnaires (BBI and DSP). In addition to the identification of the significant correlations between the variables, we aimed to find the best single variables that could be used in long-term psychophysiological wellness monitoring. We recruited 17 participants from a rehabilitation program, and targeted to improve the working ability. Participants of this program commonly report increased levels of work exhaustion and long-term stress, among other health and work-related complaints. The participants were monitored for three months both at home and during the rehabilitation period in the rehabilitation institute.

The methods used are described in Section II. Results of the study are presented in Section III. Section IV discusses the results. Conclusions of the study are presented in Section V.

II. Methods

A. Data Collection

The goal of data collection was to study whether the selected home monitoring devices provide information that as such, or after processing, reflect the participant's stress level. This study was conducted in real-life settings. The data collection equipment consisted of ten hardware units: 1) activity monitor (actigraph) worn on the wrist; 2) HR monitor; 3) mobile phone; 4) step counter; 5) blood pressure monitor; 6) personal weight scale; 7) movement sensor in bed; 8) wireless sensor node with temperature and illumination sensors; 9) laptop PC; and 10) central server (Fig. 1). Data from activity monitor worn on the wrist, movement sensor in bed, HR monitor, and wireless sensor node with temperature and illumination sensors were sent via the laptop to central server. Data from blood pressure monitor, step counter, personal weight scale, and daily self-assessments (e.g., stress, sleep quality) were sent via Wellness Diary mobile phone application to the central server (Fig. 1).

Fig. 1. Data collection equipment (details in text): (1) IST wrist activity monitor (actigraph). (2) Suunto HR monitor. (3) Nokia mobile phone with Wellness Diary application for self-assessments and measurement results from (4) Omron step counter, (5) Omron blood pressure monitor, and (6) weight scale, (7) Emfit bed sensor, (8) environmental sensors, (9) laptop PC, and (10) central server. Collected HR data were processed with Firstbeat PRO WAS.

The laptop (HP nc6000, Hewlett-Packard Company, Palo Alto, CA) was used as a data storage unit, to which the data were transferred from devices that acquired data continuously. A computer program was written especially for this study to automatically send all collected data from the laptop to a central server once per week using a modem connection.

A mobile phone (Nokia 6670, Nokia Corporation, Helsinki, Finland) was used as a second data storage unit. The user manually entered measurement results (e.g., blood pressure) and self-assessments (e.g., stress level) into the Wellness Diary program [24], [25] installed on the mobile phone. The Wellness Diary is a stand-alone program that was not connected to the laptop. Instead, the user could send the collected data as a multimedia messaging service (MMS) message to a central server. The user was instructed to do so once per week. The Wellness Diary application is a tool for self-assessments, which is a basic concept of cognitive behavioral therapy. It relies on the user entering the inputs to make the user aware of changes in his health status, to subsequently learn what has positive effect on health, and as a consequence start changing his behavior. This is why all measurement results were not transferred automatically to the Wellness Diary application. The Wellness Diary used in this study was a modified version of the one available on the Internet [24].

The activity monitor (actigraph) [26] worn on the wrist (IST Vivago WristCare, IST International Security Technology Oy, Helsinki, Finland) continuously measured the user's activity as reflected by hand movements. The device was attached to the wrist of the nondominant hand. Its data were transferred automatically, via IST wireless base station, to the laptop. The range of the wireless connection is 20–30 m, which covers a typical apartment.

The movement sensor in the bed (Emfit SafeBed, Emfit Ltd., Vaajakoski, Finland) was a thin-film ferroelectret sensor, which was placed below the mattress and which measures pressure changes. The sensor measured continuously and the data were transferred automatically to the laptop via a wireless sensor node (SoapBox [27]).

The HR monitor (Suunto T6 wrist-top computer, Suunto Oy, Vantaa, Finland) was used to make one-day, beat-to-beat HR recordings three days a week. The user started the normal one-day-measurement in the morning and ended it in the evening by stopping the measurement and downloading the data to the laptop. HR data were analyzed using Firstbeat PRO Wellness Analysis Software (WAS) (Firstbeat Technologies Oy, Jyväskylä, Finland).

The blood pressure monitor (Omron 705IT, Omron, Kyoto, Japan) was used to measure blood pressure every morning and evening. Users manually entered the blood pressure readings to the Wellness Diary.

The step counter (Omron Walking Style II, Omron, Kyoto, Japan) measured the number of steps taken during a day, and reflecting "lifestyle activities." The user read the step count from the step counter display in the evening and entered the step count into the Wellness Diary.

Bedroom temperature and illumination levels were continuously measured using the SoapBox sensor platform [27], which also wirelessly sends data to the laptop. The SoapBox unit was placed close to bedroom reading lamp for easy detection of lights-on during night. Data collection and transfer were automatic.

A personal weight scale was used to measure the user's weight every morning before breakfast. The users' own scales were used; thus, many different models were used. The user manually entered the weight to the Wellness Diary.

Additionally, each participant filled in a BBI-15 questionnaire once, a DSP questionnaire four times, and used the "day-type paper form" to write down the day type (work, free, sick, rehabilitation) for each day of the measurement period. Fig. 2 depicts the study protocol.

Participants were recruited from vocational rehabilitation groups (rehabilitation funded by KELA, The Social Insurance Institution of Finland). The participants had lowered working ability (due to mental stress), and their application to the rehabilitation had been approved by KELA. The time from application to the actual rehabilitation took several months. The study was conducted in two parts, first eight people participated in the measurements simultaneously in May–August 2005 and nine more people took part in September–December 2005. The 17 participants included 14 females and 3 males. Average age was 54.5 years (standard deviation (SD) 5.4 years, range 40–62). The participants were white-collar workers, representing mostly university employees and health care professionals. The average BBI was 49.2 (SD 12, range 26–72). Nine participants had increased values of DSP (total stress score $\geq$50) during the study. The study has been approved by local ethics committee (hospital district of North-Ostrobothnia, Finland).

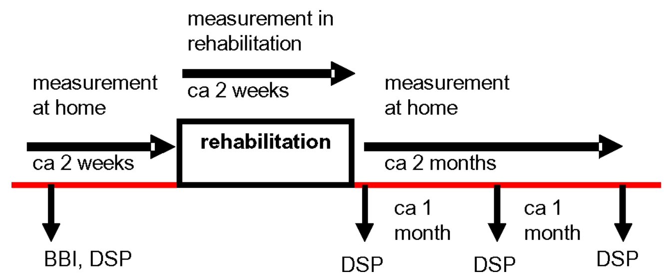


Fig. 2. Study protocol describing both measurements and questionnaires (Bergen burnout indicator—BBI, Derogatis stress profile—DSP).

Fig. 3. Wellness Diary stress assessment form (in Finnish, "Kiire/stressi/paineet" = busyness/stress/pressure and "Väsymys" = tiredness).

The study protocol contained a two-week rehabilitation in a rehabilitation center. The participants used the self-monitoring equipment for two weeks before the rehabilitation, during the two-week rehabilitation, and for two months after the rehabilitation (Fig. 2). In the beginning of the study, the participants filled in the BBI, and as the devices were installed into their home, they filled in the DSP questionnaire for the first time. The DSP was filled in the second time after the rehabilitation, the third time one month after that, and the fourth time at the end of the measurement period. The DSP total stress score (DSPtss) was used as the reference of stress level in this study.

In addition to DSP, stress assessments were done daily by means of Wellness Diary. The stress form (Fig. 3) in the Wellness Diary software consisted of four assessments: 1) busyness/stress/pressure; 2) tension/anxiety/fear; 3) tiredness; and 4) trouble/irritation/anger. The participant was allowed to answer as many of the four assessments as she/he felt relevant. The stress level was assessed in the evening. The participants were instructed to do the assessment so that the given value covers the feelings of the whole day. The assessment was done using a slider with scale from 0 to 10 and -1 for "no assessment." In addition to DSPtss, the daily stress level

TABLE I
DAILY MEASUREMENT ROUTINES (WD = WELLNESS DIARY, HR = HEART RATE)

Morning	Actions needed
Weight	Measure and fill weight form in WD
Blood pressure	Measure and fill blood pressure form in WD
Sleep	Estimate length and quality of previous night's sleep and fill in sleep form in WD
Steps	Place step counter in pocket
Heart rate	Start HR measurement. Measure HR 3 days a week (2 on work days, 1 on weekend)
Evening	**Actions before going to bed**
Heart rate	Stop measurement and transfer data to laptop
Blood pressure	Measure and fill blood pressure form in WD
Stress	Assess day's stress level and fill stress form in WD
Steps	Check day's step count on step counter and fill steps form in WD
Others	
Exercise	Fill in exercise form in WD for each sports activity
Wellness Diary	Send measurement results from mobile phone to research server once a week

"busyness/stress/pressure" assessed using the Wellness Diary was used as a second reference of stress level. The daily measurement and assessment routines are summarized in Table I. Other measurements were done automatically and did not require user interaction.

B. Feature Signals

After data collection, feature signals were computed from the raw data. The generated feature signals have a sampling rate of one per day. A brief description of each feature is given in Table II.

Wellness Diary features are raw data as entered into the application by the user. Outliers are removed by comparing the values to known acceptable ranges. For blood pressure measurements, morning was defined as 5–12 O'clock (5–12 AM), and evening as 14–24 O'clock (2–12 PM). Feature "*WD weight change*" was defined as weight of next morning minus weight on the current day morning.

The bed sensor features were computed from presence and HR data that the device gives as output once per minute. The presence signal describes how many seconds the bed has been occupied during the past 60 s. HR is an average over the past minute. A filtered presence signal was obtained by low-pass filtering the presence signal using a filter length of 10 min. The low-pass filtering removes short-duration absences from the bed during night. Filtered values greater than 30 s were taken into account, and the signal was turned into a Boolean signal describing presence (1) and absence (0) from bed. "*Bed time start*" and "*Bed time end*" times are searched from the filtered signal, and defined as the start and end of the longest continuous bed presence block during the night. "*Bed wakeups*" is computed from the unfiltered bed occupancy signal. A Boolean presence signal is generated by giving 1 for values 60 s and 0 for others. Number of wakeups is computed as the number of breaks during night (number of zero blocks between sleep start and sleep end times). "*Bed avg sleep 3 nights*" is computed as the average sleep length over 3 consecutive nights. Night sleep time is computed from the filtered signal. "*Bed 20 to 10 bed time*" is a direct sum of bed presence between 20 O'clock in the evening and 10 O'clock in the morning, computed from the unfiltered signal. Similarly "*Bed 10 to 20 bed time*" is a direct sum of bed presence during day time. "*Bed avg night HR*" is the average HR over the self-reported sleep time. HR signal is available directly from the bed sensor.

TABLE II
FEATURE SIGNALS (FS = 1 PER DAY)

Feature	Description
WD busyness/stress/pressure	Self assessment of "busyness/stress/pressure" from WD
WDTrouble/irritation/anger	Trouble/irritation/anger self assessment
WD Tiredness	Tiredness self assessment
WD Tension/anxiety/fear	Tension/anxiety/fear self assessment
WD BP sys & dia morning	Systolic & diast. blood pressure morning
WD BP sys & dia evening	Systolic & diast. blood pressure evening
WD sleep quality	Sleep quality self-assessment (0…10)
WD sleep length	Sleep length self-assessment
WD steps	Step count from step counter
WD exercise	Duration of the day's exercise sessions
WD weight	Weight
WD weight change	Weight difference from this to next day
WD NumExerciseEntries	Number of exercise entries to WD
WD NumWeightEntries	Number of weight entries to WD
Scored sleep length	Sleep length as scored by medical doctor
Scored sleep quality	Sleep quality as scored by medical doctor
Bed time start	Bed time start offset from 22:00 (10 pm) (in minutes). Positive = later than 22, negative = earlier than 22.
Bed time end	Bed time end offset from 06:00 (6 am) (in minutes). Positive = later than 06, negative = earlier than 06.
Bed wakeups	Number of wakeups per night
Bed avg sleep 3 nights	Average sleep length of 3 consecutive nights.
Bed 20 to 10 bed time	Time in bed between 20:00 and 10:00 (night)
Bed 10 to 20 bed time	Time in bed between 10:00 and 20:00 (day)
Bed avg night HR	Average heart rate during night, measured with bed sensor
Wrist night activity	Wrist activity monitor: night activity
Wrist night activity SD	Wrist activity monitor: night activity SD
Wrist day activity	Wrist activity monitor: day activity
Wrist day activity SD	Wrist activity monitor: day activity SD
Wrist act ratio N to prev D	Wrist activity monitor: ratio of night activity per previous day activity
Wrist act ratio N to next D	Wrist activity monitor: ratio of night activity per next day activity
Wrist sleep length	Wrist activity monitor: sleep length
Wrist num sleep periods	Wrist activity monitor: number of sleep periods
Illumination avg, median, var N	Average, median and variance of night illumination
HR relaxation time	Relaxation duration computed using Firstbeat PRO WAS
HR stress time	Stress duration computed using Firstbeat PRO WAS
HR sport time	Sport duration computed using Firstbeat PRO WAS
HR min HR	Minimum heart rate from HRM (day)
HR average HR	Average heart rate form HRM (day)

WD = wellness diary, HRM = heart rate monitor.

Activity monitor (actigraph) features were computed from the wrist activity monitor data. The activity monitor signal is an activity count that is high when there are a lot of hand movements and low when the hand is not moving. The activity monitor detects sleep automatically based on low activity, and gives as one output signal a Boolean vector, which indicates sleep [26]. Low activity can occur during day as well as during night; thus, time limits are required in order to reliably detect the sleeping time during night. By default, the wrist activity monitor used in the study detects sleep during the night 23–07 O'clock (11 PM–07 AM). These limits are designed for elderly users. In this study, we had working age users, who sleep more irregularly. In order to analyze the effect of stress on night activity and day activity, we took the scored sleep onset and wakeup times, and used these to define night. Day was defined to begin 1 h after wakeup and night 1 h before sleep onset. The feature "*Wrist sleep length*" represents the sleep length as detected by the activity monitor during the scored night. Feature "*Wrist number of sleep periods*" represents the number of sleep periods the activity monitor algorithm detects during the scored night. "*Wrist act ratio N to next D*" represents the average activity ratio of previous night and current day. This feature is high if night activity is high or day activity is low. Similarly, the feature "*Wrist act ratio N to prev D*" represents the activity ratio of previous night and previous day. The "*Wrist night activity*" feature represents the average activity during the scored night. "*Wrist night activity SD*" represents the SD of activity during the scored night. Similarly, features "*Wrist day activity*" and "*Wrist day activity SD*" represent those of the day time.

Illumination data were first preprocessed (clear outliers were replaced with NaNs). Then average, median, and variance were computed for the night time data.

HRV features were computed from ambulatory beat-to-beat heartbeat recordings. Analysis was carried out by using the Firstbeat PRO WAS. It computes stress and relaxation states of the autonomic nervous system by using HR and HRV signals, and indexes derived from these. The software segments the recording into stationary segments of sports, stress, and relaxation. "*Stress time*" feature gives the time the body is in a stress state, and the sympathetic nervous system activity is dominating over parasympathetic activation. "*Relaxation time*" feature gives the time the body is in a relaxed state and parasympathetic activation is dominating.

Features "*Scored sleep length*" and "*Scored sleep quality*" are the sleep length and sleep quality as assessed by a medical doctor by examining in one plot the activity monitor signal, bed presence signal, and the WD self-reported sleep onset and wakeup times [28].

An example (Fig. 4) of one case shows changes in selected variables. Self-reported stress is higher in the beginning as the participant is working. The daily stress level falls during rehabilitation, but DSPtss increases even despite the summer holiday. Diastolic morning blood pressure decreases during summer holiday (from 80 to 70 mmHg). Wrist sleep length feature varies more during work and rehabilitation than during summer holiday. Bed time between 20 and 10 O'clock (during night) is in many nights 1–2 h longer than detected sleep length. This could indicate sleeping problems, although this person sleeps normally 6–9 h per night. HR stress time is detected from daytime HR recordings. Stress time per day is shortest during rehabilitation. Elsewhere, there seems to be 200–700 min of stress daily, as detected using HR recordings. Walking (or running) has been done a lot during rehabilitation. Step counts exceed 15 000 steps during a couple of days in the rehabilitation period.

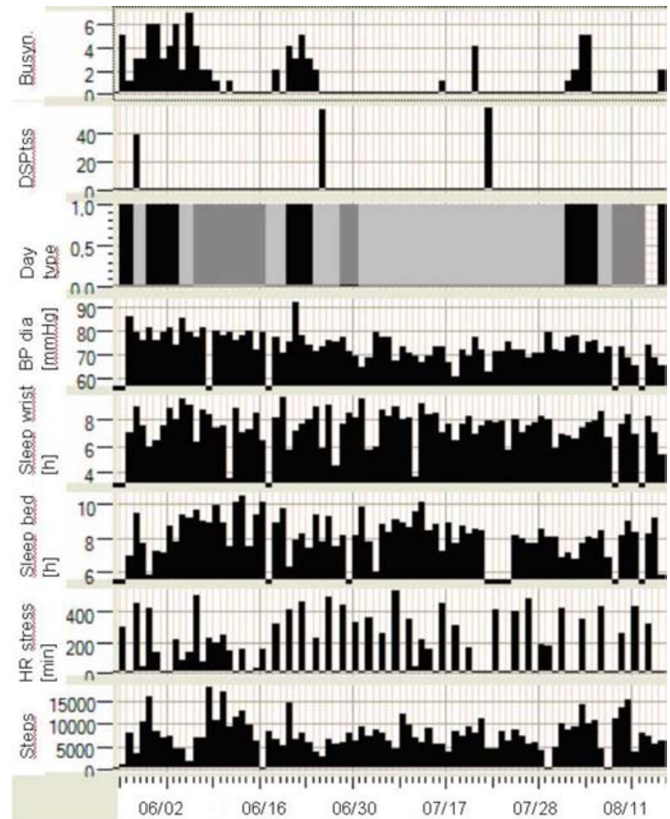


Fig. 4. Example of collected data from one case. Selected features from top to bottom. (1) Busyness/stress/pressure [0–10]. (2) DSPtss. (3) Workday (black), rehabday (dark gray), freeday (light gray). (4) WD BP diastolic morning [in millimeter of mercury]. (5) Wrist sleep length [in hours]. (6) Bed time 20–10 [in hours]. (7) HR stress time [in minutes]. (8) WD steps. Horizontal axis represents calendar time [month/day].

III. RESULTS

A. Self-Assessed Stress and Psychophysiological Variables

Spearman correlations were computed between assessed busyness/stress/pressure and the measured variables. Only workdays were included in this correlation calculation, because the focus was on work-related stress, not on holiday stress. The personal average of the variable value was subtracted first before pooling the data together. Only case-wise significant ($p < 0.05$) correlations are shown (Table III). Mathematical computations were done using Matlab (The Mathworks, Inc., Natick, MA).

The psychophysiological variables that have the most significant correlation with self-assessed stress level (from Wellness Diary) also show a visually apparent change when ordered by stress level (Fig. 5). The median of the ordered feature signals over 15 days was computed to highlight level change in feature signals as function of stress level. Sleep length, both self-assessed and wrist-activity-monitor-detected, decreases by ca. 1 h when comparing the least stressful (on the left) and the most stressful days (on the right). Also, SD of day activity as

TABLE III
OVERALL AND CASEWISE SPEARMAN CORRELATIONS, ρ (WITH SIGNIFICANCE, p) TO BUSYNESS/STRESS/PRESSURE SELF-ASSESSMENT FOR SELECTED FEATURES

CaseID			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
BBI			48	54	50	60	53	38	48	33	43	26	72	48	70	58	43	54	39
Feature	P	RHO																	
WD Trouble/irritation/anger	0.00	0.53	0.65	0.40	0.57		0.54	0.53	0.57	0.83	0.63		0.30	0.42	0.69	0.55	0.40		0.71
WD Tiredness	0.00	0.45	0.60	0.50	0.38			0.54		0.53	0.42		0.49	0.50	0.62	0.68	0.52		0.49
WD Tension/anxiety/fear	0.00	0.41	0.72	0.43	0.41		0.34	0.37	0.37	0.69	0.39		0.44		0.64		0.76		
WD sleep length	0.00	-0.26		-0.26	-0.22						-0.24	-1.00	-0.55	-0.81			-0.43		-0.53
Wrist sleep length	0.00	-0.22											-0.40	-0.68	-0.31		-0.27		-0.40
WD NumWeightEntries	0.00	0.13		0.30			0.38	0.34	0.31				-0.21		0.29				
Wrist day activity SD	0.00	-0.15								-0.36	-0.36								-0.36
Illumination median N	0.00	-0.14	-0.52								-0.40			-0.48					-0.56
HR sport time	0.00	-0.33											-0.55						
Illumination variance N	0.00	0.13							0.34				0.62	0.32					
WD NumExerciseEntries	0.00	-0.10									-0.28		-0.31						-0.33
Scored sleep length	0.00	-0.12												-0.42					
Wrist day activity	0.00	-0.12								-0.32				-0.35					
Scored sleep quality	0.00	0.21						0.44		0.28									
Illumination average N	0.00	0.12	-0.36										0.62						-0.49
WD sleep quality	0.00	-0.10			-0.22			-0.28		-0.36	-0.25		-0.24			0.53			-0.52
WD BP diast morning	0.01	0.10						0.26		0.52	0.24								
HR relaxation time	0.01	0.15	-0.44				0.49			0.33			0.40		0.84				
HR stress time	0.01	0.15																0.59	
Bed time start	0.07	-0.08					0.42	-0.31						-0.53					
HR average HR	0.07	-0.11											-0.46					-0.77	
Bed 20 to 10 bed time	0.10	0.06											0.31	0.64			-0.35		-0.44
HR min HR	0.10	-0.10		0.30						-0.32			-0.53		-0.75			-0.63	
Wrist act ratio N to next D	0.14	0.05														0.29			
Bed time end	0.17	-0.06					0.51	-0.42		-0.36									
Wrist night activity	0.29	-0.04																	
WD weight change	0.32	0.03		0.34				0.30											
WD steps	0.32	0.03											-0.26	0.68	-0.30				0.38
Bed average night HR	0.33	-0.07											-0.34						
Wrist act ratio N to prev D	0.36	0.03														0.32			
WD BP diast evening	0.36	0.03		0.47															
Wrist num sleep periods	0.47	0.02			0.51		0.37			-0.32				-0.42	0.29				0.28
WD exercise	0.51	-0.03																	-0.50
Wrist night activity SD	0.54	-0.02																	
Bed wakeups	0.56	0.03	-0.40	-0.68	0.44				-0.45										
Bed 10 to 20 bed time	0.57	-0.03							-0.40				0.32						
WD BP systolic morning	0.77	0.01		-0.26						0.39									
WD weight	0.85	-0.01								0.30	0.36			-0.52		-0.33	-0.27		
Bed avg sleep 3 nights	0.95	0.00									0.80					-0.35			

Only significant casewise correlations are shown. Workdays included. Most significant features are listed first. N = night, D = day, WD = wellness diary, HR = heart rate. Increased DSPtss was found in cases 3, 6, 7, 8, 11, 13, 14, 16, and 17.

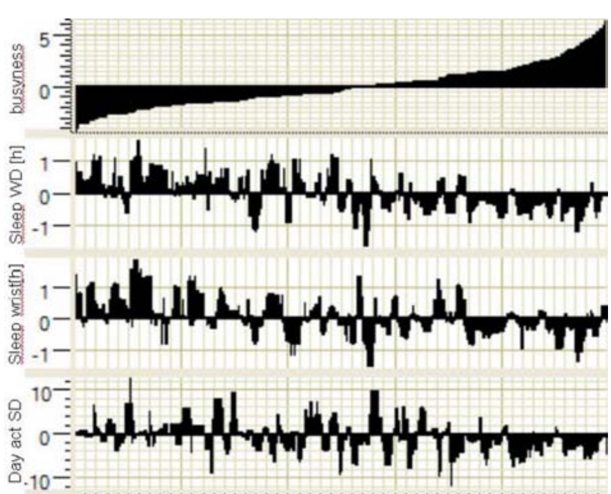


Fig. 5. Feature signals sorted by busyness/stress/pressure [0–10]. Personal average has been subtracted before pooling data together. (1) Busyness/stress/pressure. (2) WD sleep length [in hours]. (3) Wrist activity monitor sleep length [in hours]. (4) Wrist day activity SD [activity count]. Feature signals are median filtered (15 days) to highlight changes in signal level as a function of stress. Data include working days only. Horizontal axis represents working days in order of increasing stress level.

TABLE IV
STATISTICALLY SIGNIFICANT CORRELATIONS TO DSPtss

Feature	p	rho
Bed avg night HR	0.03	0.59
Wrist day activity SD	0.03	0.35
Wrist night activity SD	0.03	0.34
WD exercise	0.04	-0.36
WD weight change	0.04	0.30

Seven-day median values of variables were correlated to DSPtss.

measured using the wrist activity monitor decreases on the most stressful days.

B. DSP and Psychophysiological Variables

Correlations to DSPtss points were obtained by computing the median of psychophysiological variables over seven days (the day people filled in the DSP questionnaire and six days before). Significant ($p < 0.05$) Spearman correlations were found in five variables (Table IV). The personal average was not subtracted, before pooling the data from different cases together, because there are at most four DSPtss measurements per case.

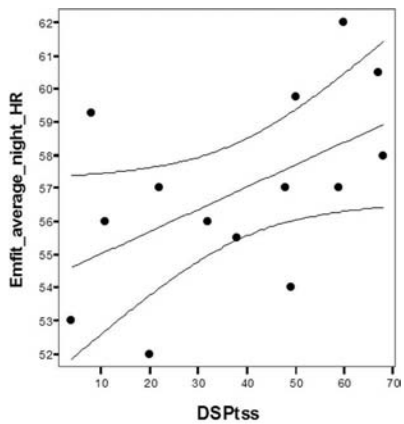


Fig. 6. Average night HR as measured using the bed sensor (y) plotted versus DSPtss (x) ($p = 0.03$, $\rho = 0.59$). Regression line and 95% confidence interval for regression line are shown.

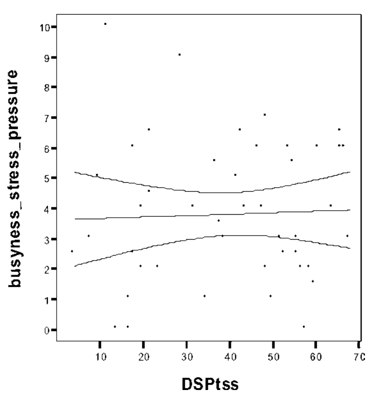


Fig. 7. Busyness/stress/pressure (y) plotted versus DSPtss (x) ($p = 0.64$, $\rho = 0.07$). Regression line and 95% confidence interval for regression line are shown.

Average night HR as measured using the bed sensor showed the highest significant correlation with DSPtss (Fig. 6). This variable is not available in the data of all of the cases as the measurement of HR using the bed sensor was started only in the latter part of the measurements. Even if there are fewer samples in this data, it shows the highest significant correlation with the self-assessed stress level.

C. Correlation of References (DSP, Self-Assessment)

Correlation of DSPtss and self-assessed stress level is visualized in Fig. 7. Self-assessed stress level is computed by taking a seven-day median over six days before and the day when DSP questionnaire is answered.

D. Effect of Rehabilitation

Effect of the two-week intervention (rehabilitation) was assessed by comparing daily stress level (from WD) before and after rehabilitation. The values for comparison were average computed over seven days before intervention and 14 days after intervention. “Busyness/stress/pressure” changes from prerehabilitation 0.90 above personal average to postrehabilitation 0.02 above personal average. Similarly, e.g., sleep length as detected by wrist activity monitor changes from -0.11 h to $+0.11$ h.

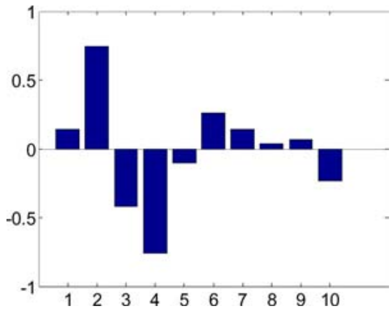


Fig. 8. Average busyness/stress/pressure over all participants—effect of rehabilitation. Rehabilitation periods are on weeks 3 and 4. The personal average was subtracted before pooling data together. Scale is that of WD self-assessment [0–10].

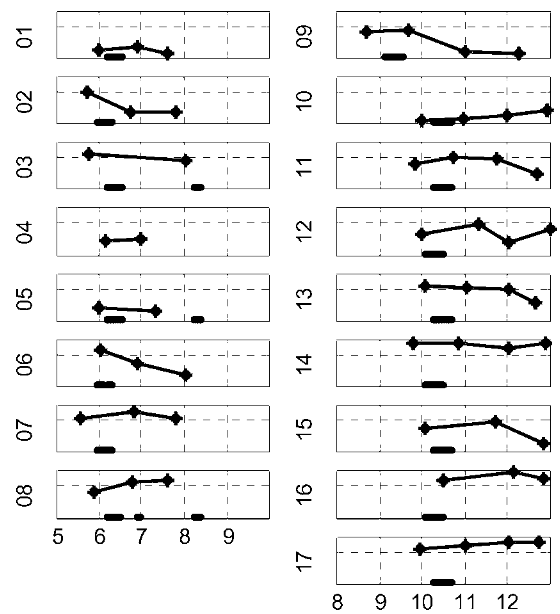


Fig. 9. DSPtsss as function of time. Each graph represents one case. Rehabilitation periods are shown on bottom of each panel (black horizontal line). The Y-axis range is from 0 to 75 DSP points (grid line at 50 points). Increased DSPtss ($\geq$50) was found in cases 3, 6, 7, 8, 13, 14, 16, and 17 at least in one of the measurements. X-axis scale is indicated at the bottom panels (month of year 2005).

Weekly changes in busyness/stress/pressure of all users were aligned so that weeks 3 and 4 are rehabilitation weeks. Average busyness/stress/pressure over all participants is low during rehabilitation, and shows a decreasing trend after rehabilitation (Fig. 8).

The individual DSPtss changes during the study show decrease in 9 of 17 cases possible (Fig. 9).

IV. Discussion

The measurements used in this study were well applicable to monitoring during a rehabilitation program. Even the current prototype, consisting of many devices from different manufacturers, was seen by the clinical expert as an applicable tool providing extra information about the patients. The patients were mostly excited about the prototype, and felt that it gave them objective information about their health.

A. Self-Assessed Stress and Psychophysiological Variables

The results show moderate, but significant overall correlations to daily stress level on working days and DSPtss. Even moderate correlations can be considered important in case of these kinds of data [29]. However, it must be kept in mind that many of the variables (e.g., BP variables) are not independent. Thus, when the value of one variable changes, the dependent variable changes as well. Strong correlations can be found at individual level, but the correlations of different individuals can even be conflicting, which weakens the overall correlations. Thus, based on the results of this study, it can be said that stress is an individual phenomenon. Different people react to stress in different ways. One person may react with blood pressure, while another reacts with disturbed sleep, etc. Thus, finding very specific variables that always indicate stress, when it is present, is a demanding task for future research.

WD self-assessments "trouble/irritation/anger," "tiredness," and "tension/anxiety/fear" positively correlate with "busyness/stress/pressure." This is natural because all four assessments were done usually at the same time. However, the participants gave feedback on daily stress level self-assessment, saying that it was rather difficult to give a number to one's own stress level. They mentioned that the number given to a certain stress level could have changed over time as they got used to assessing their stress level. Self-assessed sleep length (WD sleep length) also shows significant negative correlation with daily stress level. High stress levels are associated with shorter sleep length the night before the stressful day according to self-reported sleep length, wrist activity monitor sleep length, and sleep length as scored by the MD. Individual correlations between stress and sleep length are highly significant. However, the changes in body weight do not significantly correlate with daily stress level. Lower number of exercise entries on stressful days is probably associated with long working hours and tiredness at home, and thus, less exercise. WD self-assessments also show that stressful days are associated with poorer sleep quality the night before. Higher diastolic morning blood pressure is also associated with higher daily stress level later on the day.

Bed sensor variables do not show significant overall correlations with daily stress level. However, there are strong individual correlations, but with different signs. Before stressful days, some people go to bed earlier, some later than normally. Also, time spent in bed before a stressful day is for some people longer and for others, shorter than normally. Features from the bed sensor indicate the time spent in bed, not the time slept. Other sleep length features aim at measuring the total sleep time.

Wrist activity monitor features include features about both day and night. Higher daily stress level is significantly associated with shorter sleep length the night before, and less day activity and day activity deviation on the stressful day. The day activity changes may indicate longer working days and less exercise.

HR features "sport time," "relaxation time," and "stress time" show significant correlations to daily stress level. Sport time clearly decreases on stressful days. Both stress and relaxation time increase on stressful days. Stress time is natural, but longer relaxation time may be caused by inactivity.

Illumination features correlate significantly with daily stress level. The median of night illumination decreases, while average and variance of night illumination increase when stress level increases. This indicates more stress during darker autumn nights, more lights on during night, and thus, wake ups during night.

Some hardware problems were experienced with the continuously measuring devices during the study. The problems encountered caused breaks in the measurement process or data transmission (e.g., a modem was broken due to a thunder storm). Missing data may weaken the significance of observed correlations of variables.

B. DSP and Psychophysiological Variables

Significant correlations between DSPtss and psychophysiological variables were found in five features. High DSP score correlates strongly with higher night HR as measured with the bed sensor. The night HR feature has fewer samples than other variables (HR measurement was not available in all bed sensors). High DSPtss is also associated with larger daytime and nighttime SD of activity. This may indicate more restless behavior when DSPtss is high. Higher DSPtss is also associated with less exercise and increasing weight. The DSPtss was correlated to seven-day median of the feature variables. The set of significantly correlating variables does not remain the same if the representing value is computed over a different time range (say 14 days). This indicates that the actual changes are small compared to noise.

C. Correlation of References (DSP, Self-Assessment)

Correlation between DSPtss and daily stress level self-assessment from WD is poor. This may indicate that the two self-assessments measure different things. WD busyness/stress/pressure is a self-assessment of one day's stress level. However, DSP measures more long-term stress level and even personality. When filling in the questionnaire, the participant assesses how she/he usually behaves, not how she/he behaved today. However, it is not clear which time interval DSP reflects. Knowing this would ease selecting the time range over which to compute the representative values. Also, DSP was seldom measured in this study, only once per month.

D. Effect of Rehabilitation

Two-week rehabilitation that contains, e.g., lectures about stress management and exercise, had an effect on self-assessed stress level. Stress level is clearly lower during rehabilitation than before it. Also, after rehabilitation, stress level is lower than before rehabilitation. A decreasing trend can be found in stress level after the rehabilitation.

E. Comparison to Other Studies

The findings on DSP are in line with [19]. In that study, DSPtss did not correlate significantly with HR in different stressful tasks: 1) counting backwards; 2) cold pressor task; and 3) oral presentation. The authors conclude that in reality, DSP does not measure stress as a dynamic process, but considers stress as a static phenomenon. They mention that DSP should be accompanied with instructions that tie an individual's responses to a particular situation.

Correlations between mental strain and HR variables used in this study have also been studied in [30]. In that study, mental strain and HR variables, computed using Firstbeat PRO WAS, were compared on postal workers. HR was monitored for 24 h, starting from the beginning of a working day. Self-reported mental strain was collected after the recordings by asking the participants to draw a graph of their mental strain during the day on a paper. Of 27 participants, self-reported mental strain was found to have a significant positive Pearson correlation with absolute stress vector (ASV) for 18 participants, and non-significant or negative correlation for 9 participants. ASV is a second-by-second index vector that is computed from HR, HRV high-frequency power, HRV low-frequency power, and respiratory variables derived from HRV. Similarly, there is an absolute relaxation vector (ARV). ASV represents activity of the sympathetic nervous system and ARV that of parasympathetic nervous system. Thus, the findings are in line with the findings of our study. In our study, stress time and relaxation time variables were found to have significant, positive Spearman correlation ($p < 0.01$, $\rho = 0.15$) with self-reported busyness/stress/pressure.

Short stress tests were performed in [31]. The authors found HRV to be a more sensitive and selective measure of mental stress than blood pressure. Our findings partially support these findings. HR variables have stronger groupwise correlation with self-reported stress than blood pressure variables. Our findings also show that of blood pressure variables, diastolic morning pressure has the strongest correlation with self-reported stress. According to [32], after 7.5 years, cumulative job strain increased systolic blood pressure modestly, but significantly. The change was stronger for men than for women.

Activity monitoring was used in [33] together with cortisol and subjective stress and sleepiness scaling. High-stress and low-stress weeks were compared. Results show that during the high-stress week, total sleep time decreased and restlessness at bedtime was significantly increased. Our findings support these findings. In our study, before stressful days, sleep length measured with wrist activity monitor was short, and day activity during a stressful day was low.

In future research, our aim is to concentrate on measurement of recovery periods (sports, night, etc.). The recovery periods play an important role in successful stress management. We believe that any improvement in measurement of these would improve accuracy of automatic stress measurements.

V. Conclusion

Psychophysiological variables measured with unobtrusive wireless sensors in a home environment show significant, but modest, overall correlation with self-assessed stress level when data from different participants are pooled together. Correlations are strong, but sometimes conflicting on individual level. On a daily basis, sleep length variables have moderate negative correlation with self-reported busyness/stress/pressure of the following day. On a longer term basis, higher seven-day average night HR was associated with higher stress level as reflected by the DSPtss. Overall, the nighttime variables and daytime exercise variables gave promising correlations with stress level, highlighting the importance of recovery periods after periods of stress. Further developments in measurement of recovery periods might improve the accuracy of automatic stress measurements. In conclusion, psychophysiological wellbeing may be monitored at home and by using wireless sensors in long-term settings, but data interpretation requires focus on individual patterns rather than a groupwise approach.

Acknowledgment

The authors would like to express their gratitude to participants to the wSense project measurements. They would also like to thank IST Oy, Firstbeat Technologies Oy, Nokia Oyj, Emfit Ltd., Suunto Oy, Normomedical Oy, and Omron Corporation for the study equipment.

References

[1] W. J. Verberk, A. A. Kroon, J. W. M. Lenders, A. G. H. Kessels, G. A. van Montfrans, A. J. Smit, P.-H. M. van der Kuy, P. J. Nelemans, R. J. M. W. Rennenberg, D. E. Grobbee, F. W. Beltman, M. A. Joore, D. E. M. Brunenberg, C. Dirksen, T. Thien, and P. W. de Leeuw, "Self-measurement of blood pressure at home reduces the need for antihypertensive drugs," *Hypertension*, vol. 50, pp. 1019–1025, 2007.

[2] R. Lappalainen, P. Pulkkinen, M. van Gils, J. Pärkkä, and I. Korhonen, "Long-term self-monitoring of weight: A case study," *Cogn. Behav. Ther.*, vol. 34, pp. 108–114, Jun. 2005.

[3] J. M. McGinnis, P. Williams-Russo, and J. R. Knickman, "The case for more active policy attention to health promotion," *Health Aff.*, vol. 21, pp. 78–93, 2002.

[4] S. N. Haynes and D. T. Yoshioka, "Clinical assessment applications of ambulatory sensors," *Psychol. Assess.*, vol. 19, no. 1, pp. 44–57, 2007.

[5] M. T. Tuomisto, T. Terho, I. Korhonen, R. Lappalainen, T. Tuomisto, P. Laippala, and V. Turjanmaa, "Diurnal and weekly rhythms of health-related variables in home recordings for two months," *Physiol. Behav.*, vol. 87, pp. 650–658, 2006.

[6] A. Conrad, F. H. Wilhelm, W. T. Roth, D. Spiegel, and C. B. Taylor, "Circadian affective, cardiopulmonary and cortisol variability in depressed and nondepressed individuals at risk for cardiovascular disease," *J. Psychiatr. Res.*, vol. 42, pp. 769–777, 2008.

[7] S. S. Intille, E. Munguia Tapia, J. Rondoni, J. Beaudin, C. Kukla, S. Agarwal, L. Bao, and K. Larson, "Tools for studying behavior and technology in natural settings," in *UbiComp 2003* (LNCS 2864), A. K. Dey *et al.*, Eds. Berlin, Germany: Springer-Verlag, 2003, pp. 157–174.

[8] F. H. Wilhelm, M. C. Pfaltz, and P. Grossman, "Continuous electronic data capture of physiology, behavior and experience in real life: Towards ecological momentary assessment of emotion," *Interact. Comput.*, vol. 18, pp. 171–186, 2006.

[9] R. Kalimo and S. Toppinen, "Työuupumus Suomen työikäisellä väestöllä," in *Burnout in the Finnish Working-age Population, in Finnish*. Helsinki, Finland: Finnish Institute of Occupational Health, 1997.

[10] A. Aromaa and S. Koskinen, Eds. (2004). *Health and Functional Capacity in Finland—Baseline Results of the Health 2000 Health Examination Survey*, Helsinki, Finland: Publications of the National Public Health Institute [Online]. B12. Available: http://www.ktl.fi/halsa2000/index.uk.html

[11] A. Shirom, "Reflections on the study of burnout," *Work Stress*, vol. 19, no. 3, pp. 263–270, 2005.

[12] H. Herrman, S. Saxena, and R. Moodle, Eds. (2005). *Promoting Mental Health—Concepts, Emerging Evidence, Practice*. Geneva, Switzerland: WHO [Online]. Available: http://www.who.int/mental_health/evidence/MH_Promotion_Book.pdf

[13] M.-L. Kinnunen, "Allostatic load in relation to psychosocial stressors and health," Ph.D. dissertation, Jyväskylä Univ., Jyväskylä, Finland, 2005.

[14] C. Maslach and S. E. Jackson, "The measurement of experienced burnout," *J. Occup. Behav.*, vol. 2, pp. 99–113, 1981.

[15] T. Honkonen, K. Ahola, M. Pertovaara, E. Isometsä, R. Kalimo, E. Nykyri, A. Aromaa, and J. Lönnqvist, “The association between burnout and physical illness in the general population—Results from the Finnish health 2000 study,” *J. Psychosom. Res.*, vol. 61, pp. 59–66, 2006.

[16] F. Jones and B. C. Fletcher, “An empirical study of occupational stress transmission in working couples,” *Hum. Relat.*, vol. 46, no. 7, pp. 881–903, 1993.

[17] P. Näätänen, A. Aro, S. B. Matthiesen, and K. Salmela-Aro, *Bergen Burnout Indicator 15*. Helsinki, Finland: Edita, 2003.

[18] L. R. Derogatis, “The Derogatis stress profile (DSP): Quantification of psychological stress,” *Adv. Psychosom. Med.*, vol. 17, pp. 30–54, 1987.

[19] P. L. Dobkin, R. O. Pihl, and C. Breault, “Validation of the Derogatis stress profile using laboratory and real world data,” *Psychother. Psychosom.*, vol. 56, pp. 185–196, 1991.

[20] D. Shapiro, L. D. Jamner, I. B Goldstein, and R. J. Delfino, “Striking a chord: Moods, blood pressure and heart rate in everyday life,” *Psychophysiology*, vol. 38, pp. 197–204, 2001.

[21] M. Partinen, “Sleep disorders and stress,” *J. Psychosom. Res.*, vol. 38, suppl. 1, pp. 89–91, 1994.

[22] S. J. Linton, “Does work stress predict insomnia? A prospective study,” *Brit. J. Health. Psychol.*, vol. 9, pp. 127–136, 2004.

[23] G. Kecklund and T. Åkerstedt. (2004). Report on methods and classification of stress, inattention and emotional states. Deliverable D1.1.2 of Sensation EU project (IST-507231) [Online]. Available: http://www.sensation-eu.org/deliverables.html

[24] Wellness Diary Application for Mobile Phones. (2007, Sep.) [Online]: Available: http://research.nokia.com/research/projects/WellnessDiary/index.html, accessed 24

[25] E. Mattila, J. Pärkkä, M. Hermersdorf, J. Kaasinen, M. Vainio, K. Samposalo, J. Merilahti, J. Kolari, M. Kulju, R. Lappalainen, and I. Korhonen, “Mobile diary for wellness management—Results on usage and usability in two user studies,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 4, Jul. 2008.

[26] J. Lötjönen, I. Korhonen, K. Hirvonen, S. Eskelinen, M. Myllymäki, and M. Partinen, “Automatic sleep-wake and nap analysis with a new wrist worn online activity monitoring device Vivago Wristcare,” *Sleep*, vol. 26, no. 1, pp. 86–90, 2003.

[27] E. Tuulari, “Enabling ambient intelligence research with soapbox platform,” *Ercim News*, no. 47, Oct. 2001.

[28] J. Merilahti, A. Saarinen, J. Parkka, K. Antila, E. Mattila, and I. Korhonen, “Long-term subjective and objective sleep analysis of total sleep time and sleep quality in real life settings,” in *Proc. 29th Annu. Int. Conf. IEEE/EMBS, EMBC 2007*, Lyon, France, Aug. 22–26, pp. 5202–5205.

[29] K. Antila, M. van Gils, J. Merilahti, and I. Korhonen, “Associations of psychological self-assessments and HRV in long-term measurements at home,” presented at the IFMBE, EMBEC 2005 Conf., Prague, Czech Repulic, Nov. 20–25, vol. 11.

[30] M.-L. Kinnunen, H. Rusko, T. Feldt, U. Kinnunen, T. Juuti, T. Myllymäki, K. Laine, P. Hakkarainen, and V. Louhevaara, “Stress and relaxation based on heart rate variability: Associations with self-reported mental strain and differences between waking hours and sleep,” in *Proc. NES 2006—38th Annu. Congr. Nordic Ergonom. Soc.*, Hämeenlinna, Finland, 2006, pp. 136–139.

[31] N. Hjortskov, D. Rissén, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, “The effect of mental stress on heart rate variability and blood pressure during computer work,” *Eur. J. Appl. Physiol.*, vol. 92, pp. 84–89, 2004.

[32] C. Guimont, C. Brisson, G. R. Dagenais, A. Milot, M. Vézina, B. Mâsse, J. Moisan, N. Laflamme, and C. Blanchette, “Effects of job strain on blood pressure: A prospective study of male and female white-collar workers,” *Amer. J. Public Health*, vol. 96, pp. 1436–1443, 2006.

[33] A. Dahlgren, G. Kecklund, and T. Åkerstedt, “Different levels of work-related stress and the effects on sleep, fatigue and cortisol,” *Scand. J. Work Environ. Health*, vol. 31, pp. 277–285, 2005.

Juha Pärkkä (M’07) received the M.Sc. (Tech.) degree in information technology (digital signal processing) from Tampere University of Technology, Tampere, Finland, in 1997.

He is currently a Senior Research Scientist at Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, Tampere. His current research interests include biomedical signal processing, ubiquitous computing, and personal health systems.

Juho Merilahti received the M.Sc. (Tech) degree in information technology (software engineering) from Tampere University of Technology, Tampere, Finland, in 2006.

He is currently a Research Scientist at Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, Tampere. His current research interests include biomedical signal processing and personal health systems.

Elina M. Mattila was born in Kankaanpää, Finland, in 1980. She received the M.Sc. (Tech.) degree in electrical engineering from Tampere University of Technology, Tampere, Finland, in 2004.

She is currently a Research Scientist at Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, Tampere. Her current research interests include information and communication technology (ICT) in personal health management and biosignal processing.

Esko Malm received the M.Sc. (Tech.) degree in electrical engineering from the University of Oulu, Oulu, Finland, in 1996.

He is currently a Senior SW Designer at Finwe Ltd., Oulu. His current research interests include embedded software engineering and pervasive computing.

Kari Antila received the M.Sc. (Tech) degree in engineering physics and mathematics from Helsinki University of Technology, Helsinki, Finland, in 2003.

He was with Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, Tampere, Finland. He is currently with the Laboratory of Mathematics in Imaging, Brigham and Women’s Hospital, Harvard Medical School. His current research interests include biomedical models, statistical analysis, and visualizations.

Martti T. Tuomisto received the Ph.D. from Karolinska Institute Medical University, Stockholm, Sweden, in 1995.

He is currently the Director of Research and Education at the Institute for Extension Studies and a Adjunct Professor in the Department of Psychology, University of Tampere, Tampere, Finland. His current research interests include behavioral medicine and physiology, behavior analysis and therapy, and processes and methods of intervention in these areas. He is the author or coauthor of more than 60 original scientific papers in English and more than 80 conference presentations.

Ari Viljam Saarinen received the M.D. degree from Helsinki University, Helsinki, Finland, in 1974, and the M.D. licensed medical specialist in neurology from Oulu University, Oulu, Finland, in 1980.

He is the Head of the Neurological Department, Kainuu Central Hospital, Kajaani, Finland. He current research interests include sleep medicine.

Mark van Gils (M'99) received the M.Sc. and Ph.D. degrees from the Technical University of Eindhoven, Eindhoven, the Netherlands, in 1990 and 1995, respectively.

He is currently with Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, Tampere, Finland, as Senior Research Scientist. He holds an appointment as Docent in the area of physiological signal processing at Helsinki University of Technology, and lectures biomedical signal processing at that university as well as at Tampere University of Technology. His current interests include include biomedical signal analysis and interpretation, statistical analysis, data mining, pattern recognition, education, and scientific publishing processes.

Ilkka Korhonen (M'98) was born in Hankasalmi, Finland, in 1968. He received the M.Sc. and Dr.Tech. degrees in digital signal processing from Tampere University of Technology, Tampere, Finland, in 1991 and 1998, respectively.

He is currently the Chief Research Scientist on information and communication technology (ICT) for health area, leading a team in Pervasive Health Technologies at Valtion Teknillinen Tutkimuskeskus (VTT) Technical Research Centre, Tampere. He is a Docent in medical informatics (with specialty on biosignal processing) at Tampere University of Technology, Tampere. His current research interests include use of ICT for health and wellness, personal health systems, and biosignal interpretation methods, and especially their application in critical care patient monitoring, wearable biomedical monitoring, home health monitoring, and eHealth/mHealth. He has published more than 100 original papers in international scientific journals and conference proceedings, and has several patents in the area.

Dr. Korhonen is a member of the IEEE Engineering in Medicine and Biology Society (EMBS) Technical Committee on Wearable Biomedical Sensors and Systems.

Series title, number and report code of publication

VTT Publications 765
VTT-PUBS-765

Author(s)
Juha Pärkkä

Title

Analysis of Personal Health Monitoring Data for Physical Activity Recognition and Assessment of Energy Expenditure, Mental Load and Stress

Abstract

Personal health monitoring refers to the long-term health monitoring that is performed in uncontrolled environments instead of a laboratory, for example, at home or by using wearable sensors. The monitoring is done by individuals alone, usually without guidance from health care professionals. Data produced by personal health monitoring (for example, actigraphy, heart rate, etc.) are currently used more in personal wellness monitoring rather than in clinical decision-making, because of challenges in the interpretation of the long-term and possibly unreliable data. Automatic analysis of long-term personal health monitoring data could be used for the continuous recognition of changes in individual's behavior and health status, and to point out which everyday selections have a negative effect on health and which have a positive effect. This can not be achieved by using sparse measurements in controlled environments.

In this thesis, data analysis was carried out for the recognition of physical and mental load using data from wearable sensors and other self-measurements. Large, annotated data libraries were collected in real-life or realistic laboratory conditions for the purpose of the development of practical algorithms and the identification of the most information-rich sensors and signal interpretation methods. Time and frequency domain features were computed from raw sensor data for the correlation analysis and the automatic classification of the personal health monitoring data. The decision tree, artificial neural network, K-Nearest Neighbor and a hybrid of a decision tree and artificial neural network classifiers were used.

Automatic activity recognition aims at recognizing individual's activities and postures using data from unobtrusive, wearable sensors. Similarly, the unobtrusive, wearable sensors can be used for the assessment of energy expenditure. The quantities measured in this thesis include acceleration, compass bearings, angular rate, ECG, heart rate, respiratory effort, illumination, temperature, humidity, GPS location, pulse plethysmogram, skin conductance and air pressure. The results indicate that several everyday activities, especially those with regular movements, can be recognized with good accuracy. The energy expenditure estimate obtained using movement sensors was found accurate in activities involving regular movements. The sensors that react to the change of activity type without delay were found the most useful for activity recognition. These include accelerometers, magnetometers, angular rate sensors and GPS location sensors.

Automatic assessment of mental load aims at measuring the level of mental load during everyday activities using data from wearable sensors. The assessment of long-term stress aims at finding measures that reflect the perceived stress level, either directly or as observed through changes in behavior. Data were collected with people suffering from long-term work-related stress and participating in a rehabilitation program. Automatic measurements of recovery, measured with a bed sensor, actigraphy and bedroom illumination sensors were found to correlate best with the self-assessed stress level.

Careful selection of sensor types, sensor locations and input features played a more critical role in successful classification than the selection of a classifier. Computational complexity of the classifier's classification phase has an impact on the power consumption of a hosting mobile terminal. Power consumption is one of the bottlenecks in long-term personal health monitoring solutions today.

ISBN
978-951-38-7740-8 (soft back ed.)
978-951-38-7741-5 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN	Project number
VTT Publications 1235-0621 (soft back ed.) 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)	

Date	Language	Pages
June 2011	English, Finnish abstr.	103 p. + app. 54 p.

Name of project	Commissioned by

Keywords	Publisher
Personal health monitoring, biosignal processing and classification, physical activity, activity recognition, energy expenditure, mental load, stress	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

Julkaisun sarja, numero ja raporttikoodi

VTT Publications 765
VTT-PUBS-765

Tekijä(t)
Juha Pärkkä

Nimeke

Henkilökohtaisessa terveydentilan seurannassa syntyvän mittaustiedon analyysiä fyysisten aktiviteettien tunnistamista sekä energiankulutuksen, henkisen kuormituksen ja stressin arviointia varten

Tiivistelmä

Henkilökohtaisen terveydentilan seuranta viittaa pitkäaikaismittauksiin, joita tehdään laboratorion sijaan kontrolloimattomissa oloissa, esimerkiksi kotona tai käyttäen puettavia antureita. Mittauksia tekee yksilö itse, yleensä ilman terveydenhuollon ammattilaisten ohjausta. Henkilökohtaisen terveydentilan seurannasta kertyvää dataa, esimerkiksi aktigrafiaa tai sykettä, käytetään tällä hetkellä enemmän henkilökohtaiseen terveyden seurantaan, kuin kliiniseen päätöksentekoon, johtuen haasteista paikoin epäluotettavan pitkäaikaisdatan tulkinnassa. Pitäkaikaismittauksilla voidaan kuitenkin jatkuvasti arvioida muutoksia yksilön käyttäytymisessä ja terveydentilassa ja osoittaa, millä valinnoilla on terveyden kannalta positiivisia, millä negatiivisia vaikutuksia. Tähän ei päästä harvoilla yksittäismittauksilla kontrolloiduissa oloissa.

Tässä työssä käytettiin puettavien antureiden ja muiden henkilökohtaisten mittausten avulla kerättyä dataa yksilön fyysisen aktiivisuuden ja henkisen kuorman profilointiin automaattisen data-analyysin avulla. Tutkimuksissa kerättiin laajoja, annotoituja datakirjastoja jokapäiväistä elämää vastaavissa ympäristöissä. Datakirjastojen avulla tunnistettiin parhaita antureita sekä kehitettiin käytännönläheisiä algoritmeja datan automaattista tulkintaa varten. Aika- ja taajustason piirteitä laskettiin antureiden tuottamasta raakadatasta korrelaationanalyysiä ja henkilökohtaisen terveysdatan automaattista luokittelua varten. Työssä käytettiin luokittelijoina binäärisiä päätöspuita, neuraaliverkkoja, k-lähimmän naapurin luokittelijaa (KNN) sekä päätöspuun ja neuraaliverkon yhdistelmäluokittelijaa.

Automaattisen aktiviteettien tunnistuksen tavoitteena on tunnistaa käyttäjän aktiviteetit ja asennot päälle puettavien, mutta huomaamattomien ja liikkumista häiritsemättömien antureiden avulla. Samoja antureita voidaan käyttää myös automaattiseen energiankulutuksen tunnistamiseen. Tässä työssä mitattiin kiihtyvyyksiä, kompassisuuntaa, kulmanopeutta, EKG:ta, sykettä, hengitysliikkeitä, valoa, lämpötilaa, kosteutta, GPS-paikkaa, pulssipletysmogrammia, ihon johtavuutta sekä ilmanpainetta. Tulosten perusteella useita arkipäiväisiä aktiviteetteja, erityisesti toistuvaa liikettä sisältäviä aktiviteetteja voidaan tunnistaa automaattisesti hyvällä tarkkuudella. Liikeantureiden datasta laskettava energiankulutusarvio toimii hyvällä tarkkuudella toistuvaa liikettä sisältävillä aktiviteeteilla. Aktiviteettien tunnistuksen kannalta parhaiksi antureiksi osoittautuivat ne, joiden ulostulosignaali muuttuu välittömästi aktiviteettityypin vaihtuessa. Näitä antureita olivat kiihtyvyysanturit, magnetometrit, kulmanopeusanturit sekä GPS-paikannin.

Henkisen kuormituksen automaattisen arvioinnin tavoitteena on henkisen kuormituksen tason mittaaminen puettavilla antureilla, osana arkipäivän elämää. Pitkäaikaisen stressin automaattisen arvioinnin tavoitteena on löytää mittareita, jotka kuvastavat yksilön kokeman stressin voimakkuutta joko suoraan tai käyttäytymismuutoksia seuraamalla. Työssä kerättiin dataa pitkäaikaisesta työstressistä kärsivien ja kuntoutusohjelmaan osallistuvien henkilöiden avulla. Automaattisesti palautumista mittaavien antureiden, sänkyanturin, aktigrafin sekä makuuhuoneeseen sijoitetun valoanturin, tuottama data korreloi voimakkaimmin itseraportoidun stressin kanssa.

Anturityypin, anturin sijainnin sekä piirteiden valinnalla oli suurempi rooli onnistuneessa luokittelussa kuin luokittelijan valinnalla. Luokittelijan luokitteluvaiheen laskennallinen monimutkaisuus vaikuttaa akkukäyttöisen laitteen tehonkulutukseen. Tehonkulutus on tällä hetkellä yksi pitkäaikaisen, henkilökohtaisen terveydentilan seurannan pahimmista pullonkauloista.

ISBN
978-951-38-7740-8 (nid.)
978-951-38-7741-5 (URL: http://www.vtt.fi/publications/index.jsp)

Avainnimeke ja ISSN
VTT Publications
1235-0621 (nid.)
1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

Projektinumero

Julkaisuaika
Kesäkuu 2011

Kieli
Englanti, suom. tiiv.

Sivuja
103 s. + liitt. 54 s.

Projektin nimi

Toimeksiantaja(t)

Avainsanat
Personal health monitoring, biosignal processing and classification, physical activity, activity recognition, energy expenditure, mental load, stress

Julkaisija
VTT
PL 1000, 02044 VTT
Puh. 020 722 4520
Faksi 020 722 4374

VTT PUBLICATIONS

749 Toni Ahonen, Markku Reunanen & Ville Ojanen (eds.). Customer value driven service business development. Outcomes from the Fleet Asset Management Project. 2010. 43 p. + app. 92 p.

750 Tiina Apilo. A model for corporate renewal. Requirements for innovation management. 2010. 167 p. + app. 16 p.

751 Sakari Stenudd. Using machine learning in the adaptive control of a smart environment. 2010. 75 p.

752 Evanthia Monogioudi. Enzymatic Cross-linking of β-casein and its impact on digestibility and allergenicity. 2010. 85 p. + app. 66 p.

753 Jukka-Tapani Mäkinen. Concurrent engineering approach to plastic optics design. 2010. 99 p. + app. 98 p.

754 Sanni Voutilainen. Fungal thermostable cellobiohydrolases. Characterization and protein engineering studies. 2010. 98 p. + app. 55 p.

755 Pirjo Näkki, Asta Bäck, Teemu Ropponen, Juha Kronqvist, Kari A. Hintikka, Auli Harju, Reeta Pöyhtäri & Petri Kola. Social media for citizen participation. Report on the Somus project. 2011. 112 p. + app. 11 p.

756 Tuomas Pensala. Thin Film Bulk Acoustic Wave Devices. Performance Optimization and Modeling. 2011. 97 p. + app. 73 p.

757 Seppo Uosukainen. Foundations of acoustic analogies. 2011. 34 p. + app. 69 p.

758 Tomi Haatainen. Stamp fabrication by step and stamp nanoimprinting. 2011. 70 p. + app. 59 p.

759 Jukka Kääriäinen. Towards an Application Lifecycle Management Framework. 2011. 103 p. + app. 81 p.

760 Maria Antikainen. Facilitating customer involvement in collaborative online innovation communities. 2011. 94 p. + app. 97 p.

761 Petteri Alahuhta. Technologies in Mobile Terminals Enabling Ubiquitous Services. 2011. 127 p. + app. 100 p.

762 Raimo Hyötyläinen. Cellular-networked industrial enterprises in innovation paradigm. 2011. 208 p.

763 Greta Faccio. Discovery of oxidative enzymes for food engineering. Tyrosinase and sulfhydryl oxidase. 2011. 101 p. + app. 672 p.

764 FUSION YEARBOOK. ASSOCIATION EURATOM-TEKES. Annual Report 2010. Eds. by Seppo Karttunen & Markus Airila. 164 p. + app. 13 p.

765 Juha Pärkkä. Analysis of Personal Health Monitoring Data for Physical Activity Recognition and Assessment of Energy Expenditure, Mental Load and Stress. 2011. 103 p. + app. 54 p.

ISBN 978-951-38-7740-8 (soft back ed.)
ISSN 1235-0621 (soft back ed.)
ISBN 978-951-38-7741-5 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)