VTT



EuroXR 2026

Proceedings of Application, Poster, and Demo Tracks of the 23rd EuroXR International Conference

Kaj Helin | Sara Arlati | Armin Grasnick (eds.)

VTT TECHNOLOGY 449

VTT TECHNOLOGY 449

Proceedings of Application, Poster, and Demo Tracks of the 23rd EuroXR International Conference

Kaj Helin (ed.)

VTT Technical Research Centre of Finland Ltd

Sara Arlati (ed.)

STIIMA - CNR, Italy

Armin Grasnick (ed.)

University of Applied Sciences, Germany



ISBN 978-951-38-8814-5

VTT Technology 449

ISSN-L 2242-1211
ISSN 2242-122X (Online)
DOI: 10.32040/2242-122X.2026.T449

Copyright © VTT 2026

JULKAISIJA – PUBLISHER

VTT
PL 1000
02044 VTT
Puh. 020 722 111
https://www.vtt.fi

VTT
P.O. Box 1000
FI-02044 VTT, Finland
Tel. +358 20 722 111
https://www.vttresearch.com

Cover image: The photographers are mentioned in the images.

EuroXR 2026: Valencia, Spain

The EuroXR 2026 conference is held in conjunction with the Virtual Worlds Partnership (VWP) Annual Forum 2026, creating a convergence of scientific excellence, industrial innovation and policy leadership within Europe's immersive technologies ecosystem. The Conference will focus on cutting-edge advancements in eXtended Reality (XR) technologies. This includes innovations in software systems, display technologies, interaction devices, and various applications. In addition to presenting papers on cutting-edge scientific results and highlights from numerous application fields, the EuroXR conference series aims to create a unique human-dimension framework. This framework will interconnect European and international XR communities, fostering knowledge cross-fertilization between researchers, technology providers, and end-users.

3 to 6 November 2026

Valencia, Spain

Conference organizers

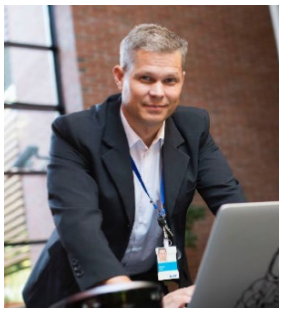
Preface

We are pleased to present these conference proceedings in the VTT Technology series, which contains the papers accepted for the Application and Poster & Demo Track of EuroXR 2026, the 23rd annual EuroXR conference, being hosted from November 3rd to November 6th 2026, at Polytechnic City of Innovation (CPI), Universitat Politècnica de València. This publication is thus a collection of the application papers (talks, posters and demonstrations) presented at the conference. It provides an interesting perspective into current and future applications of VR/AR/MR. EuroXR Conference 2026 is held in conjunction with the Virtual Worlds Partnership (VWP) Annual Forum 2026, creating a convergence of scientific excellence, industrial innovation, and policy leadership within Europe's immersive technologies ecosystem.

In previous years, under the name EuroVR, the Conference has been held in Bremen (2014), Lecco (2015), Athens (2016), Laval (2017), London (2018), Tallinn (2019) and Valencia (2020); and as EuroXR in Milan (2021), Stuttgart (2022), Rotterdam (2023), Athens (2024), and Winterthur (2025). The focus of the EuroXR conferences is to present, each year, novel eXtended Reality (XR) technologies, including software systems, display technologies, interaction devices, and applications, to foster engagement between industry, academia, and the public sector, and to promote the development and deployment of VR/MR/AR technologies in new, emerging, and existing fields. This annual event of the EuroXR association (https://www.euroxr.org/) provides a unique platform for exchange between researchers, technology providers, and end users around commercial or research applications.

We would like to warmly thank the industrial committee chairs for their great support and commitment to the conference, and special thanks go to the local organizing committee for their great effort in making this event happen 😊.

On behalf of the organising committee,



Kaj Helin
EuroXR Advisoboard member
Principal Scientist at VTT Technical Research Centre of Finland Ltd, Finland



Sara Arlati
EuroXR Application track
Researcher at STIIMA, CNR, Italy



Armin Grasnick
EuroXR Poster and demo,
Professor at International University of Applied Sciences, Germany

Application Program Co-chairs

- **Sara Arlati,** STIIMA, CNR, Italy
- **Kayvan Mirza,** Optinvent, France
- **Kaj Helin**, VTT, Finland
- **Tina Katika,** ICCS, Greece
- **Barbara Schiavi,** Immersion, France
- **Konstantinos Smagas,** Geoimaging, Cyprus
- **David Nahon,** Dassault Système, France
- **Javier Marín-Morales,** LabLENI, Universitat Politecnica de Valencia, Spain

Posters and Demos Program Co-chairs

- **Armin Grasnick,** International University of Applied Sciences, Germany
- **Michele Gattullo,** Polytechnic University of Bari, Italy
- **Despina Michael-Grigoriou,** Cyprus University of Technology, Cyprus
- **Dang Ngoc-Son,** University of Minhon, Portugal
- **Arcadio Reyes-Lecuona,** Universidad de Málaga, Spain
- **Jorge León Posada Velásquez,** Vicomtech, Spain
- **Maria Eleonora Minissi,** LabLENI, Universitat Politecnica de Valencia, Spain

General Co-chairs

- **Mariano Alcaniz Raya,** LabLENI, Universitat Politecnica de Valencia, Spain
- **Albert "Skip" Rizzo,** University of Southern California, USA
- **Maria Roussou**, National and Kapodistrian University of Athens, Greece

Scientific Program Co-chairs

- **Mario Lorenz,** TU Chemnitz, Germany
- **Gabriel Zachmann**, University of Bremen, Germany
- **Rachel Farrell,** UCD, Dublin, Ireland
- **Voicu Popescu,** Purdue University, USA
- **Maki Sugimoto,** Keio University, Japan
- **Patrick Bourdot**, CNRS & University Paris-Saclay, France

Organizing Team

- **CEEI, Valencia** - Laura Sánchez Membrives, Carmen Renovell
- **Euro XR** –Patrick Bourdot, Mónica González, Jerome Perret, and Gabriel Zachmann
- **LabLENI, UPV -** Mariano Alcaniz Raya, Javier Marín-Morales, Maria Eleonora Minissi, Fernando López, Paz Sebastiá.

Table of contents

APPLICATION TRACK

EuroXR 2026: Valencia, Spain 1
Preface 3
Table of contents 7
Scaling Immersive Technologies in Youth Work: A Large-Scale XR Case Study for Training, Onboarding, and Implementation in Non-Formal Education 13
Immersive Training in the Metaverse:A Multi-User VR System for Contact-Center Operator Onboarding 18
Training Without Moving Aircraft: A Modular Virtual Reality Platform for Airport Operational Teams 22
Virtual Reality as a Tool for Improving Interpretation of Machine Component Drawings with Consideration of Spatial Ability Differences 26
Large-scale Design of VR-based Trainings for E-Mobility at Robert Bosch GmbH: Method and Experiences 41
Embodied Biosignals: Holographic Biofeedback and Interactive Spatial Visualization for Physiology Education 46
Smart Glasses versus Hand Scanning for Automotive Parts Sorting in a Circular Economy Testbed: A Pilot Evaluation 50
CoBotXpeRience: An XR-Based Training Framework for Collaborative Robot Programming and Industrial Upskilling 53
Overcoming Workforce Shortages with a VR Digital Twin: The Liege Airport Case Study 58
XR Aircraft Dismantling Assistant for Aerospace Circularity 62
AI-GLASS: Physics-Informed AI and Pixel-Streamed Digital Twin for Glass Manufacturing 66
Toward a Digital Gardener: An AI-Enabled XR Cross-Platform for Hydroponic Culture Inspection 70
Immersive Collaborative VR for Astronaut Training to Mars missions 74
Mobile XR-based C2 for better situation awareness – Field trials in Bulgaria 82
Training the Mission, Not Only the Aircraft: A Co-Designed VR Simulator for UAV-Assisted Avalanche Search and Rescue 86
GeoBoardXR: A WebXR-Based Immersive Terrain and Geospatial Data Viewer 91
SPACE+: Prototyping a Spatial AI Ecosystem for Participatory Spatial Management in XR 96
ITI MIND: Translating Five Years of Mixed Reality Research for Mental Health and Wellbeing into a Platform for Real-World Care Settings 100
Evaluating a Co-Designed Immersive Virtual Reality Prototype for Social Anxiety Support in Young People 104
From Virtual Operating Room to Training Tool: User Experience and Acceptance of an Immersive VR Anesthesia Simulation 109

A Virtual Reality-augmented treadmill platform to improve walking patterns in patients with vestibular diseases 113

A Virtual Reality Application for Enhancing BCI Motor Imagery Training through Multisensory Stimulation 117

ACTA-VR: Iterative Design of Ecological VR Tasks for Cognitive Flexibility in Older Adults 121

XR Compliance Matrix: Linking Application Contexts, Requirements and Verification Pathways 127

Expressive Virtual Humans for Conversational Simulation in XR: Challenges and Solutions 131

Cross-Platform Dual-Haptic Integration in XPBD with Real-Time Data Synchronization 135

An open-source framework for prototyping virtual reality applications with force-feedback 139

MetaCity Liverpool: A Comprehensive, Multi-Modal Digital Twin Framework 144

Integration of Artificial Intelligence into Virtual Reality Training Systems for Public Speaking 148

XREL: a Hybrid eXtended REality Learning experience 152

Interactive VR360 for Scientific Heritage: A Gaze-Based Approach to Demonstrating Historical Instruments 156

Inmersidra: Bridging Virtual Gestures and Physical Action in XR for Asturian Cider Pouring 160

Building the Ecosystem: A Multi-Level Approach to Scaling XR in Higher Education 164

Instructor-led Mobile Augmented Reality for Synchronous Lessons 169

Dog Assisted Virtual Environment (DAVE) 173

POSTERS AND DEMOS

GhostLights: Investigating Fog-of-War Alternatives in Asymmetric Collaborative Mixed Reality .. 179

Navigating XR Standardization: A Living, Multidimensional Standards Register for XR and Virtual Worlds 183

From Concept to Deployment – A Physical AR Workstation for Washing Machine Disassembly, Inspection and Component Recovery 187

Appearance Control via Environmental Light Modulation: Performance Limits with a Transmissive LCD 191

LIMIAR: A VR Decision Simulator for Residential Care Professionals 195

Influence of Material Surfaces on the Quality of Spatial Maps Generated by the Microsoft HoloLens 2 198

Toward Reliable Pupillometry in eXtended Reality: A Video-Based Pipeline for Isolating Cognitive Workload from Pupil Data 202

Immersive Cultural Heritage Experiences Through an Extended Reality Digital Twin of the Ancient Hydraulis 206

From Physical Organ Pipes to Immersive Digital Heritage: Reverse Engineering, Digital Libraries, and XR Exploration 208

Development of a 3D Digital Twin for a Filament Winding Industrial Facility: An Immersive User Interface for Industry 4.0 211

Exploring Local Dorsolateral Prefrontal Cortex Dynamics During Virtual Navigation Using Virtual Reality and Stereo-Electroencephalography 214

Perceived Motion Fidelity in Hyper-Realistic VR Avatars: A Preliminary Comparison of IK-Only and IK+MediaPipe Tracking 218

Behaviourally expressed, behaviourally silent: attention and presence in VR telemetry 224

HaptiSense: A Wearable Device for Structured Vibrotactile Interaction toward Semantic Haptics in XR..... 228

MIDTOWN: a Model based Immersive Digital Twin Overview With Novice-friendly visualizations..... 232

Feasibility of Simultaneous EEG Recording During PCVR-Based Extended Reality Experiences: An Assessment of Signal Quality, Spectral Measures, and Event-Related Potentials..... 236

DemoXR: A Feasibility Study of Measurement-Grounded AR Instruction Drafting from Robot Demonstrations..... 240

Externalising and Regulating Emotion through Breath: A Therapist-Supervised Multi-User Mixed Reality Intervention..... 245

Enabling Natural Interaction with Everyday Physical Objects in a Mixed Reality Animal-Assisted Therapy Scenario 250

Synthetic Environment for Navigation, Simulation and Operations for Robotics 255

Perceptual Limits of Avatar Movement Distortion in Embodied VR 258

Floor-projected UI for hands-free interaction with AGVs in intralogistics processes..... 263

From Automated Guided Vehicle to Telepresence Robot: Retrofitting an In-House AGV for VR-Controlled Telepresence 267

A chance to act: Bystander behaviour after an embodied experience of harassment..... 271

Assessment methodologies for the perceived quality of split-domain light field visualization in single- and simultaneous-participant studies 275

Direct comparison of 3D content on glasses-free light field displays and in extended reality: experimental design alternatives and preliminary observations..... 279

An AI Data Assistant for Collaborative XR Inspection of Building Digital Twins 283

An Interactive Virtual-Reality Platform for Cultural Heritage Exploration, Risk Assessment and Emergency Preparedness..... 287

Impact of Virtual Reality Immersion on Upper-Limb Object Manipulation..... 291

A Virtual Reality System for Enhanced Indoor Fire Evacuation Training and Management..... 296

A.R.C.O.: A Multimodal XR-LLM Cognitive Assistant for Technical Inspections in Virtual Worlds..... 300

LoomaXR**:** An Engine-Agnostic Hub for Collaborative XR Scene Creation..... 304

XR Skills Between Campus and Workplace in Ten Countries..... 309

Feel the Hue: Exploring Colour-Vibrotactile Correspondences in VR for Affective Regulation..... 312

Supporting Low Vision in VR: Accessibility Challenges and Design Principles 317

Material Degradation and Structural Behavior of Cultural Heritage Burial Stone Column through XR..... 321

Voice-Driven Generation of 3D Assets and Immersive Environments in Mixed Reality 324

XR NeedStruct: From Free-Text to Structured XR User Requirements for Industrial XR Prototyping..... 330

HoloHaptic Calls: Powering real-time, bidirectional haptic-augmented holographic communications using EU-based commercial-ready XR technology..... 334

A Traffic Interception Framework for Auditing Biometric Data in VR Applications..... 338

Device Tracking Error and Headset Slip: Two Error Sources in HMD-Reported Head Pose for Rehabilitation 343

One Twin, Many Realities: A Collaborative Physics AI Digital Twin Across XR Virtual Worlds 347

Immersive FVV Live Platform: enabling Multiuser Cultural Experiences with AI and Volumetric Video 351

Embodied Safety Moderation in Social VR 355

From Digital Resources to Immersive Learning: The Alliance4XR Platform 359

Virtual Embodiment in Healthy Participants and Individuals with Spinal Cord Injury: Preliminary Evidence from a Multimodal VR Protocol 363

Experiencing the Binaural Rendering Toolbox: An Interactive Demonstration of 3D Binaural Audio for XR 367

Towards Dynamic Auralisation of Historic Church Organs in Virtual Environments 371

Towards Bioadaptive VR for Autism: An EDA-Driven Adaptation of a Conversational Virtual Human 374

Forgenie: A multi-input augmented reality interface for creating 3D models 378

From Voice-Coaching Metaphors to Integrated Breathing, Voice and Gesture Practice in Virtual Reality 382

Extending the Surgical View: A Collaborative XR Interface for Spatial Communication in Robotic Surgery 388

Design and Evaluation of an XR-Based Immersive Teleoperation System with Multimodal Operator Monitoring 392

Introducing an IEC 62559-Based Use Case Repository and Architecture Model for Virtual Worlds 397

A Virtual Reality System for Pericardiocentesis Training 403

Composing RFC 8628, RFC 7636 and OpenID Connect for Standalone XR Training: Sign-In Outside the Application and Per-User xAPI Attribution 407

Author Index 414

Tables of Figures 422

Lists of Tables 429

Application track

Scaling Immersive Technologies in Youth Work: A Large-Scale XR Case Study for Training, Onboarding, and Implementation in Non-Formal Education

Adam Leech, Tom Rickard, and Dr. Eemer Eivers

Foróige, The National Youth Development Organisation, Ireland

Corresponding author: adam.leech@foroige.ie

Keywords: Training and Education, Youth Work, Non-Formal Education, Large-Scale Deployment, Social Skills, Inclusion, Digital Youth Work.

Abstract

Traditional youth work and non-formal education frameworks frequently encounter physical, geographic, and social barriers when attempting to reach isolated demographics. While Extended Reality (XR) offers an alternative digital environment capable of bypassing these limitations, widespread deployment is structurally restricted by a baseline digital divide among practitioners. This paper presents a large-scale case study evaluating Foróige GoVirtual (Eivers. 2026), a national initiative that successfully upskilled over 300 youth workers (from 2023 - 2025). Practitioners had little to no virtual reality (VR) experience and were trained into proficient XR facilitators confident in delivering targeted Youth Work Interventions. Evaluating pre- and post-training data alongside multi-stakeholder feedback, we show that targeted training significantly elevated frontline staff confidence (from baseline averages as low as 2.4/10 to over 7.0/10). Consequently, these upskilled practitioners effectively facilitated immersive sessions for over 1,500 young people annually, fostering measurable improvements in self-confidence, leadership, communication and empathy among other skills. This case study offers an empirical foundation for structural hardware management and advocates for the formal integration of digital youth work into non-formal education qualification pathways.

Introduction and Context

Equitable access to in-person services remains a challenge in youth work. Socioeconomic obstacles, sparse populations, lack of public transport, and geographic isolation often leave rural or marginalised young people disconnected from in-person youth work services. Furthermore, young people experiencing severe social anxiety, neurodivergence (e.g., Autism Spectrum Disorder, ADHD), or physical mobility limitations often find physical face-to-face environments challenging, overwhelmingly stressful or inaccessible.



***Figure 1**. Young People and Youth Workers engaging remotely in a safe virtual space.*

Digital youth work has emerged as a critical frontier to meet young people where they are. Virtual Reality (VR) platforms can establish safe, immersive environments that mitigate barriers to participate. However, a significant operational challenge exists: frontline practitioners typically have low technical confidence and lack formal training in XR hardware, software, and online safeguarding. This case study examines Foróige GoVirtual, a national initiative aimed at scaling organisational capability by equipping youth workers with the technical competencies required to facilitate virtual spaces (Foróige, 2025) for marginalised young people.

Training and Capacity Building

The GoVirtual scaling structure was designed to establish an organisation-wide baseline of digital literacy. To transition frontline staff from novice observers to active digital facilitators, a multi-tiered training curriculum was deployed nationally between 2023 and 2025:

- Level 1 Training (Foundational): This initial phase introduced participants to foundational VR architecture. Core competencies included hands-on with the hardware (setting up Meta VR headsets), account management, basic safety parameters (guardian boundaries), virtual safeguarding, and introductory application navigation utilising games like First Steps for Quest II. The level 1 training also included matching XR use to youth work interventions, and allowed practitioners to explore and identify the added value of XR technology to their practice and interventions.
- Level 2 Training (Advanced): Aimed at expanding practitioners' skillsets into multiplayer virtual environments, this level focused on facilitating youth work inside approved, private multiplayer environments (e.g., Rec Room). Training activities emphasised collaborative real-time participation, troubleshooting, and peer-to-peer practical learning.
- Hybrid Integration: A combined approach rolled out in 2025 to integrate Level 1 foundational training and Level 2 advanced training to ensure participants were upskilled promptly. This training was tailored to the specific staff roles attending the training. By tailoring to specific staff roles, the XR technology became more relevant to the projected outcomes of specific youth projects, youth work provision and targeted interventions.

Quantitative Evaluation of Staff Training and Development

To evaluate training efficacy, independent mixed-methods analysis tracked practitioners' self-rated confidence across six core domains on a 10-point metric. Prior to foundational training, baseline staff confidence was limited, particularly regarding application facilitation (Mean = 2.4) and locating resources and supports (Mean = 3.0) (Table 1 below). Post-training evaluations revealed an exponential increase in technical capability, with mean confidence tracking well above 7.0 across all recorded performance indicators.

***Table 1**. Aggregate internal evaluation metrics charting foundational training groups (N=108 pre-training; N=91 post-training)*

GoVirtual Competencies	Pre-Training Confidence Mean (1-10)	Post-Training Confidence Mean (1-10)
Meta Account Configuration	3.8	7.5
VR Headset Setup/Calibration	3.5	7.4
Using Headsets Safely	3.5	8.1
Safeguarding Young People	3.2	7.3
Facilitating Young People in VR Environments	2.4	7.1
Accessing Resources & Support	3.0	8.0

Intervention Delivery and Practical Application

Following systemic upskilling, approximately two-thirds of the trained cohort successfully embedded VR applications into their professional youth work toolkits. Operational tracking from 2023 to 2025 indicated that the model scaled effectively, moving from an early baseline of 400 participants up to an estimated 1,500 young people annually by 2024 and 2025.

While a primary objective of GoVirtual was remote facilitation, surveyed Youth Workers indicated integration of the technology for localised, in-person delivery as well: 40% utilised VR in-person on a regular basis (weekly/bi-weekly), and 39% deployed it routinely for face-to-face group sessions.

Trained practitioners tailored specialised applications to explicitly map onto targeted Youth Work goals:

1. **Socio-Emotional Dysregulation and Anxiety Management**: Frontline staff utilised custom programmes like Dala (Adaptas, 2023) (co-developed with clinical child and adolescent mental health teams) to guide students through school-related anxiety and avoidance behaviours, alongside immersive mindfulness platforms like Tripp
2. **Transversal Skill and Employability Development**: Practitioners leveraged software like Bodyswaps to create safe, repeatable testing environments for public speaking, conflict resolution, and job interview simulation.
3. Collaborative and Creative Spaces: Shared creation applications (e.g., Gravity Sketch, Kingspray, PatchWorld and TribeXR) were integrated into youth work settings to support asset creation, music production, and interactive collaborative teamwork. Collaborative Spaces also gave the opportunity to both youth workers and young people to showcase their creations (Spatial galleries, n.d), customised Rec Room Spaces, live music performance, and real-world application of the assets and skills developed)

Analysis of Targeted Youth Outcomes

The multi-stakeholder evaluation tracked clear socio-emotional outcomes across the Foróige GoVirtual initiative from 2023 - 2025. This outlined a number of outcomes related to the integration of Virtual Reality into Youth Work practice.

A primary success metric of the GoVirtual model was its accessibility. For young people experiencing physical barriers (wheelchair users for example) the virtual space offered a completely level playing field. Avatars removed physical barriers, enabling young people to connect with one another in a safe and accessible space. Similarly, geographically isolated young people (islanders, rural isolation, young carers) overcame transport constraints to establish consistent social circles. Over 80% of practicing youth workers confirmed that VR was an exceptionally effective tool for resolving geographic and social isolation.

Youth Workers reported extensive longitudinal benefits regarding participants' soft skills and self-confidence:

- **Self-Confidence & Problem Solving:** At least 70% of active Youth Workers noted self-confidence increases and heightened problem-solving capabilities in a majority of the young people they were working with.
- **Real-World Social Skill Application:** 43% of surveyed Youth Workers noted noticeable improvements across almost all participating young people regarding communication skills. By replacing the pressure of direct eye contact and hard-to-read facial expressions with a shared digital objective, socially anxious and neurodivergent participants became more comfortable in practicing communication skills. Qualitative data verified that these skills transferred back into real-world settings, with parents noting decreased social anxiety and re-established community friendships.

Human Factors, Friction Points, and Operational Insights

An essential component of the evaluation was documenting systemic human-technological friction points to refine subsequent operational frameworks. Despite strong systemic outcomes, approximately 37% of the trained Youth Workers reported rare or non-existent utilisation of the hardware. The independent evaluator isolated three main barriers:

- **Facilitator Confidence Deficits (30%):** A primary obstacle was localised confidence. Youth Workers noted that if an intervention did not occur immediately following a training session, their technical abilities naturally decreased, creating a barrier to starting later.
- **Hardware Maintenance Overhead (22%):** Managing localised hardware deployment presented noticeable administrative challenges. Keeping standalone headsets charged, managing software updates, and navigating unstable internet connections in remote rural areas added overhead to already busy workloads.
- **Physiological Discomfort (26%):** Sensory needs and motion sickness were identified as significant personal constraints for over a quarter of non-utilising staff members, directly preventing them from maintaining long-term presence within virtual spaces.

Qualitative feedback isolated a compelling dynamic unique to immersive facilitation, termed ‘sideways communication’. In traditional one-to-one Youth Work settings, face-to-face eye contact can feel confrontational or intensely stressful for a vulnerable young person. In contrast, within a virtual environment, the focus shifts to a shared interactive task (e.g., collaborative graffiti art or a virtual game). By engaging in a shared task side-by-side, participants felt less scrutinised, allowing them to communicate openly and candidly with their youth worker from their own comfortable space.

Significance and Recommendations

The findings of this evaluation offer actionable pathways for organisations seeking to operationalise large-scale XR deployments in public and non-formal education domains:

- **Transition to Centralised Device Management**: To minimise local administrative overhead, organisations should move from localised headset maintenance to centralised Mobile Device Management software. Centralised cloud deployment permits bulk configuration updates, safety parameter locking, and automated application pushing, removing technical burdens from local facilitators.
- **Establishing Continuous Communities of Practice**: To combat confidence dips among newly trained staff, organisations must supplement point-in-time training with ongoing peer-support structures. Weekly technical open-office hours and peer mentoring networks ensure that practitioners can troubleshoot emerging technical challenges in a supportive environment.
- **De-linking Funding from Strict Local Geographies:** Traditional community work funding frameworks are tied to physical, localised jurisdictions. However, isolated target groups (e.g., rare physical conditions, young carers, or extreme school avoiders) are geographically dispersed. To optimise XR's unique benefits, funding models must adapt to support practitioners operating with national or regional digital remits.
- **Academic Integration into Non-Formal Education Curriculum:** Given the deeply digital reality of many young people, digital youth work should no longer be treated as an optional specialty. Higher education programmes leading to professional qualifications in youth work, social work, and community development should formally integrate digital tools and immersive pedagogy as standard components of their core qualification paths.

References

Adaptas (2023). At the forefront of Developing VR for Learning and Wellbeing. [online] Adaptas, Leadership Training and Development. Available at: https://adaptastraining.com/can-virtual-reality-help-reduce-anxiety/ [Accessed 14 July 2026].

Bodyswaps®. (n.d.). Bodyswaps: VR Soft Skills Training Provider. [online] Available at: https://bodyswaps.co/ [Accessed 14 July 2026].

Eivers, E. (2026). Evaluation of Foróige GoVirtual. Report. Available at: https://www.foroige.ie/app/uploads/2026/04/GoVirtual-review-with-appendices.pdf [Accessed 14 July 2026]

Foróige. (2025). Case Study 1 - Foróige. [online] Available at: https://www.foroige.ie/our-work/programmes/foroige-go/govirtual/case-study-1/ [Accessed 14 July 2026].

Foróige. (2025). Case Study 2 - Foróige. [online] Available at: https://www.foroige.ie/our-work/programmes/foroige-go/govirtual/case-study-2/ [Accessed 14 July 2026].

Gravity Sketch. (2024). Home - Gravity Sketch. [online] Available at: https://gravitysketch.com/ [Accessed 14 July 2026].

Meta. (2026). Kingspray Graffiti. [online] Available at: https://www.meta.com/en-gb/experiences/kingspray-graffiti/2082941345119152/?srsltid=afmboordobk_kpzv6jgfoa3mot-lqhwmmefquv-vgl8f8hbpeaqr1gqj [Accessed 14 July 2026].

PatchXR (2026). PatchWorld. [online] patchworld. Available at: https://patchxr.com/ [Accessed 14 July 2026].

TribeXR DJ in VR. (n.d.). TribeXR DJ in VR. [online] Available at: https://www.tribexr.com/ [Accessed 14 July 2026].

TRIPP. (n.d.). TRIPP: Fitness for your Inner Self. [on line] Available at: https://www.tripp.com/ [Accessed 14 July 2026].

Spatial. (n.d.). Spatial - Create, Share & Experience Your Creativity In 3D. [online] Available at: https://www.spatial.io/ [Accessed 14 July 2026].

Rec Room. (n.d.). Rec Room. [online] Available at: https://recroom.com/ [Accessed 14 July 2026].

Immersive Training in the Metaverse:A Multi-User VR System for Contact-Center Operator Onboarding

Seyed Mahed Mousavi[1], Raffaele Croce[2], Olha Khomyn[1], Stefania Viapiana[2], Eugenio Fumarola[2], Vincenzo Giliberti[2], and Giuseppe Riccardi[1]

[1]*Signals and Interactive Systems Lab, University of Trento, Italy*

[2]*TP Italy*

Corresponding author: Giuseppe.riccardi@unitn.it

Keywords: Immersive Training, Collaborative and Distributed XR, Multi-Modal Education

Introduction

Contact-center operators work in high-variability environments where performance depends not only on procedural knowledge but on real-time decision-making, communicative competence, and emotional regulation. Traditional training methods (e.g., classroom instruction, role-play, and shadowing) are limited in their capacity to reproduce the complexity and pressure of operational scenarios in a controlled manner (Schulz et al., 2024). Virtual Reality (VR) addresses this gap by providing a substrate for situated, experiential learning in which trainees practice within realistic simulations of their operational context (Zechner et al., 2023). Studies show that immersive environments increase engagement, social presence, and skill transfer (Walenta et al., 2026), and promote the development of communicative and relational competencies (Moore et al., 2022), that are central to contact-center performance. Multi-user environments extend this potential by enabling collaborative training scenarios, promoting competence development through co-presence and joint activity (Batra et al., 2025). The organizational relevance of this approach is confirmed by large-scale industry deployments such as Accenture[1] and Boston Consulting Group[2]. This paper presents HOPE Metaverse Training (Figure1), a multi-user VR environment designed for the onboarding and training of contact-center operators, supporting role-differentiated trainer and trainee avatars, one-to-one and one-to-many session configurations, first-person embodiment, and gesture-based interaction. The system integrates real-time synchronized audio with multi-party speaking, lip-synced avatars, NPCs, and in-scene access to interactive multimodal documents, and digital 3D replicas of operational elements within a shared immersive scene.



***Figure 1**. A training session in HOPE Metaverse Training: the trainer explains the specifications of electricity meter, interacting with a digital 3D replica of the object, while trainees follow along with the corresponding multi-modal manual displayed on their screen within the shared virtual scene.*

HOPE Metaverse Training

HOPE Metaverse Training is a multi-user VR environment developed in Unity for Meta Quest, designed to support the training through realistic simulation of operational scenarios. The system reproduces a contact-center setting with operational workstations, NPCs populating the environment, multimodal documents, and digital 3D replicas of operational elements, enabling trainees to interact with content and other participants within a shared immersive scene. The system implements two distinct roles: trainer and trainee, Figure 2. The trainer configures the session, loads scenario content, and intervenes in real time within the shared scene. The trainee participates in a first-person perspective, navigating the environment and engaging with content and peers. Two session configurations are supported: one-to-one and one-to-many, accommodating both individual and group training modalities. Interaction is gesture-based: users navigate the virtual environment via walking and signal participation via hand-raising. Real-time synchronized audio and multi-party speaking enable natural communicative dynamics between trainer and trainees within the shared scene. Avatar lip sync, driven by voice amplitude, ensures audiovisual coherence during verbal interaction, reinforcing social presence and turn-taking legibility (Peixoto et al., 2021).

Multimodal documents and digital 3D replicas are placed by the trainer within the scene as interactive elements, spatially anchored to the operational environment and preserving presence continuity during training activity. The digital 3D replicas represent the physical products operators are trained to support (e.g., an electricity meter in an energy provider campaign, a mobile device in a consumer electronics campaign). Operators examine these replicas from multiple angles within the immersive environment, directly alongside the corresponding multimodal documentation, building product familiarity and linking technical knowledge to a physical referent without leaving the immersive context. The architecture is modular, separating responsibilities across avatar management, locomotion, networking, audio, and lip sync components, ensuring extensibility for future training scenarios and content updates.

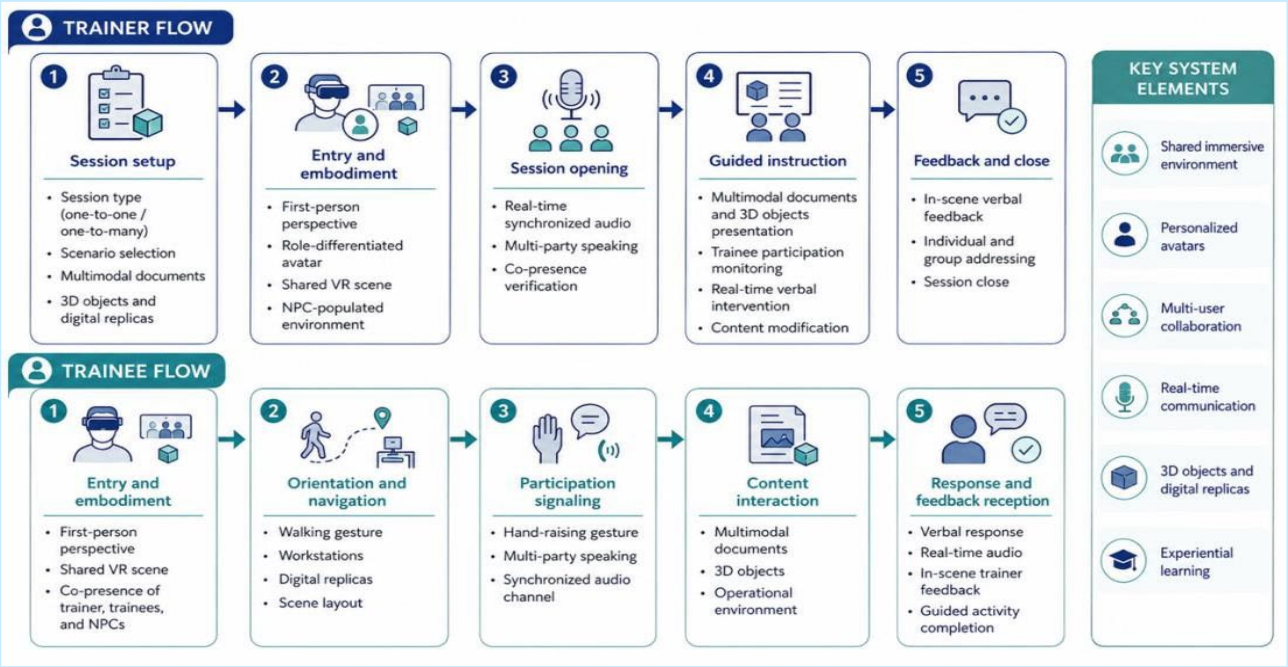


***Figure 2**. Interaction flows in HOPE Metaverse Training. The trainer flow (top) covers session configuration and opening with real-time synchronized audio, guided instruction through multi-modal documents and 3D objects, and in-scene feedback delivery. The trainee flow (bottom) covers first-person embodiment, spatial navigation, participation signaling via gesture, content interaction, and verbal response with feedback reception. Key system elements are summarized on the right.*

Deployment and Planned Evaluation

The beta version of HOPE Metaverse Training has undergone internal testing between two partner institutions located in different cities, connected remotely in the multi-user context. Testing sessions validated the core pipeline across multi-user synchronization, avatar instantiation, locomotion, real-time audio, and lip sync. Latency was assessed through human observation during testing sessions; avatar movement, voice communication, and lip sync were judged by participants as sufficiently fluid and coherent to sustain co-presence and natural interaction.

A human evaluation with approximately 15 participants is planned following the internal validation phase. The evaluation will measure usability, user satisfaction, perceived presence, and operator competence. These measures are selected to assess the relationship between immersive training conditions and the perceived empowerment of contact-center operators, in line with evidence on the mediating role of engagement and immersion in VR-based training effectiveness (Zechner et al., 2023).

Conclusion and Future Work

This paper presented HOPE Metaverse Training, a multi-user VR environment for contact-center operator training, integrating first-person embodiment, real-time synchronized audio, multi-modal documents, and 3D objects within a shared immersive scene. Internal testing between two remote partner sites validated the core pipeline. A user evaluation with approximately 15 participants is planned.

Acknowledgment

This work has been funded by Puglia Region (Italy), Project “High Optimization Patterns for operations Excellence (H.O.P.E)”.

References

Schulz, F., Nguyen, Q., Baetzner, A., Sjöberg, D., & Gyllencreutz, L. (2024). Exploring medical first responders' perceptions of mass casualty incident scenario training: a qualitative study on learning conditions and recommendations for improvement. BMJ Open. https://doi.org/10.1136/bmjopen-2024-084925

Zechner, O., Kleygrewe, L., Jaspaert, E., Schrom-Feiertag, H., Hutter, R.I.V., & Tscheligi, M. (2023). Enhancing Operational Police Training in High Stress Situations with Virtual Reality: Experiences, Tools and Guidelines. Multimodal Technologies and Interaction, 7(2), 14. https://doi.org/10.3390/mti7020014

Walenta, D., Hendriks, P., Jacquemin, P., & Gräf, M. (2026). New Hires, New Worlds: Metaverse Onboarding and Its Impact on Organizational Socialization. In Proceedings of the 59th Hawaii International Conference on System Sciences (p. 4611). https://doi.org/10.24251/HICSS.2026.550

Moore, N., Ahmadpour, N., Brown, M., Poronnik, P., & Davids, J. (2022). Designing Virtual Reality-Based Conversational Agents to Train Clinicians in Verbal De-escalation Skills: Exploratory Usability Study. JMIR Serious Games, 10(3), e38669. https://doi.org/10.2196/38669

Batra, K., Zhang, Z., Yang, S., Agrawal, A., Gu, Y., Benes, B., Magana, A., & Popescu, V. (2025). XRXL: A System for Immersive Visualization in Large Lectures. In 2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR) (pp. 370–380). IEEE. https://doi.org/10.1109/VR59515.2025.00061

Peixoto, B., Melo, M., Cabral, L., & Bessa, M. (2021). Evaluation of Animation and Lip-Sync of Avatars, and User Interaction in Immersive Virtual Reality Learning Environments. In 2021 International Conference on Graphics and Interaction (ICGI) (pp. 1–7). https://doi.org/10.1109/ICGI54032.2021.9655283

Training Without Moving Aircraft: A Modular Virtual Reality Platform for Airport Operational Teams

Álvaro Monzón, and Alejandro Castillo

DeuSens Hyperxperience, S.L., Zaragoza, Spain

Corresponding author: amonzon@deusens.com

Keywords: Virtual Reality training, haptic feedback, airport operations, industrial training, aviation safety, scalable XR deployment

Video-case: https://www.youtube.com/watch?v=fAO9FWT5zUQ

Abstract

Aena, the world's largest airport operator by passenger volume, launched Aena Ventures, an open innovation programme that received more than 500 proposals from 40 countries. DeuSens was selected to design a virtual reality (VR) training platform for Aena's operational teams — firefighters, ground marshallers and maintenance crews — replacing training methods that previously required moving real aircraft and relocating staff across the airport network. Following a six-month pilot and validation phase, the project was approved for a two-year deployment across Aena's network. The platform integrates TouchDIVER Pro haptic gloves (WEART) to deliver tactile feedback during interaction with 25 recreated aircraft models, including the ATR 72 and the Airbus A350, enabling trainees to locate and operate critical safety elements — doors, compartments, systems, access routes — under realistic conditions. To date, the platform has supported more than 675 training exercises, and the network-wide deployment is expected to train 2,640 professionals across Aena's operational teams. Its modular architecture allows new roles and procedures to be added without redesigning the system, changing the economics of recurring operational training. This paper presents the platform's design process, technical implementation, and expected operational impact, discussed as a case of scalable, reduced-logistics XR deployment for high-stakes industrial training; broader claims of hardware independence remain to be validated across headset models.

Introduction

Aena is the world's largest airport operator by passenger volume and a constituent of the Ibex 35 index. It manages 46 airports and two heliports in Spain and holds stakes in airport infrastructure in Brazil, the United Kingdom, Mexico, Colombia and Jamaica. Training operational staff — firefighters, ground marshallers, maintenance crews — traditionally required physically relocating personnel across the network and, in some cases, immobilising real aircraft for the duration of the exercise, with direct costs in logistics, aircraft downtime and travel-related emissions.

To address this, Aena launched Aena Ventures, an open innovation programme inviting the international startup ecosystem to reinvent airport operations through emerging technologies.

Challenge and Application Description

More than 500 startups from 40 countries entered Aena Ventures' second edition. DeuSens was selected to solve its most ambitious challenge, "Aeropuerto Virtual" (Virtual Airport): to design the virtual environment where the next generation of airport professionals would be trained. The proposal was not conceived as a one-off simulator but as a common platform on which Aena could build, for years to come, every training programme it might need. Aena approved the concept, funded a pilot, and, after a six-month validation period, opened a two-year deployment contract across its airport network.

The platform was designed with multiple operational roles in mind. Firefighters, ground marshallers and maintenance teams train on the same technological base, each with a dedicated module defining objectives, interactions and assessment criteria tailored to their professional profile.

Technical Implementation

Immersive training reaches a decisive step when it moves from visual to tactile interaction. TouchDIVER Pro haptic gloves, developed by WEART, were integrated into the platform so that trainees press buttons, hold tools, and perceive whether a surface is hot or cold with a response comparable to real conditions. The platform currently recreates 25 aircraft models — including the ATR 72 and the Airbus A350 — so that the virtual aircraft stops being an image and becomes an environment the trainee can physically interact with, matching the diversity of aircraft actually operated across Aena's network.

On that foundation, the platform reproduces the critical points that, in an emergency, must be located within seconds: doors, compartments, systems and access routes. Trainees repeat each procedure until it becomes an automatic response, so that when a real situation occurs — a rescue, a fire — the trained reflex determines the difference between arriving late and arriving on time.

The platform was developed iteratively together with Aena's operational staff during the six-month pilot phase, combining usability walkthroughs with functional testing of each training module ahead of the scale-up decision. This close collaboration with end users throughout the pilot, rather than a single end-of-project evaluation, is what allowed the platform's design to be refined before the two-year deployment was approved.

Expected Impact and Early Observations

Before this platform, training Aena's operational teams meant gathering professionals from across the network in one location and keeping a real aircraft unavailable for other use for the duration of the exercise. With the VR platform, the same training is designed to take place at any work centre, with a headset and a pair of haptic gloves: the intention is that people no longer travel to the training, but that the training travels to them. To date, the platform has supported more than 675 training exercises across its 25 recreated aircraft models, and the ongoing network-wide deployment is expected to train 2,640 professionals across Aena's operational teams. Based on early observations from the pilot phase, this is expected to translate into more teams trained in less time, more consistent assessments, contained logistics costs, and a reduction in emissions tied to mass staff travel. Beyond the exercise count and the projected trainee population reported above, these remain expected benefits rather than measured outcomes at this stage, as formal efficiency, cost and emissions data from the pilot are still being collected.

Limitations and Validation Needs

The main limitation of this work is that, at the time of writing, only scale figures — aircraft models recreated, exercises completed, professionals targeted for training — are measured; the efficiency, cost and emissions benefits of the platform are still described in terms of expected rather than measured outcomes, as formal quantitative results from the pilot phase remain under analysis. In

addition, the claim that the platform offers hardware-independent deployment has only been validated with the specific headset and haptic glove combination used in the pilot; extending this claim to other XR hardware configurations would require further testing. Safety and wellbeing aspects of extended VR and haptic use, together with a detailed account of the lessons learned during the pilot, were not the focus of this paper and are left for future work.

Conclusions and Future Work

This project represents, for DeuSens, the world's first virtual reality training platform built for an airport operator. Its modular, role-based architecture means that when Aena needs to train a new role or roll out a new operational procedure, it does not commission a project from scratch: it adds one more module to a living system, and the initial investment pays off with every new addition. Given Aena's international footprint, future work will explore extending the platform to additional operational roles and to airports within Aena's wider network. Future publications will report quantified pilot results and the acceptance criteria used to validate each training module, examine safety and wellbeing aspects of extended VR and haptic use, capture lessons learned from the pilot cohort, and provide further evidence supporting the platform's hardware-independence claims.

Platform in Use



Figure 1. *Photomontage of Aena's training platform, integrating the user into the project's virtual reality.*



Figure 2*. Trainee using Aena's training platform in real life.*

Platform Modularity by Role

Table 1. *Operational roles trained on the shared VR platform.*

Operational Role	Training Focus	Key Skills Reinforced
Firefighters	Emergency response inside real aircraft models (ATR 72, A350)	Door/compartment operation, haptic hazard recognition
Ground marshallers	Aircraft-side signalling and access procedures	Spatial awareness under time pressure
Maintenance crews	Technical systems inspection on virtual aircraft models	Precision, procedure sequencing under haptic feedback

References

Aena. (2026). Aena Ventures: results of the second edition open innovation programme. Available at: https://www.aena.es/en/press/aena-ventures-closes-the-second-edition-with-more-than-500-proposals---from-startups-from-all-over-the-world.html&p=1575086693589

DeuSens Hyperxperience, S.L. (2026). Immersive VR training for Aena's airport teams. Project case study. Zaragoza: DeuSens. Available at: https://deusens.com/en/portfolio/vr-training-aena

WEART S.r.l. (2025). TouchDIVER Pro haptic glove — technical specifications. Available at: https://weart.it/touchdiver-pro-haptic-gloves/

Virtual Reality as a Tool for Improving Interpretation of Machine Component Drawings with Consideration of Spatial Ability Differences

Fate S. Show, Simon Fletcher, Duke Gledhill, and Andrew Longstaff

School of Computing and Engineering, University of Huddersfield, United kingdom

Corresponfding authors:
fate.show@hud.ac.uk,
s.fletcher@hud.ac.uk,
d.gkedhill@hud.ac.uk,
a.p.longstaff@hud.ac.uk

Keywords: Virtual reality, spatial ability, manufacturing, technical drawings, engineering design

Abstract

Engineering students often lack the hands-on experience that is required for accurately designing machine components, while accounting for scale and end-user requirements (Prabhu, et al., 2025; Hung Chang, et al., 2020). Consequently, this is thought to contribute to an industry-wide skills gap (Baukal, 2024). This study was designed to observe the effects of virtual reality (VR) on spatial perception within real-world applications such as aerospace tool design and manufacturing. User engagement and perceived mental workload were also evaluated to assess participants' acceptance of VR as a learning method.

Introduction

Spatial ability encompasses a range of skills involved in mentally visualising, processing, and manipulating 3D objects and spaces. Individuals with lower spatial reasoning can experience difficulty interpreting 3D objects when they are presented in 2D formats, such as computer screens and paper-based drawings. This skill is essential for technical environments as it has shown to correlate strongly with academic achievement in STEM disciplines (Sorby & Veurink, 2019; Engineering UK, 2025; Engineering Council, 2020; EngineeringUK, 2022).

Research into spatial ability has reported gender differences in which males often outperform females on tests (Sorby & Veurink, 2019; Engineering UK, 2025; Engineering Council, 2020), and these gender differences may contribute to the wider engineering industry gender imbalance (Sorby & Veurink, 2019). Within the UK technology and engineering sector, women account for only 21% of degree graduates EngineeringUK, 2025) and 14–17% of careers (Engineering Council, 2020; EngineeringUK, 2022). Consequently, lack of spatial reasoning presents a significant challenge for mechanical engineering students (Dorrington et al., 2019), particularly when interpreting and understanding technical drawings of machine components.

Advancements in VR technology have enabled the development of immersive training environments that can target specialised skills that would traditionally require extensive experience and physical resources to develop. Spatial ability is a malleable skill that can be improved through training Hoffer (1977), and previous research has found comparable rates of improvement between genders (Uittal et al., 2012). This suggests that targeted training of spatial skills could potentially reduce gender disparities in engineering-related abilities.

Literature Review

CAD Programs in Engineering

Computer-Aided Design (CAD) programs are a standard tool in engineering design, providing a platform for visually developing concepts and creating precise designs for manufacture. However, effective use of CAD requires a strong spatial understanding, with users commonly experiencing difficulties in interpreting perspective and scale(Dorrington et al., 2019).

A study by Kösa and KarakuşKösa & Karakuş, 2018) examined the effects of CAD programs on spatial ability within an undergraduate engineering drawing course. They identified a strong relationship between spatial reasoning scores and performance in CAD training, with participants with higher spatial ability demonstrating better graphical understanding and course outcomes. Similarly, Onyancha et al. (2009) observed improvements in spatial ability following a CAD workshop, with targeted training having the greatest impact on participants with lower initial spatial ability. They attributed this pattern to a ceiling effect, where participants with higher initial spatial ability showed less improvement than those with lower spatial ability. This suggests that individuals with greater spatial skills may require fewer visual aids when solving spatial problems, providing preliminary support for the ability-as-compensator hypothesis.

The Effects of Colour on Cognitive Load

Cognitive load refers to the mental effort required to process information during a technical task. Two types of cognitive load (intrinsic and extraneous) can simultaneously contribute to overall mental demand. Extraneous cognitive load relates to how information is presented, including interface and question design. For example, complex wording may hinder task comprehension. In contrast, intrinsic load relates to the inherent difficulty of the material and subject specific knowledge (Sweller et al., 2011). Colour can be used for signalling important information by guiding the viewer's attention to specific details. By highlighting elements that require attention, signalling can reduce extraneous cognitive load, demonstrating positive outcomes for students completing STEM activities.

Waliullah and Ekman (2019) designed a mathematics task using colour coding to guide participants in answering questions. Their results indicated that colour coding improved both response accuracy and efficiency among schoolchildren aged 12–16. Koh et al. (2023) adopted a different approach with first-year medical students who had no prior anatomy experience. Comparing colour-coded anatomical cross-sections of organs against photorealistic models, they found that colour coding supported greater anatomical understanding without increasing the perceived mental load. Participants with lower spatial ability performed worse with the photorealistic models than with the colour-coded models, suggesting that colour coding may support the interpretation of 3D structures without increasing perceived cognitive demands.

VR as a Training Tool

Advancements in virtual reality (VR) technology have enabled the development of immersive training environments that can target specialised skills (such as understanding technical drawings) that would traditionally require extensive experience and physical resources to develop. A meta-analysis by Villena-Taranilla et.al. (2022) found that on average, interactive learning methods such as VR encourages higher success rates than purely theoretical learning. VR is widely agreed to be engaging which adds a learning incentive, with long-term VR intervention also providing increased benefits over some traditional methods (Pande et al., 2021).

Such a training program was developed by Chen et.al. (2011), who created an educational augmented reality (AR) based system for an engineering graphics course. This system utilised mobile phones that scanned unique markers assigned to machine parts on technical drawing documents. Once the code was scanned, the corresponding virtual 3D object would generate on the screen. These models aid the understanding of the students by adding additional visual information, which can be particularly helpful for those with lower spatial skills as it reduces their reliance on mental projections. Another similar AR mobile application was applied to a graphics training course. The scanned fiducial markers generated geometric models that targeted spatial ability training specifically. The results showed students preferring AR aids to traditional paper drawings with greater post-test gains in metal rotation and visualisation scores (Contero et al., 2012).

Research Gap

Factors such as cognitive load and learner engagement may influence motivation when undertaking complex subject matter. Therefore, the use of gamification (Krath et al.., 2021), alongside VR technology, may contribute to a more human-centred training platform by increasing motivation. The use of colour within technical drawings may also provide a visual guide to support the interpretation of complex object geometry. This approach may be particularly beneficial for students who experience greater difficulty interpreting technical drawings, as colour coding and the VR platform may provide additional spatial support without substantially increasing perceived cognitive load. Consequently, the combination of these techniques may result in improved engineering outcomes.

Methodology

Participants and Design

This pilot study explored the use of a VR-based environment designed to train technical drawing comprehension and the identification of defects in mis-manufactured components. The components used provide real-world context, as some of the models were real machine probing artefact designs made for an aerospace company. The aim of the study was to observe how VR aids spatial understanding for technical tasks related to engineering design and the willingness of technological adoption through positive engagement.

***Table 2**. Characteristics of Participants*

Characteristics	Control (PC)	Experimental (VR)
Sex: Female / Male	3 / 5	2 / 8
Department: Engineering / Computing	7 / 1	6 / 4
Age (years): Mean / Range	32 / 25 - 39	36.5 / 20 - 59

The study involved 18 participants (5 female and 13 male), all of whom were students or faculty members within the department of Computer Science and Engineering at the University of Huddersfield, UK. Three-quarters of participants had little to no prior experience with VR or technical drawings prior to the experiment. Informed consent was obtained, and the study received ethical approval. G*Power (v3.1.9.7) was used to estimate the required sample size for an independent-measures, two-tailed test, assuming a medium effect size (d = 0.5) and 80% statistical power. This indicated a required total sample of 128 participants. As the study was intended to provide preliminary findings and inform future research rather than establish a definitive effect of VR, the smaller sample of 18 participants was considered appropriate for this exploratory investigation.

Design

This three-part process was completed in one session lasting 30-60 minutes. The study began with the Mental Rotation Test (MRT), followed by the novel virtual environment referred to as the Engineering Skills Test (EST), and concluded with a qualitative feedback form that gathered a

workload assessment using the NASA Task Load Index (NASA-TLX). Prior to entering the EST, all participants were shown an example of a technical drawing alongside an explanation of the required terminology. Microsoft Form data was automated into a spreadsheet using Power Automate, and statistical analysis was completed using Microsoft Excel.

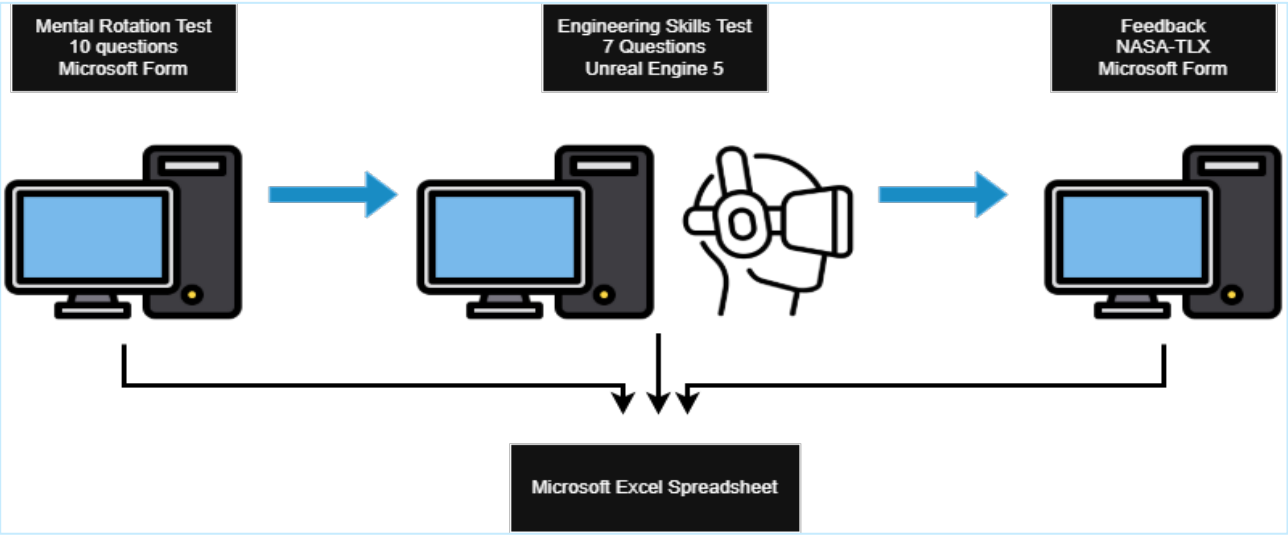


***Figure 3**. The experimental method.*

Mental Rotation Test

To determine initial spatial ability as a baseline for the experiment, the MRT was used. The original Vandenberg and Kuse MRT (1978) presents participants with a reference image and asks them to select two images that correspond to it. This study adapted the task into an odd-one-out format to align with the EST. The original MRT consists of 24 questions with a 10-minute time limit; however, the adapted MRT contained 10 questions with no time limit. The decision to remove the time limit was intended to reduce potential gender differences in responses due to time pressure. As females have been found to respond more negatively to countdowns, whereas males may become more competitive (Stoevenbelt et al., 2023). Therefore, to maintain a similar overall duration to the original MRT, the number of questions was reduced from 24 to 10, resulting in an average completion time of just under 9 minutes. Furthermore, as the EST session followed the MRT, a prolonged MRT duration could have resulted in participant fatigue and potentially influenced subsequent EST performance.

Engineering Skills Test

The training environment was developed using Unreal Engine for deployment on both the Meta Quest Pro headset and a traditional computer-based setup. The environmental design was identical between conditions however, the traditional control condition represented CAD programs commonly used by engineers, whilst the experimental VR condition used Newtonian physics and was comparable to a real manufacturing scenario. The gamified EST software, incorporated data collection systems to monitor individual performance. These performance metrics were then analysed alongside MRT scores to investigate the relationship between performance in the EST and participants' spatial ability.

Seven EST questions were included in total, with four presenting technical drawings in monochromatic black and white and three incorporating colour. This enabled comparison of performance and participant preference between the two conditions. Four colours were selected based on Four Colour Theory (Rogers, 2011), comprising the three primary colours, which provide distinct chromatic separation, and brown as a neutral fourth colour produced by mixing the primary colours. A similar colour-coding approach was also used by Waliullah and Ekman (2019) to investigate the effects of colour on cognitive load during mathematics tasks.

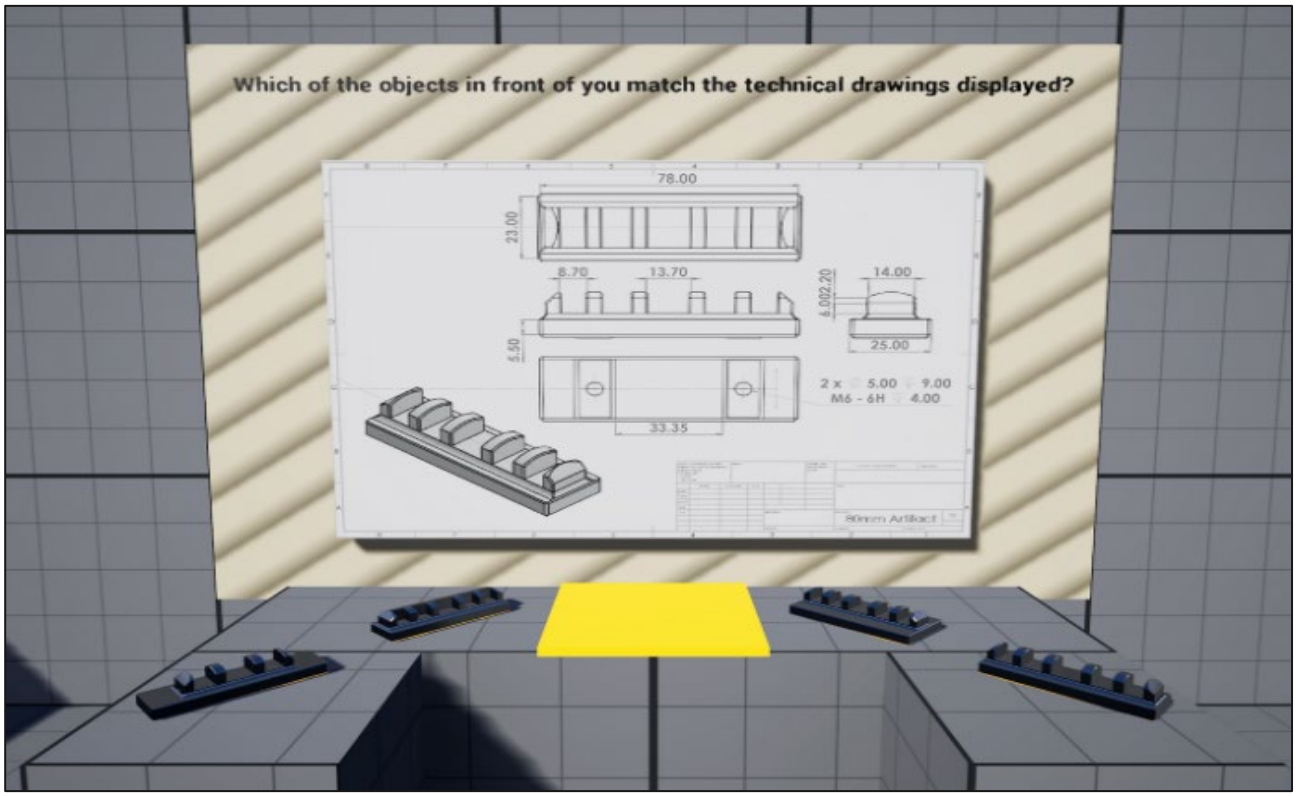


***Figure 4**. User's view of the training environment from the PC.*

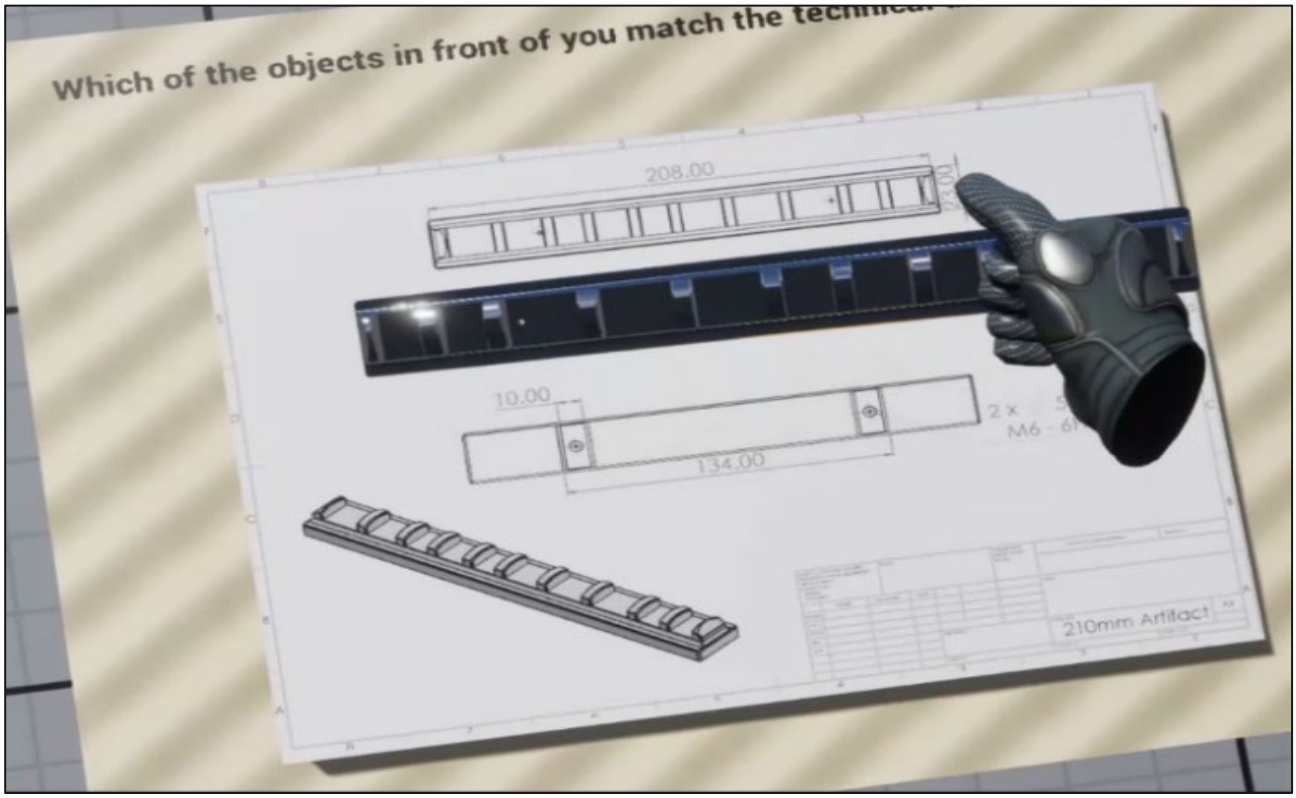


***Figure 5**.User's view of the training environment from the VR headset.*

Colour is commonly used in mechanical engineering drawings to identify materials or distinguish systems within a machine. However, established drawing standards restrict the use of colour in technical drawings, partly because printing coloured drawings in black and white can reduce visual clarity. Therefore, colour was included in this study as an exploratory feature rather than a representation of standard engineering practice, examining whether it could provide an intermediary aid for students developing technical drawing interpretation skills.

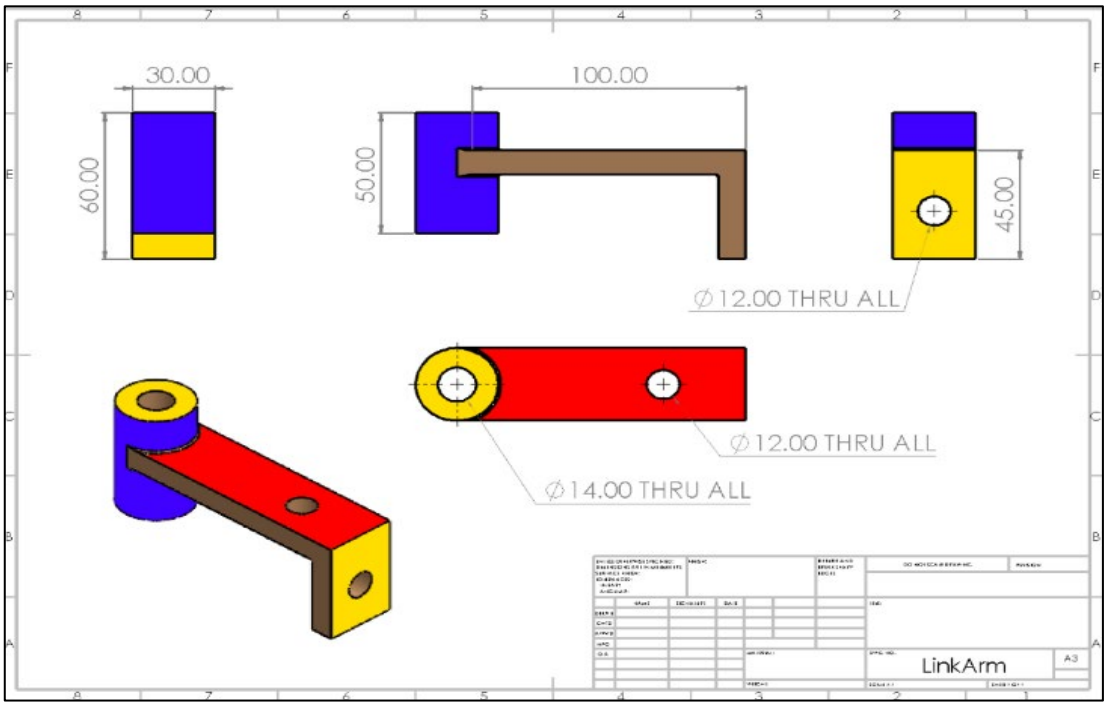


Figure 6. *Coloured Drawing Extract*

Method

The MRT models were generated using Blender (v5.0.0), with shadows disabled and lighting set to 'flat' to replicate the visual appearance of the original Vandenberg and Kuse MRT. The test was implemented using Microsoft Forms, with each question worth 10 points.

The machine component models were created in SolidWorks. To approximate the complexity of each 3D model, a crystallographic symmetry categorisation technique was applied, whereby objects exhibiting fewer symmetries were considered more complex. SolidWorks models are constructed using the Front, Top, and Right reference planes, which were used as the basis for identifying and counting the symmetries present in each object.

Workload was assessed using the NASA-TLX, a commonly used measure in VR-based experiments(Dorrington et al., 2019; Koh et al., 2023; Contero et al., 2012). Participants rated workload from Very Low to Very High, alongside a separate Likert-scale assessment of engagement. Demographic information and qualitative feedback were also collected at the end through a semi-structured interview.

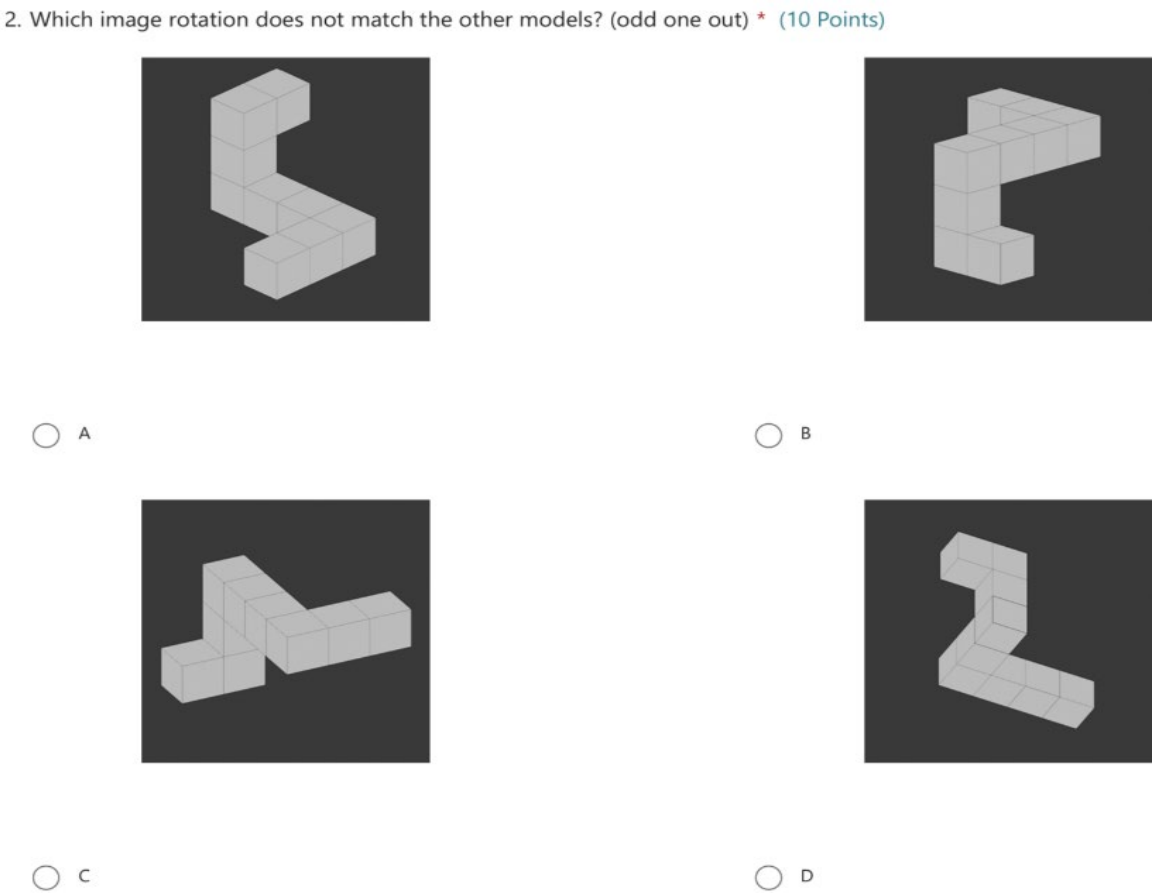


***Figure 7**. A sample MRT question.*

Results

MRT Results

The participants achieved a mean score corresponding to the 70th percentile, which is above the population median of the 50th percentile. This met the expected range as adults pursuing technology and engineering careers have been reported to demonstrate above-average spatial reasoning ability (Gold et al., 2018).

Initial Spatial Ability

An independent-samples, two-tailed Student's *t*-test was used to compare MRT scores between conditions. This approach was consistent with similar studies observing spatial ability (Koh et al., 2023; Contero et al., 2012; Oren et al., 2012).The null hypothesis was stated as H_0: there is no significant difference in initial spatial ability between the conditions. A significance level of $\alpha = 0.05$ was used for all statistical analyses.

The analysis found no statistically significant difference in initial MRT performance between the two groups ($p = 0.08$). The calculated *t*-value of 1.89 was below the critical value of 2.12, further indicating that the result was not statistically significant. Therefore, the null hypothesis was not rejected, suggesting that the groups did not differ significantly in their initial spatial ability.

MRT Question Difficulty

MRT question difficulty was analysed using the Item Difficulty Index (IDI), calculated as the number of correct responses divided by the total number of participants. The IDI was converted to a percentage for readability and can be understood as how often a question was answered correctly. The questions appeared to fall into three difficulty categories: difficult (40–60%), moderate (60–80%), and easier (80–100%). Of the 10 questions, three were classified as difficult, five as moderate, and two as easier.

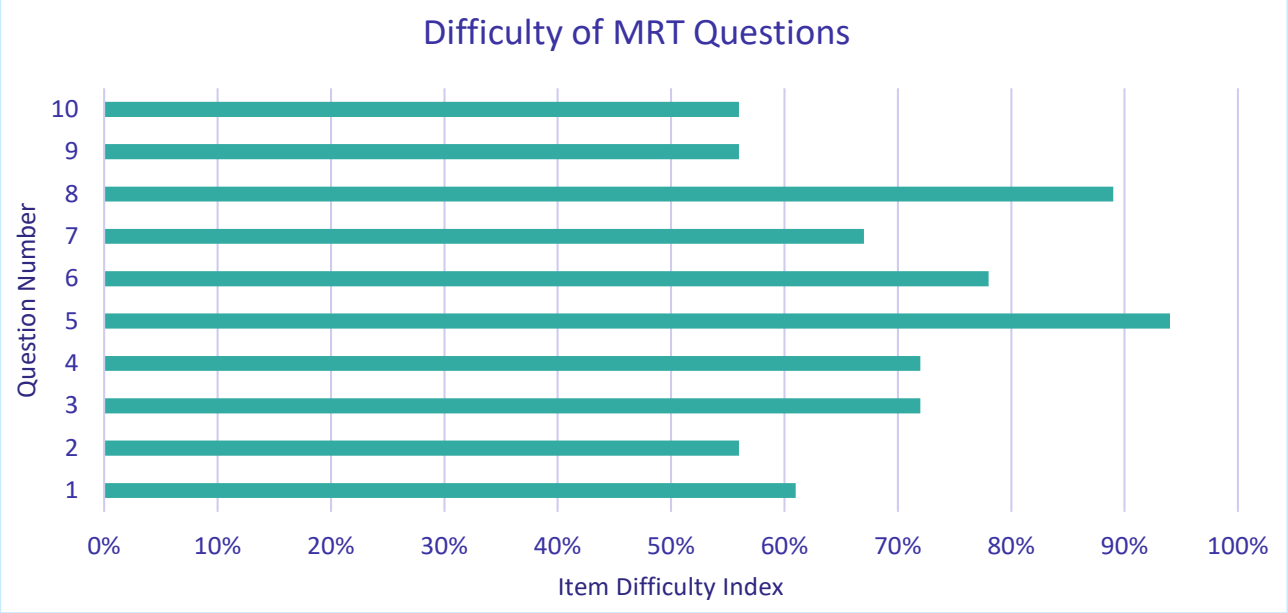


***Figure 8**. MRT Question Difficulty*

EST Results

Was There Better Spatial Understanding in the VR Condition?

Previous research examining the effects of VR on spatial ability has reported predominantly positive outcomes; therefore, a similar effect was anticipated in the present study. An independent one-tailed *t*-test was conducted to determine whether the difference in scores between the MRT and EST was significantly greater using VR training than traditional training. The null hypothesis was stated as H_0: there was no significant difference in spatial understanding between the two conditions, while the alternative hypothesis stated H_1: spatial understanding would be greater in the VR environment.

After scaling the EST scores to correspond with the 100-point MRT scale, the VR condition demonstrated a mean difference of 16.57 points (*SD* = 21.01) between MRT and EST scores, compared with a mean difference of 0.36 points (*SD* = 12.20) in the traditional condition. The analysis indicated that this difference was statistically significant (p = 0.036), supporting the alternative hypothesis that participants in the VR condition demonstrated a greater difference between their MRT and EST performance than participants in the traditional condition. Additionally, female participants achieved an average score difference of 42.86 points between the MRT and EST, compared with a 1.90-point decrease in the traditional condition.

Did Initial Spatial Ability Affect EST Task Performance?

Pearson's correlation coefficients were calculated to examine the relationship between participants' pre-existing spatial ability and their task performance in each condition. A stronger positive correlation was observed in the PC condition (r = 0.73), compared to a weaker positive correlation (r = 0.46) in VR.

The regression analysis indicated a stronger association between spatial ability and EST performance in the PC condition (r^2 = 0.53) than in the VR condition (r^2 = 0.22). Furthermore, MRT score was a statistically significant predictor of EST performance in the PC condition (p = 0.04), but not in the VR condition (p = 0.18).

Which Condition Facilitated Better Efficiency?

Participants in the VR condition completed the exercise 22% faster on average than those in the traditional condition; however, this difference was not statistically significant (p = 0.26, *SD* = 182.28). The use of colour in the drawings (Q5–7) was also associated with shorter response times among PC participants, who answered these questions an average of 26.58 seconds faster than the greyscale questions. However, this difference was not statistically significant (p = 0.85, *SD* = 27.46)

therefore, the observed differences in completion time cannot be attributed to the use of colour with statistical confidence.

EST Question Difficulty

To assess the difficulty of each question, the IDI was first calculated. This value was then used to calculate the Inverse Efficiency Score (IES), where difficulty was defined as the mean reaction time divided by accuracy.

For the greyscale drawings, the participants found that the test questions were easier in the VR condition, with an anomalous result observed for Question 3, where PC participants found the question over three times as difficult. Conversely, participants found the colour drawings easier in the PC condition.

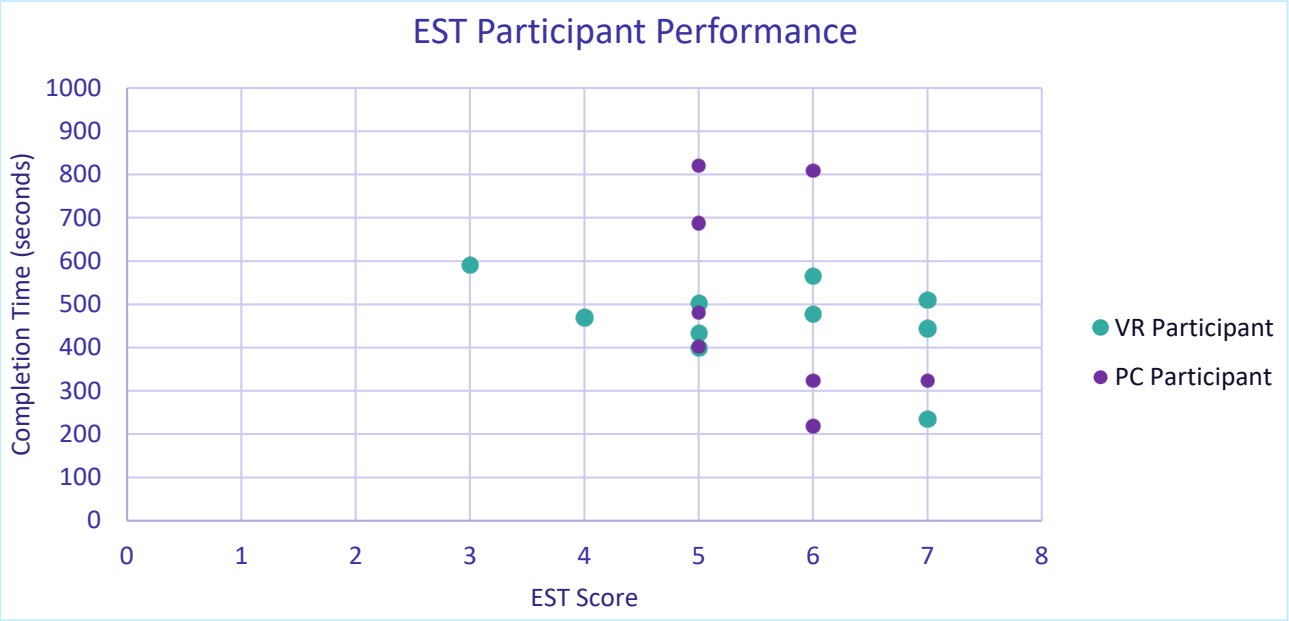


***Figure 9**. EST Participant Performance.*

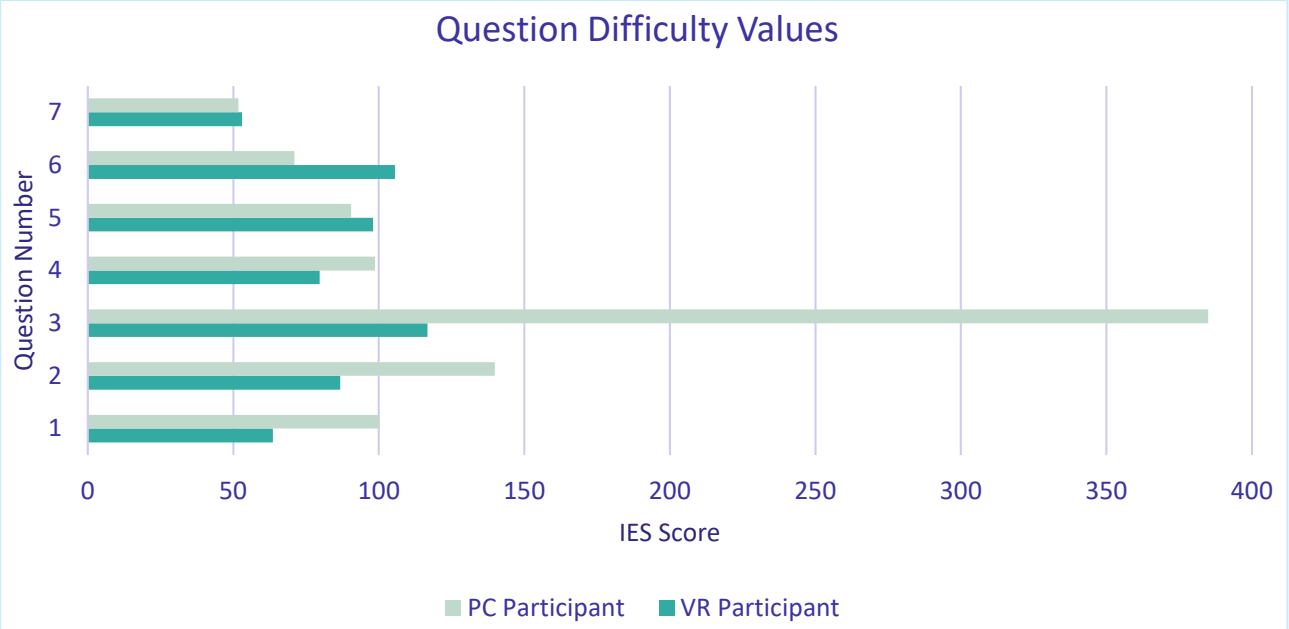


***Figure 10**. EST Question Difficulty Values.*

Second Round Results

A second EST trial was conducted immediately after the first, following an AB–BA design in which participants who completed the first trial in one condition subsequently experienced the other condition. However, data from the second trial were excluded from the main statistical analysis because the questions were identical to those in the first trial, potentially introducing a recall effect. Consequently, the second-round results were not used as evidence of improved performance.

Observations from the second trials showed a statistically significant reduction in completion time between the first and second rounds in both the VR condition ($p = 0.03$) and PC condition ($p = 0.04$). Interestingly, the reduction in completion time was not consistently accompanied by improved accuracy. Participants who initially completed the task in VR showed an 11% reduction in accuracy when subsequently completing the PC condition, whereas participants who initially completed the task in PC showed an 11% increase in accuracy when subsequently completing the VR condition.

Workload Results

Mental workload scores were comparable between the two conditions, with the VR condition producing slightly more favourable results within the frustration and performance categories. This suggests that VR did not substantially increase perceived cognitive workload within the presented study.

Engagement was assessed using three questions addressing enjoyment, engagement, and willingness to reuse the software, scored on a 1–6 scale from “Very Strongly Disagree” to “Very Strongly Agree”. The VR condition achieved a mean score of 5.6 for each question, compared with 3.75–5.13 for the traditional condition. All 16 participants who experienced both conditions also expressed a preference for VR.

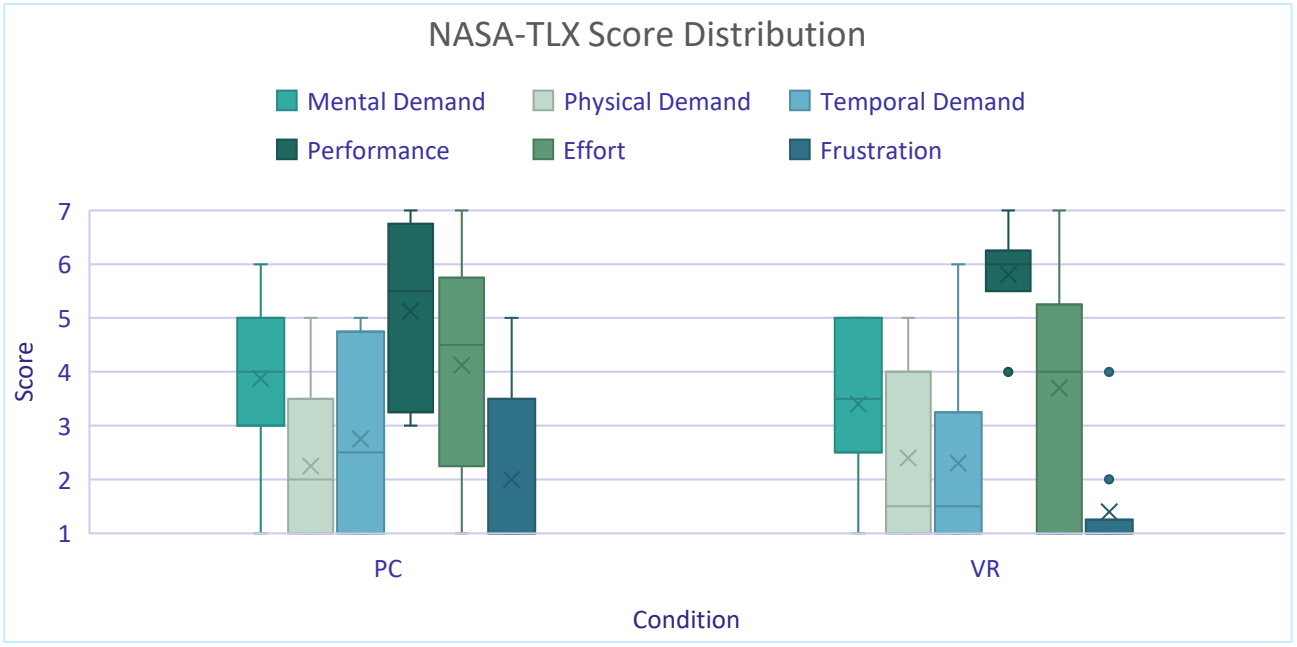


Figure 11. *NASA-TLX Score Distribution*

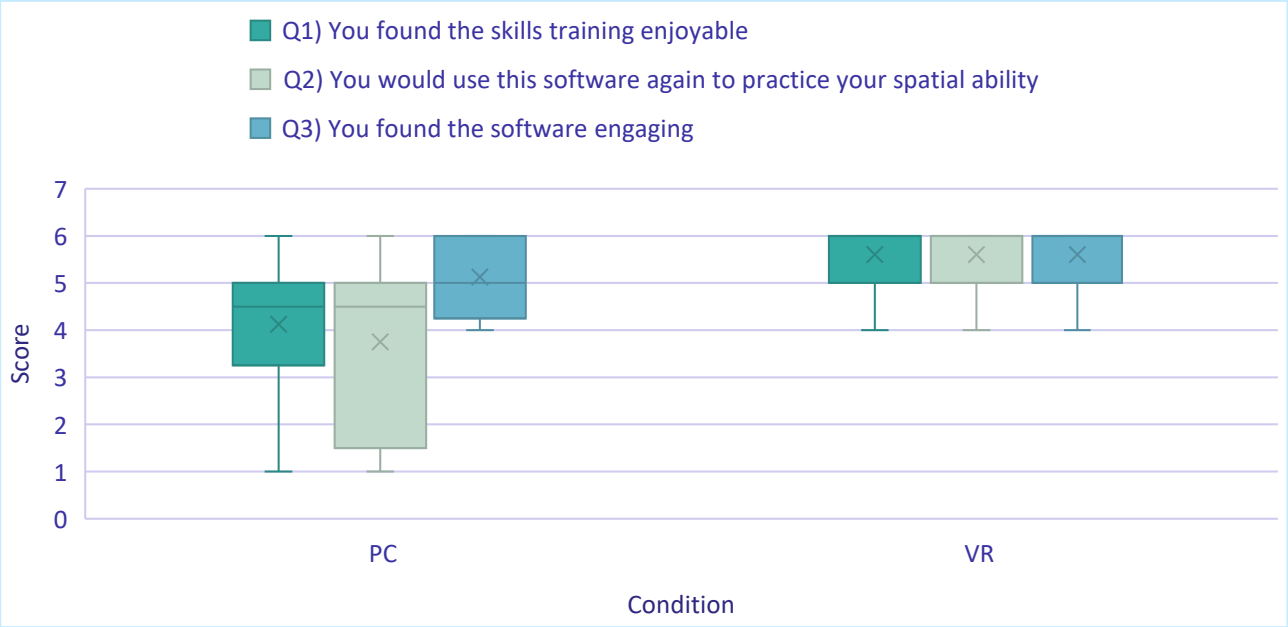


***Figure 12**. Platform Engagement Feedback*

Usability and Strategic Reasoning

Despite some participants having prior experience using CAD programs, VR was still perceived as more intuitive and more closely representative of a real-world environment. Participants with manufacturing experience described the VR training exercise as comparable to a real-world scenario. Some participants commented that some CAD programs were easier to use than the study's PC interface however, the VR interface was still much easier to use than most PC CAD programs, with less of a skill barrier.

Across both conditions, participants employed similar strategies, including comparing thru holes, chamfers, and component scales. Participants with manufacturing experience tended to rely more heavily on the technical drawings, whereas those without prior manufacturing experience would first focus on the isometric model images before using the drawings for additional details. Despite these similarities, the VR condition appeared to support a wider range of strategies. For example, participants could pick up two models simultaneously, compare them directly, and place them aside in a specific order to support their reasoning. In contrast, models in the PC condition returned to their original positions after being moved, limiting participants' ability to arrange models spatially for comparison.

Discussion

Reliance on Initial Spatial Ability

The correlation between initial spatial ability and EST task performance was stronger in the traditional PC condition (r = 0.73) than in the experimental VR condition (r = 0.46). Similarly, the regression analysis indicated that MRT performance accounted for approximately 53% of the variance in EST scores in the PC condition, compared with approximately 22% in the VR condition. This suggests that the relationship between initial spatial ability and EST performance was weaker in the VR condition. The observed correlation in the VR condition is comparable to that reported by Alpiste et al. (2019), who identified a similar positive correlation (r = 0.46, p < 0.001) when investigating the effect of a VR engineering graphics course on spatial ability. This consistency with previous research provides further support for the observed reduced reliance on pre-existing spatial ability on performance in VR-based engineering tasks. However, the small sample size limits the strength of this interpretation and its generalisability.

EST Efficiency and Question Difficulty

Participants in the VR condition completed the exercise 22% faster on average than those in the traditional condition. This observed trend towards shorter completion times contrasts with previous research indicating that VR can increase task completion times (Oren et al., 2012; Mullen & Davidenko, 2021). Although some participants reported a reduced perception of time while using VR, the shorter completion times observed in the present study may be associated more with the interface, which participants perceived as more intuitive.

For the greyscale drawings, the participants found the test questions easier in the VR condition, with those accustomed to technical drawings generally having a preference for the traditional greyscale. An anomalous result observed for Question 3, where PC participants found the question over three times as difficult. This may have been due to the small details required to solve the question, which may have been more difficult to navigate or visualise using the computer interface. Furthermore, Questions 3 and 4 had no planes of symmetry, which may have also contributed to their higher difficulty.

Conversely, several participants (including some with previous technical drawing experience) stated finding the colour drawings easier in the PC condition and this was supported by question difficulty and completion time data. However, there was little effect of colour on performance in the VR condition. This may suggest a ceiling effect, whereby colour provided an additional visual aid for the PC interface and potentially reduced the extraneous cognitive load. In contrast, the VR interface may have already provided sufficient spatial reasoning support for interpreting the drawings, meaning that the additional use of colour had less influence on performance.

Performance of Female Participants

Another observation found that female participants appeared to benefit more from the VR environment than the males, as they received a more substantial score difference. This aligns with McLaren-Gradinaru et al.(mcLaren-Gradinaru et al., 2023), who found that gender differences in spatial ability are influenced by video game experience, with female participants potentially achieving comparable or superior performance to males through gaming. Additionally, females were reported to prefer puzzle games in comparison to males who preferred action and navigational games. Therefore, the puzzle-based nature of the EST may have also contributed to its greater positive effect among female participants. However, this finding is only a preliminary insight and places emphasis on the need for a larger population sample.

Engagement and Cognitive Load

Both conditions in this experiment were presented in a gamified format; however, the VR condition demonstrated higher levels of engagement and was associated with a more positive learning experience. All 16 participants who experienced both conditions expressed a preference for VR. These findings show a stronger and more consistent preference for VR, despite that the traditional PC platform was also perceived as engaging.

These findings build on the predictive models proposed by Martín-Gutiérrez and González (2016), which suggest that the use of 3D tools can enhance student attention and motivation, while the use of videogames may provide additional benefits for spatial ability. The findings of the present study suggest that VR may further enhance these effects by combining an immersive 3D learning environment with interactive and gamified elements, potentially supporting both motivation and spatial understanding. Accordingly, the proposed revised model incorporates that VR as a learning platform may further enhance the positive effects of motivation and spatial reasoning.

Additionally, VR did not substantially increase perceived cognitive workload within the presented study. This finding is consistent with Koh et al. (2023), who similarly reported improvements in spatial ability during a 3D anatomy modelling task despite similar levels of cognitive load between conditions. Together, these findings suggest that improvements in spatial performance can occur without a corresponding increase in perceived cognitive workload.

Ability-as-Compensator Hypothesis

Participants in the VR condition began with a lower average MRT score but achieved comparable EST accuracy to those in the traditional PC condition. Despite the lower baseline performance, the VR group demonstrated a significantly larger difference between their MRT and EST scores. In contrast, the difference between MRT and EST performance was comparable in the PC condition. The larger change observed in the VR condition is therefore notable, as it suggests that participants were able to achieve comparable technical drawing performance despite starting with lower average spatial ability.

Compiling the results from the question accuracy, time, and difficulty alongside correlation and regression analysis, the PC participant performance appeared more strongly associated with pre-existing spatial ability. In contrast, the weaker relationship observed in VR suggests that VR performance may be less dependent on participants' initial spatial ability. This pattern was also found by Höffler and Leutner (1977) when observing the effects of animation versus static images for a chemistry task. Additionally, the results from the second trial show equal levels of improvement or regression upon entering the alternate condition. This may suggest that the PC participants who began the study with a higher spatial ability also found the VR platform a useful aid. The study findings provide preliminary support for the ability-as-compensator hypothesis (Lin & Suh, 2021), suggesting that VR may serve as an effective tool for enhancing the structural comprehension of technical drawings, particularly among individuals with lower spatial ability.

Conclusion

The findings suggest that the proposed VR training environment was comparable to a realistic manufacturing scenario such as when a machine operator inspects a product that has finished its machining procedure and been removed from the machine. Additionally, the results indicate that VR can effectively support the development of spatial and interpretative skills relevant to engineering design. Better spatial reasoning with higher observation of defects was apparent for individuals with varying spatial skills although, there was a more noticeable difference for those who started the study with lesser spatial ability.

CAD is a widely used tool in engineering design and has been associated with positive effects on spatial ability. The integration of VR into CAD and 3D design environments may potentially provide more intuitive and accessible methods of interacting with 3D models and help reduce some barriers associated with gender and the development these skills. Industry adoption of VR within 3D design is emerging, with companies such as Siemens and Autodesk incorporating VR capabilities into aspects of their design workflows, indicating growing interest in the application of immersive technologies. However, given the limited sample size, the study findings should be interpreted with caution, and further research involving larger and more diverse samples is required to establish the robustness and generalisability of these effects.

Limitations and Future Research

The smaller sample size limited the statistical analyses that could be conducted and the generalisability of the findings. However, the aim of this study was to provide evidence to inform future research by examining the relationship between spatial ability and theoretical engineering skills within a practical, immersive VR environment. These findings will inform planned future research investigating the application of technical drawings within a VR-based machine operation case study.

Acknowledgments

The authors would like to acknowledge the support from the EPSRC The Future Advanced Metrology Hub for Sustainable Manufacturing (Ref. EP/Z53285X/1) and UKRI Advanced Machinery & Productivity Initiative (Ref. 84646) for the funding for this work.

Disclosure of Interests

The author has no competing interests to declare that are relevant to the content of this article.

References

Alpiste Penalba, F., Torner Ribé, J., & Brigos Hermida, M. Á. (2019). Exploring virtual reality to improve engineering students' spatial abilities pilot study. In *Edulearn19 proceedings: 1st-3rd July 2019: 11th International Conference on Education and Learning Technologies* (pp. 6275-6284). https://hdl.handle.net/2117/168208

Baukal, C. E. (2024, June). Hands-On Activity for First-Year Engineering Students. In *2024 ASEE Annual Conference & Exposition*. https://doi.org/10.18260/1-2-47520

Chen, H., Feng, K., Mo, C., Cheng, S., Guo, Z., & Huang, Y. (2011, December). Application of augmented reality in engineering graphics education. *In 2011 IEEE international symposium on IT in medicine and education* (Vol. 2, pp. 362-365). IEEE. https://doi.org/10.1109/ITiME.2011.6132125

Contero, M., Gomis, J. M., Naya, F., Albert, F., & Martin-Gutierrez, J. (2012, October). Development of an augmented reality based remedial course to improve the spatial ability of engineering students. In 2012 *Frontiers in education conference proceedings* (pp. 1-5). IEEE. https://doi.org/10.1109/FIE.2012.6462312

Dorrington, P., Harrison, W., Brown, H., Holmes, M., & Kerton, R. (2019). Step away from the CAD station: A hands-on and immersive approach to second year teaching of Mechanical Engineering design. *In Proceedings of the Virtual and Augmented Reality to Enhance Learning and Teaching in Higher Education Conference 2018, Swansea*

EngineeringUK, “Graduate outcomes - engineering and technology,” Engineering UK, London, 2025.

Engineering Council, “Mapping the UK's Engineering Workforce,” MERCATOR, 2020

EngineeringUK, “Gender disparity in early perceptions of engineering Findings from the Engineering Brand,” 2022

Gold, A. U., Pendergast, P. M., Ormand, C. J., Budd, D. A., Stempien, J. A., Mueller, K. J., & Kravitz, K. A. (2018). Spatial skills in undergraduate students—Influence of gender, motivation, academic training, and childhood play. Geosphere, 14(2), 668-683. https://doi.org/10.1130/GES01494.1

Halpern, D. F., & Collaer, M. L. (2005). Sex differences in visuospatial abilities. *The Cambridge handbook of visuospatial thinking*, 170-212. https://books.google.fi/books?hl=fi&lr=&id=m91B8zm_1qgC&oi=fnd&pg=PA170&dq=Sex+Differences+in+Visuospatial+Abilities&ots=kqrbVwmyti&sig=kSAXM74sUdbEFnUM2hC_v6HoT6Q&redir_esc=y#v=onepage&q=Sex%20Differences%20in%20Visuospatial%20Abilities&f=false

Hoffer, A. (1977). Mathematics Resource Project: Geometry and visualization, *Creative Publications*

Khan, U. (22 06 2026) Models. [Online]. Available: https://grabcad.com/uzair.khan-49/models

Koh, M. Y., Tan, G. J. S., & Mogali, S. R. (2023). Spatial ability and 3D model colour-coding affect anatomy performance: a cross-sectional and randomized trial. *Scientific reports*, 13(1), 7879. https://www.nature.com/articles/s41598-023-35046-2

Krath, J., Schürmann, L., & Von Korflesch, H. F. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in human behavior*, *125*, 106963. https://doi.org/10.1016/j.chb.2021.106963

Kösa, T., & Karakuş, F. (2018). The effects of computer-aided design software on engineering students’ spatial visualisation skills. *European Journal of Engineering Education*, 43(2), 296-308. https://doi.org/10.1080/03043797.2017.1370578

Lin, H. C., Chang, Y. S., & Li, W. H. (2020). Effects of a virtual reality teaching application on engineering design creativity of boys and girls. Thinking Skills and Creativity, 37, *Article 100705*. https://doi.org/10.1016/j.tsc.2020.100705

Lin, Y., & Suh, A. (2021). The role of spatial ability in learning with virtual reality: a literature review. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/4dadd0ed-8fe0-4974-9db7-8e3bca2c2586/contentOnyancha, R. M., Derov, M., & Kinsey, B. L. (2009). Improvements in spatial ability

as a result of targeted training and computer-aided design software use: Analyses of object geometries and rotation types. *Journal of Engineering Education*, 98(2), 157-167. https://doi.org/10.1002/j.2168-9830.2009.tb01014.x

Martín-Gutiérrez, J., & González, M. M. A. (2016). Ranking and predicting results for different training activities to develop spatial abilities. In *Visual-spatial ability in STEM education: transforming research into practice* (pp. 225-239). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-44385-0_11

McLaren-Gradinaru, M., Burles, F., Protzner, A. B., & Iaria, G. (2023). The cognitive effects of playing video games with a navigational component. *Telematics and Informatics Reports*, *9*, 100043. https://www.sciencedirect.com/science/article/pii/S2772503023000038

Mullen, G., & Davidenko, N. (2021). Time compression in virtual reality. *Timing & Time Perception*, *9*(4), 377-392. https://brill.com/view/journals/time/9/4/article-p377_377.xml

Oren, M., Carlson, P., Gilbert, S., & Vance, J. M. (2012, March). Puzzle assembly training: Real world vs. virtual environment. *In 2012 IEEE Virtual Reality Workshops (VRW)* (pp. 27-30). https://doi.org/10.1109/VR.2012.6180873

Pande, P., Thit, A., Sørensen, A. E., Mojsoska, B., Moeller, M. E., & Jepsen, P. M. (2021). Long-term effectiveness of immersive VR simulations in undergraduate science learning: lessons from a media-comparison study. *Research in Learning Technology*, 29. https://eric.ed.gov/?id=EJ1293535

Prabhu, R., Mendonca, S., Bellairu, P. K., & Shiri, N. D. (2025). Enhancing engineering education through mini project-based learning in computer integrated manufacturing laboratory: A student-centric approach. *Innovations in education and teaching international*, 62(3), 971-985. https://doi.org/10.1080/14703297.2024.2362260

Reilly, D., Neumann, D. L., & Andrews, G. (2016). Gender differences in spatial ability: Implications for STEM education and approaches to reducing the gender gap for parents and educators. In *Visual-spatial ability in STEM education: Transforming research into practice* (pp. 195-224). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-44385-0_10

Rogers, L. (2011) .The Four Colour Theorem. *University of Cambridge*, 01 02 2011. [Online]. Available: https://nrich.maths.org/articles/four-colour-theorem. [Accessed 24 08 2026].

Sorby, S. A., & Veurink, N. (2019). Preparing for STEM: Impact of spatial visualization training on middle school math performance. *Journal of Women and Minorities in Science and Engineering*, 25(1). https://doi.org/10.1615/JWomenMinorScienEng.2018024516

Stoevenbelt, A. H., Wicherts, J. M., Flore, P. C., Phillips, L. A., Pietschnig, J., Verschuere, B., ... & Schwabe, I. (2023). Are speeded tests unfair? Modeling the impact of time limits on the gender gap in mathematics. *Educational and Psychological Measurement*, *83*(4), 684-709. https://doi.org/10.1177/0013164422111107

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, 22(2), 123-138. https://link.springer.com/article/10.1007/s10648-010-9128-5

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological bulletin*, 139(2), 352. https://doi.org/10.1037/a0028446

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills*, 47(2), 599-604. https://doi.org/10.2466/pms.1978.47.2.599

Villena-Taranilla, R., Tirado-Olivares, S., Cózar-Gutiérrez, R., & González-Calero, J. A. (2022). Effects of virtual reality on learning outcomes in K-6 education: A meta-analysis. *Educational Research Review*, 35, 100434. https://doi.org/10.1016/j.edurev.2022.100434

Waliullah, F. I., & Ekman, H. (2019). Reducing cognitive load on the working memory by signaling primed colors, can color improve the learning of mathematics.

Large-scale Design of VR-based Trainings for E-Mobility at Robert Bosch GmbH: Method and Experiences

Marvin Schobert[1], David Kunz[1], Andreas Bayer[1], Jörg Franke[1], Jens Fürst[1], Robert Köhler[2], Sven Haschke[3], and Christoph Runde[3]

[1]*Institute for Factory Automation and Production Systems, FAU Erlangen-Nürnberg*

[2]*Robert Bosch GmbH*

[3]*Virtual Dimension Center w.V.*

[*]Corresponding author: marvin.schobert@faps.fau.de

Keywords: Virtual Reality, Vocational Training, E-Mobility, Authoring Tools, Design

Introduction

Large-scale workforce upskilling challenges industrial organizations when training demand increases while expert trainer capacity remains limited. The XR-Upskill research project (www.xr-upskill.de) explores virtual reality (VR) as a scalable approach to improve and accelerate employee qualification and vocational trainings. Funded by the German Federal Institute for Vocational Education and Training (BIBB), the project addresses two e-mobility use cases: a nine-hour fundamentals training delivered more than 200 times annually and increasingly shifted towards asynchronous learning (UC1), and a VR-based refresher training for the electrical safety enabling procedure of electric vehicles (UC2). A key aspect of the approach is trainer-driven content creation. Rather than developing individual VR applications, domain trainers use authoring tools to independently design and maintain immersive learning content, similar to creating presentation material. This paper presents the methodology established at Robert Bosch GmbH at Schwieberdingen (Germany) for designing VR training at scale, including the underlying design process, authoring approach, and practical experiences from its application. Its main contribution is a practitioner-oriented methodology that enables trainers to create and maintain large-scale VR training without software development expertise.

Context and Methodology

“Introduction to Electromobility” is an education and training program for apprentices, students, and professional employees, providing a general introduction to key aspects and technologies of e-mobility. The program itself has reached approximately 100–150 participants annually, with related courses bringing the total number of participants to more than 500 per year. The program consists of three three-hour units in a predominantly instructor-led format, of which units 1.1 (Basics) and 1.3 (Power electronics) were selected for the XR-Upskill project. The integration of XR into training requires a combination of organizational, technical, and didactic competencies. The XR-Upskill framework distinguishes eleven roles across the XR-supporting training process, covering the overall process from project management and technical administration to subject matter expertise, 3D asset

preparation, training authoring, learning facilitation, and learning. This paper focuses on the activities and efforts of the core roles directly involved in content creation: training authors, 3D model creators, and trainers. The development process practiced follows the ADDIE model (Sturman et al. 2025) and builds on the existing course structure and materials (cf. **Figure 1**). As "Introduction to Electromobility" is an established instructor-led program, key elements of the Analysis and Design phases were already available. This includes defined learning objectives, characteristics of target learners, selected learning theories, and an initial storyboard. Existing course materials, such as presentations and documents, therefore provide both content references and a starting point for the XR adaptation. Selected employees were further upskilled to 3D model creation.

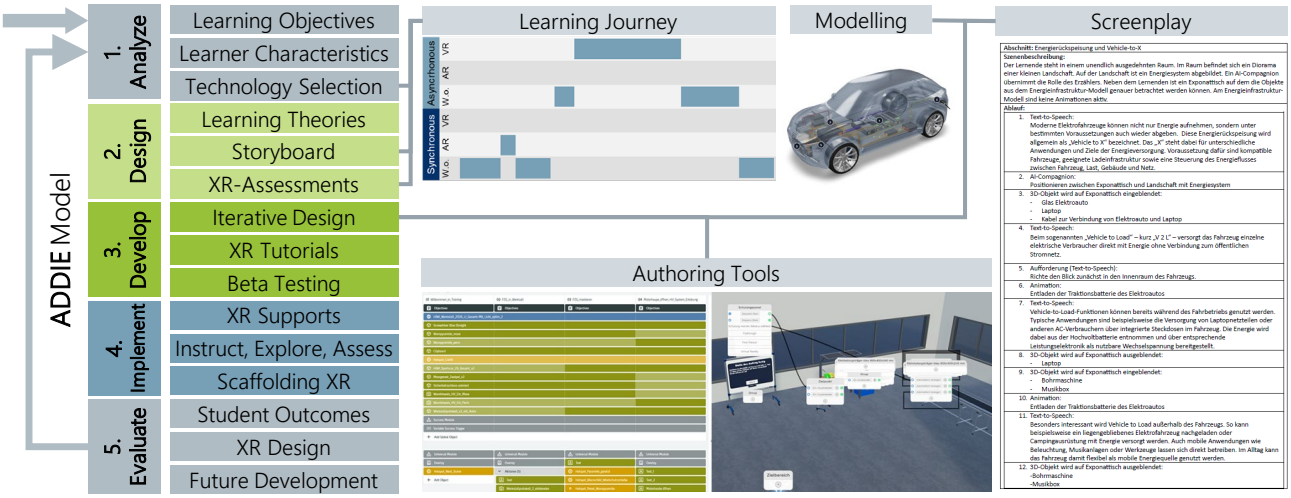


Figure 13. *ADDIE-Model according to Sturman et al. (2026) with selected tools from the XR-Upskill project*

Practical Guidelines for Analysis, Design and Development

The following steps assume that the decision to create a VR-based training has already been made. We recommend designing each learning unit for a final runtime of approximately 15 minutes to limit motion sickness and loss of attention. The method aims to enable the core team to create and evaluate each unit with less than 25 person-hours, including 3D modeling. These figures are rough estimates and depend strongly on the complexity of the unit, the models and the team's skills. Our experience shows that creation effort decreases substantially with increasing content reuse and team experience, with the rate of reduction approaching an asymptotic limit, but also increases with higher complexity or information density.

Step 1. Define Learning Objectives and Target Group and plan roughly (1 hour)

The trainer and project manager jointly define the target group, learning objectives, and role of the VR unit within the overall learning journey. For sequential units, content and overall dramaturgy should be aligned. Key questions include: What are the main learning subjects? What prior knowledge do learners have? Which learning materials already exist? How does the unit fit into the overall curriculum? How much VR experience can be expected? Which corporate design requirements should be considered?

At Bosch, the two initially mentioned use cases were selected based on existing training programs. The availability of existing learning materials significantly reduced the planning effort, resulting in approximately 60 minutes for planning a sequential series of several 15-minute units. Planning effort decreases slightly with experience but remains largely consistent.

Step 2. Definition of content and general structure (2 hours)

The unit content is roughly structured in terms of sequence, presentation format, and user interaction. This determines when and in which logical order learners are introduced to new content. Creating a title-based chapter structure as a basis for the screenplay is recommended. Key questions include: Is the overall structure guided or self-exploratory? Is the order of content

intentional? How does the unit connect to preceding and subsequent units? If multiple units are developed, can they be connected by a common thread?

UC1 was designed with a sequential structure based on the concept of a technological museum with dedicated interaction points. In contrast, UC2 follows a more explorative approach, featuring numerous interactable objects embedded in the ISO-standardized process of de-energizing electric vehicles.

Step 3. Shaping the narrative (1 hour)

Based on the structured screenplay outline created in step 2, the narrative is refined. Key questions include: How relevant is the learning objective to the learner? What should the surroundings or props look like? How should the narrative be guided, e.g., through a narrator (video- or avatar-based), task-based interactions, or an unguided/explorative approach?

While UC2 features multiple versions with different levels of guidance, ranging from task-oriented to completely free exploration, UC1 explores different narrative formats depending on the nature and vividness of the content (cf. Figure 2).



Figure 14. *Narrative examples: video-narrator-based (a), floating-avatar-based (b), exploration-based (c)*

Step 4. Detailed unit planning and authoring

The following subtasks 4.1 – 4.3 are largely performed in parallel. Step 4.4 secures quality. Further, the proposed times are rough estimates based on the scenes shown in **Figure 2**, given that experienced model creators are available. Some technical trainings might demand much more efforts than soft skill or product centered trainings, since the required information density (models and interactions) can be much higher with growing technical complexity.

Step 4.1. Structured documentation (3 hours)

Detailed planning is documented using a screenplay template, which can be provided by the authors on request. The screenplay specifies the sequence, content presentation, annotations, and progression of the unit. The final plan should be reviewed and approved by the trainer. Key questions include: What happens step by step? How much time is allocated to each step? Which interactions should be sequential or allow parallel execution? How should alternative paths and feedback be designed?

Step 4.2. Define and model 3D objects (6 hours)

Required 3D objects and animations are identified and specified based on the screenplay. Existing repositories should be considered to increase efficiency. Key questions include: How should models and animations be structured hierarchically? Which visual details are functionally necessary to support the learning objective? Which environments and background assets are required? Should different levels of detail (LoD) be realized?

Experience at Bosch shows that the overall visual consistency of the scene is more important than the accuracy of individual objects. Vertex count, texture resolution, and detail density should therefore be aligned across assets. Hero objects, i.e., objects relevant to the learning focus, can receive higher accuracy than background assets to emphasize relevant content and minimize distraction as well as computing requirements, which has high impact on the effort.

Step 4.3. Authoring (3 hours)

The unit is implemented in dedicated authoring software. 3D objects are imported and placed, then enriched with conditional appearance and behavior to implement the intended learning progression. The approved screenplay serves as the reference document. The following best practices were derived from the development process:

- Use systematic naming conventions for 3D objects, scene elements, and animations.
- Use placeholders for missing 3D objects during development.
- Maximize modularity, reconfigurability; make systematic use of hierarchies, layers, groups or tags, and reuse of objects as building blocks.
- Use learning feedback appropriately, both in-scene and through e.g. xAPI statements, to support learner engagement as well as supervision and evaluation (cf. Step 4.4).

Step 4.4. Review and iterative optimization (4 hours)

As soon as the unit can be traversed successfully for the first time, the review process should begin. xAPI messages can provide valuable insights when tracked in a third-party Learning Record Store (LRS). Planned and actual execution times can be compared to assess intuitiveness, attention span, or task difficulty, while learning success can be evaluated in greater detail. These data may also provide a basis for future dynamic adaptations during runtime. Key questions include: Is the intended learning content presented in a correct way? How does the unit respond to incorrect or unexpected learner behavior or lack of attention? Furthermore, a consistent and professional quality standard is required to ensure learner acceptance.

Step 5. Quality Approval, Acceptance Test and Integration (2 hours)

The final evaluation is conducted by the project manager before the learning unit is approved for use. The unit must be completable without errors or unintended deadlock states. Corporate quality checks should be performed and formally documented, including acceptance test results and responsibilities. Finally, the unit can be integrated into a superordinate Learning Management System (LMS), such as Moodle, and tested for correct operation (cf. **Figure 3**). Selected lessons learned include:

- Units should not exceed approximately 20 minutes; alternatively, save points should be provided to enable recovery.
- The methodology should be regularly reviewed using a Scrum retrospective approach to support continuous improvement.
- Screenplays, 3D models, and reusable assets should be structured and documented carefully, as reusability and discoverability can substantially increase efficiency.

For the pilot phase, Moodle was not deployed within the project, as the VR courses are required by legal and internal operational regulations to be conducted with trainers present, who can therefore guide learners through the units. Regular consultations with the works council are conducted to evaluate feasible solutions. As a transfer measure after the project, the publication of selected XR-Upskill learning content and experiences as Open Educational Resources is currently assessed, with the learning units possibly to be orchestrated via Moodle.

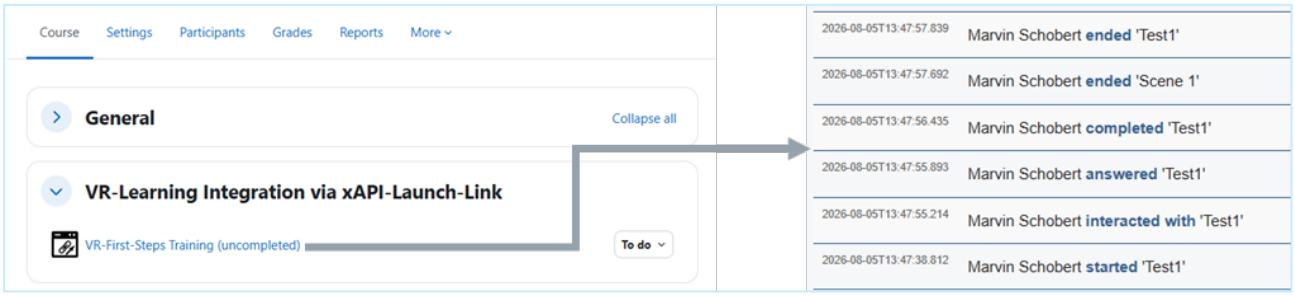


***Figure 15**. Example of Moodle integration (left) and Learning Record Store Feedback (right)*

Conclusion

This paper presented a practical methodology for the large-scale development of VR-based training at Robert Bosch GmbH in the context of the XR-Upskill research project, building on existing training structures and enabling trainer-driven content creation. The described workflow and design choices provide a structured yet flexible approach that can be adapted to different learning objectives, content types and levels of learner guidance. Initial experiences indicate that the creation of a 15-minute asynchronous VR learning unit can be achieved with roughly 22 person-hours, including 3D modeling and evaluation. This effort decreases substantially as content reuse and team experience increase, approaching a lower limit, while rising strongly with greater model complexity and information density. With the development phase now completed, the XR-Upskill project is transitioning into its implementation stage, where the learning units will be integrated into the intended training context and further evaluated.

References

Sturman, S., Sullivan, R. R., Anderson, D. & Perez, L. 2026. The KXR Instructional Design Framework: A Theory-Informed Model for Accessible Extended Reality Learning. Frontiers in Virtual Reality 7. https://doi.org/10.3389/frvir.2026.1739702

Embodied Biosignals: Holographic Biofeedback and Interactive Spatial Visualization for Physiology Education

Ghazal Hosseini[1,2], Hanif Haghtalab[1,2], and Nelson J. S. Silva[1]

[1]*Interaction and XR Labs, IT:U Interdisciplinary Transformation University Austria, Linz*
[2]*Interface Cultures, Dept. of Media Art, University of Arts Linz, Austria*

Corresponding author: ghazal.hosseini@it-u.at

Keywords: XR, education, biosignal, visualization, sensor, holographic, embodied, learning

Abstract

Learning physiological concepts such as heart rate is difficult because the signals are abstract and invisible, and numerical or two-dimensional displays convey them only in limited ways. We present a prototype XR application that renders a learner's live cardiovascular signal as an interactive holographic heart. Grounded in embodied learning, the system streams heart rate from a wearable sensor into Unity, which drives a dynamic 3D heart model on a light-field display and gesture tracking supports direct hands-on exploration. We demonstrated the application at the Open Day event at our public university. This event was attended by approximately 1,000 visitors. Our application was evaluated with ten participants using the System Usability Scale (M = 86.25, SD = 11.01), with an "excellent" rating. Participants also reported that the spatial representation supported their understanding of heart rate more than a number or a graph. We discuss implications for physiology and health-technology education.

Introduction

Recent advances in Extended Reality (XR) technologies have created new opportunities for interactive and embodied learning environments (Johnson-Glenberg, 2018). XR systems can integrate spatial visualization, biosignal data, and natural interaction to support learning experiences that extend beyond traditional 2D screen-based representations. This aligns with embodied learning theory, which holds that understanding is shaped not only by abstract reasoning but also by sensory and motor engagement with content (Nathan, 2021). In XR, this principle has a direct design implication: when a learner's physical actions are meaningfully coupled to the concept being taught, this congruence can enhance understanding more effectively than passive observation (Johnson-Glenberg, 2018). Motivating our prototype, we extend this idea to biosignals by allowing learners to see and physically interact with their own physiological data as a spatial, holographic object rather than as a graph or number. Glasses-free light-field displays have been shown to improve anatomical learning compared to conventional 2D materials (Muro et al., 2025), and immersive biofeedback interfaces have been found to increase engagement with physiological self-regulation tasks (Blum et al., 2019). Building on this evidence, this work contributes an educational XR prototype that uses spatial visualization and embodied interaction to evaluate how such approaches can enhance understanding of biosignal representations and strengthen health-technology education.

System Architecture

The system (Figure 1) comprises four layers: layer 1 - physiological sensing, layer 2 - data streaming, layer 3 - visualizations, and layer 4 - gesture-based interaction. In layer 1, a cardiovascular signal is captured using a PLUX biosignal sensor worn by the user, recording Blood Volume Pulse (BVP). Heart rate is automatically calculated using the PLUX Hear Rate Variability real-time plugin (v. 2.2.7) and visualized using a 2D dashboard (Virtanen et al., 2020). In layer 2, HR data is streamed via Lab Streaming Layer (LSL), a framework for synchronized, low-latency transmission of physiological and behavioral data (Kothe et al., 2025), into the Unity game engine. In layer 3, Unity drives a 3D heart model rendered on both a Looking Glass light-field display and a standard monitor, mapping the user's heartbeat through synchronized changes in pulsation speed, surface color, and numerical HR readout. This enables heart-rate perception through motion, color, and numeric feedback. Layer 4 uses Leap Motion Controller 2 for gesture-based interaction with the holographic heart and its physiological data.

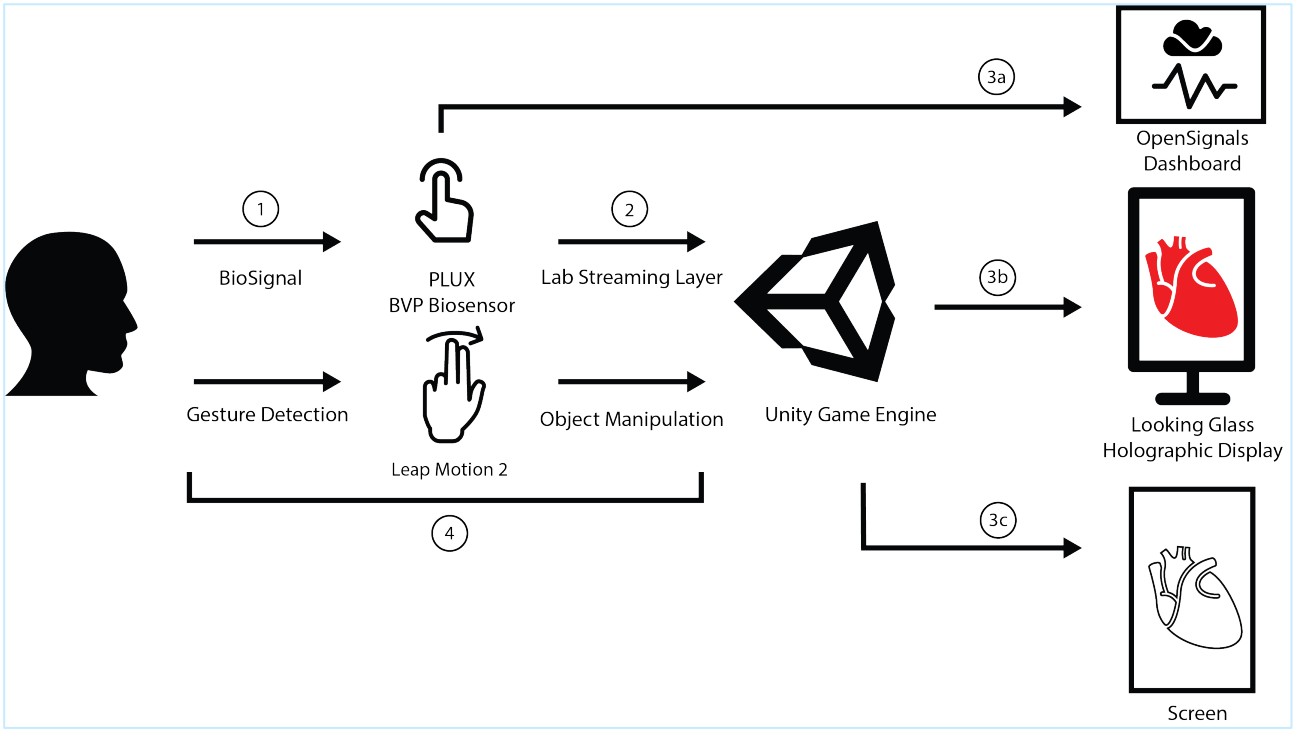


Figure 16. *System architecture: layer 1 – physiological sensing; layer 2 – data streaming; layer 3 – visualizations (2D dashboard [3a], spatial 3D heart [3b], screen-based 3D heart [3c]); layer 4 – gesture-based interaction.*

Deployment and Demonstration

The application was demonstrated at our public university's Open Day in Austria, on the 24th of April 2026 (Figure 2), an annual government supported event attended by around 1,000 visitors, including prospective students, school groups, and families. Many visitors engaged directly with the demonstration, receiving a brief introduction while wearing the sensor and exploring their heartbeat rendered as a holographic model (Figure 3) using hand gestures to rotate and inspect it. Informal feedback from observation and brief exchanges indicated strong engagement and curiosity. This initial deployment validated the system's robustness in a public setting and provided early qualitative impressions ahead of the structured usability evaluation.



Figure 17. *Public demonstration: a large screen (65”) positioned outside the lab showed the heart model to visitors.*

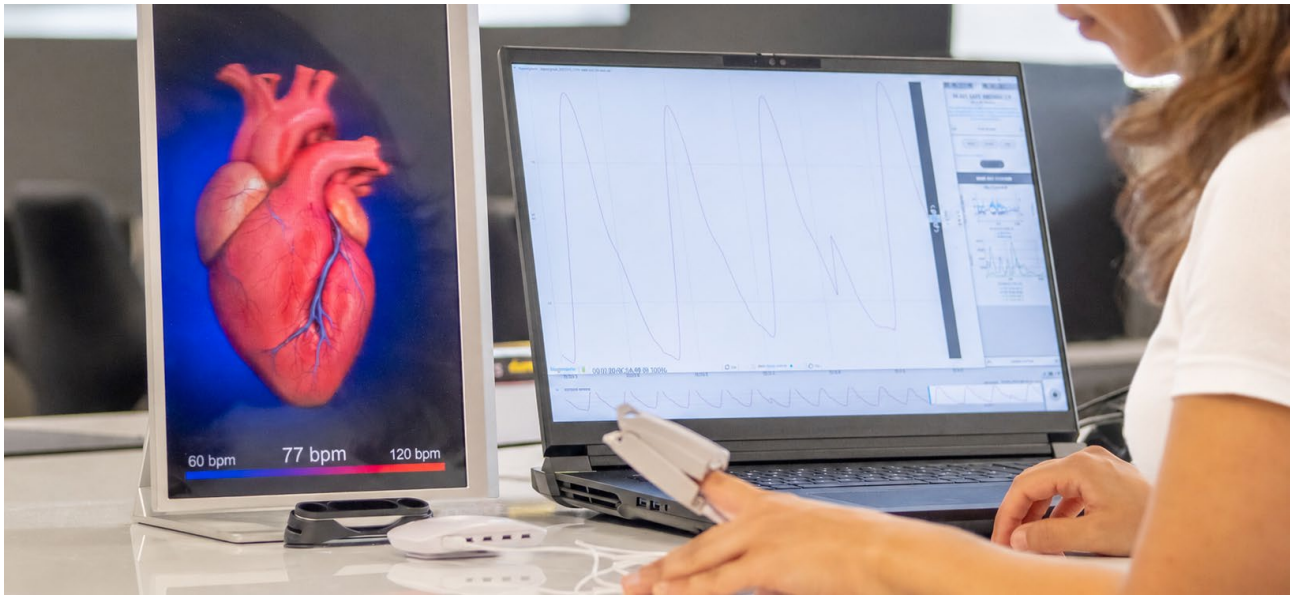


Figure 18. *A user wearing the PLUX Blood Volume Pulse (BVP) biosensor interacts with our application, it displays a 3D heart on a light-field display while streaming the user’s biosignal heart-rate data to the model.*

Preliminary Usability Evaluation

Ten participants who were learners at our university took part in our study in August 2026. Each participant wore the BVP biosensor and explored the holographic heart using hand gestures for approximately 10 minutes, then completed a SUS digital questionnaire (Brooke, 1996). One team member administered the SUS questionnaire. Demographic data were not collected. Participation was voluntary and unpaid, and their consent was obtained. Formal ethics approval was not required as no personal or physiological data were retained. The mean SUS score was 86.25 (SD = 11.01, Mdn = 85.00, range 70–100), 95% CI [78.38, 94.12]; scores were consistent with normality (Shapiro–Wilk W = 0.93, p = .406). A one-sample t test against the SUS benchmark of 68 (Sauro & Lewis, 2016) was significant, t(9) = 5.24, p < .001, Hedges' g = 1.52, 95% CI [0.61, 2.39], corresponding to grade A and an adjective rating of "excellent" (Bangor et al., 2009). Responses to perceived learning were strongly positive (Mdn = 5, M = 4.70, SD = 0.67; Wilcoxon V = 0, p = .004). Given the small, self-selected sample and reliance on self-report, these results are preliminary.

Limitations

Interacting with the 3D holographic model while wearing the sensor can itself alter heart rate, so the visualized signal is not an undisturbed physiological reading. Latency was not quantified. The system is intended for usability and educational exploration rather than clinically accurate bio-signal measurement. Because our current users' sample is limited, further evaluation studies are necessary to strengthen the reliability and generalizability of the results.

Conclusion

This paper introduced an educational XR application that visualizes a user's cardiovascular bio-signal as a holographic 3D heart, enabling spatial exploration and gesture-based interaction with one's own physiological data. Our evaluation focused on usability rather than clinical accuracy. Our participants rated the system highly and future work will extend the visualization to additional physiological signals, such as heart-rate variability. We intend to examine whether embodied holographic representations generalize well to other abstract physiological concepts in education. This would result in improvements in the learners' understanding of biosignals data, while the use of holographic display technology would free the user from wearing additional head-mounted displays.

References

Bangor, A., Kortum, P. & Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies 4(3), 114–123.

Blum, J., Rockstroh, C. & Göritz, A. S. (2019). Heart Rate Variability Biofeedback Based on Slow-Paced Breathing With Immersive Virtual Reality Nature Scenery. Frontiers in Psychology 10, 2172. https://doi.org/10.3389/fpsyg.2019.02172 .

Brooke, J. (1996). SUS: A Quick and Dirty Usability Scale. In: P. W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Eds.), Usability Evaluation in Industry. London: Taylor & Francis. pp. 189–194.

Johnson-Glenberg, M. C. (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. Frontiers in Robotics and AI 5, 81. https://doi.org/10.3389/frobt.2018.00081.

Kothe, C., Shirazi, S. Y., Stenner, T., Medine, D., Boulay, C., Grivich, M. I., Artoni, F., Mullen, T., Delorme, A. & Makeig, S. (2025). The lab streaming layer for synchronized multimodal recording. Imaging Neuroscience 3, IMAG.a.136. https://doi.org/10.1162/IMAG.a.136.

Muro, S., Miyosawa, K., Yamaguchi, K., Okamoto, K., Okamoto, S., Itamiya, T. & Akita, K. (2025). Evaluation of Glasses-Free 3D Anatomy Learning Materials Through a Randomized Control Study With a Puzzle-Like Method. Anatomical Sciences Education 18(7), 684–696. https://doi.org/10.1002/ase.70063.

Nathan, M. J. (2021). Foundations of Embodied Learning: A Paradigm for Education. New York, NY: Routledge. https://doi.org/10.4324/9780429329098.

Sauro, J. & Lewis, J. R. (2016). Quantifying the User Experience: Practical Statistics for User Research. 2nd ed. Cambridge, MA: Morgan Kaufmann.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, *17*(3), 261-272. https://doi.org/10.1038/s41592-019-0686-2.

Smart Glasses versus Hand Scanning for Automotive Parts Sorting in a Circular Economy Testbed: A Pilot Evaluation

Mario Lorenz[1], Arwa Own[1], Michael Bolech[2], Markus Wagner[2] and Konstantinos Georgopoulos[2]

[1] *Chemnitz University of Technology, Professorship for Production Systems and Processes, Reichenhainer Straße 70, 09126 Chemnitz, Germany*

[2] *Circular Economy Solutions GmbH, Lammstr. 23, 76133 Karlsruhe, Germany*

Corresponding author: mario.lorenz@mb.tu-chemnitz.de

Keywords: Augmented reality; smart glasses; circular economy; logistics; industrial sorting

Motivation and Application

Circular-economy approaches aim to replace conventional end-of-life concepts by keeping products, components and materials in productive use through strategies including reduction, reuse, recycling and recovery (Kirchherr et al., 2017). For remanufacturing supply chains, this requires not only the recovery of used components but also reliable logistics processes that identify, classify and route these components to their intended destinations.

In the C-ECO use case of the Horizon Europe project DiCiM (https://www.dicimproject.eu/), automotive cores are inspected and assigned to dedicated sorting boxes before being forwarded for remanufacturing. Conventionally, operators identify the correct sorting location from labels and manually place each core into the corresponding box. As the physical put-away step is performed without technical confirmation, repetitive work, misreading of labels and loss of concentration can result in incorrect sorting.

Wearable augmented reality devices, such as smart glasses, provide the possibility of presenting task-relevant information directly within the operator's field of view while keeping both hands available. Previous research on AR smart glasses in warehouse order picking has demonstrated potential improvements in efficiency and error reduction, while also showing that worker experience, well-being and individual characteristics are important factors for successful implementation (Windhausen et al., 2024). Consequently, industrial smart-glasses solutions should not only be assessed according to their technical feasibility but also compared with established interaction technologies from a user perspective.

To investigate this in a circular-economy setting, an AR-supported sorting application was developed for the RealWear Navigator 520 smart glasses and connected to C-ECO's CoremanNet warehouse management system. After scanning the QR code of a core, the application retrieves and displays its assigned sorting location. Following put-away, the QR code of the sorting box is scanned and the placement is validated. Core ID, box ID and timestamp can subsequently be transmitted to the backend, enabling digital confirmation and traceability of the sorting operation. The underlying interface is device-independent and therefore also permits the same process logic to be implemented using handheld scanning hardware.



Figure 19. *Interaction devices evaluated in the C-ECO core-sorting testbed: RealWear Navigator 520 smart glasses (left) and handheld scanning setup (right).*

Comparative Evaluation

A within-subject pilot evaluation with N = 3 operators compared hands-free smart glasses with hand scanning for the same QR-supported sorting task in a circular-economy testbed (Figure 1). Each participant performed the sorting process using both interaction approaches.

All participants were male, with a mean age of 34.7 years (range: 29–41 years). Regarding highest educational qualification, one participant reported a Bachelor's degree, while two selected "other". Previous experience with Pick-by-Vision was limited: two participants reported no prior experience, and one reported less than six months of experience. For Pick-by-Scan, one participant reported no prior experience, while two reported 6–24 months of experience. General experience in picking/logistics varied: two participants reported less than six months of experience, whereas one reported more than five years of experience.

Following each condition, perceived usability was assessed using the System Usability Scale (SUS) according to (Brooke, 1996). User experience was additionally evaluated using the short version of the User Experience Questionnaire (UEQ-S), comprising the dimensions Pragmatic Quality and Hedonic Quality (Schrepp et al., 2017). Participants further rated whether the respective device was helpful for sorting and whether it supported the avoidance of sorting errors. Free-text feedback was collected to identify practical advantages and limitations experienced during operation. Given the small pilot sample (N = 3), all results are interpreted descriptively and no inferential statistical conclusions are drawn.

Results

The hand-scanning condition achieved a mean SUS score of 77.5, compared with 49.2 for the smart glasses, indicating higher mean perceived usability for the handheld interaction in the present implementation.

The same tendency was observed for the pragmatic dimension of the UEQ-S. On the scale from −3 to +3, mean Pragmatic Quality was 1.67 for hand scanning compared with −0.25 for the smart

glasses. In contrast, the smart glasses achieved a higher mean Hedonic Quality score (1.83) than the hand scanner (1.25). The corresponding overall UEQ-S means were 0.79 for the smart glasses and 1.46 for hand scanning. Due to the small sample, these values are used to describe tendencies within the pilot rather than to establish statistically generalisable differences.

Despite these differences in usability and user experience, all three participants considered both technologies helpful for the sorting task and indicated that both approaches could support the avoidance of sorting errors. The qualitative responses help explain the difference between the devices. For the smart glasses, participants reported that wearing the device required acclimatisation, with one participant mentioning temporary dizziness. Connection stability and additional process time were also identified as limitations. Feedback regarding the handheld scanner was generally more favourable in terms of handling, although participants noted practical disadvantages associated with carrying the accompanying mobile device during work.

Conclusion

The pilot evaluation shows that both smart glasses and hand scanning can provide digitally validated sorting and thereby address a central weakness of the conventional sorting process: the absence of immediate technical confirmation that a core has been placed in the intended sorting box.

For the investigated testbed and the current implementation, hand scanning achieved higher pragmatic usability, whereas the smart glasses achieved higher hedonic user-experience ratings and provide the inherent advantage of hands-free interaction. The results therefore do not indicate a generally superior interaction technology. Instead, they highlight a trade-off between the practical maturity of handheld scanning and the hands-free and experiential potential of wearable AR.

Improvements in wearing comfort, connection reliability, scanning robustness and interaction speed appear particularly relevant for future iterations of the smart-glasses solution. Although the small sample size limits generalisation, the comparison demonstrates the importance of evaluating industrial XR solutions against established alternatives within the actual work process and of considering user experience alongside technical performance and error prevention.

Acknowledgement

The authors acknowledge financial support from the European Union's Horizon Europe research and innovation programme under the grant agreement No. 101091536 (DiCiM project). The authors used ChatGPT (OpenAI) for language editing and manuscript preparation support. All content was reviewed and approved by the authors, who take full responsibility for the manuscript.

References

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation and Recycling*, *127*, 221–232. https://doi.org/10.1016/j.resconrec.2017.09.005

Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, *4*(6), 103–108. https://doi.org/10.9781/ijimai.2017.09.001

Windhausen, A., Heller, J., Hilken, T., Mahr, D., Di Palma, R., & Quintens, L. (2024). Exploring the impact of augmented reality smart glasses on worker well-being in warehouse order picking. *Computers in Human Behavior*, *155*, 108153. https://doi.org/10.1016/j.chb.2024.108153

CoBotXpeRience: An XR-Based Training Framework for Collaborative Robot Programming and Industrial Upskilling

Lorenzo Turani[1], Leonardo Cattaneo[1], Tommaso Salgarelli[1], Elena Mossali[1], Samuele Bonalumi[2], Stefano Parma[2], and Sona Arevshatyan[2]

[1]Consorzio Intellimech, Bergamo, Italy

[2]MADE Competence Center, Milano, Italy

Corresponding author: lorenzo.turani@intellimech.it

Keywords: Virtual Reality, Collaborative Robots, Cobot Programming, XR Education, Training, Industry 5.0

The Industrial Challenge

The adoption of collaborative robots (cobots) and advanced automation technologies is reshaping modern manufacturing, creating an increasing demand for workforce upskilling and reskilling. Unlike conventional industrial robots, cobots operate in shared workspaces with human operators and require technicians to understand not only robot programming but also human–robot collaboration, safety procedures, task sequencing, motion planning, and workspace organization. As Industry 5.0 promotes human-centric automation, vocational education and industrial training institutions face the challenge of preparing learners with practical competencies that can be effectively transferred to real production environments (Breque et al., 2021). Despite the growing need for robotics education, practical training remains a significant challenge. Traditional hands-on approaches depend on physical robotic work cells that are expensive, available only for limited training sessions, and subject to strict safety constraints. Furthermore, conventional robot programming often requires previous technical expertise and direct interaction with industrial robot controllers. These limitations reduce training scalability and provide few opportunities for learners to experiment, make mistakes, and progressively develop confidence before operating real robotic systems (Mulero-Peréz et al., 2026; Ipsita et al., 2025). Extended Reality (XR) technologies offer an effective solution to these challenges by providing immersive, safe, and repeatable learning experiences. Rather than replacing conventional teaching, XR enables learners to bridge the gap between theoretical concepts and practical application through realistic industrial simulations where they can observe robot behaviour, understand safety mechanisms, test programming decisions, and experience collaborative workflows without risks to people, equipment, or production activities (Mulero-Peréz et al., 2026; Ipsita et al., 2025; Mulero-Peréz et al., 2025). To address these challenges, the CoBotXpeRience project developed an XR-based educational platform for collaborative robotics within the Open Call of the European MASTER-XR project (Barz et al., 2024).

XR Solution and Use Cases

CoBotXpeRience is an XR-based training framework designed for vocational learners, university students, technicians, and industrial operators. Developed in Unity using the XR Interaction Toolkit and deployed through the VIROO platform, the application runs on Meta Quest 3 headsets and incorporates a digital twin of a Universal Robots UR10 collaborative robot. The educational approach combines guided instruction with immersive experimentation. Participants interact directly with

industrial environments, execute robotic operations, and analyse the outcomes of their decisions within a safe virtual setting. Simulation enables users to explore robot behaviour, test alternative solutions, and understand programming logic without risks to personnel or equipment.

Two industrial scenarios were developed. The first reproduces a collaborative motorbike assembly workstation through a realistic industrial digital twin, enabling learners to explore the complete production cell and become familiar with its main components, including collaborative robots, the operator station, assembly parts, and the AGV (Figure 1). Users can interact with the robotic system through manual jog operations based on both forward and inverse kinematics, perform pick-and-place and target-positioning activities, and execute collaborative assembly tasks. The experience promotes understanding of production workflows, workspace organisation, and human-robot interaction, while also visualising safety-related behaviours and operational constraints within a shared collaborative environment.



Figure 20. *CoBotXpeRience first scenario: collaborative motorbike assembly workstation.*

The second scenario focuses on a flexible pick-and-place Cobot Workstation dedicated to robot programming (Figure 2), replicating the Smile Laboratory robotic cell within an immersive XR environment. Participants interact with a virtual cobot work cell to define waypoints, select motion trajectories, configure grippers, and execute automated manipulation tasks. Through visual and spatial interaction, users can explore robot motion planning, launch predefined routines, and observe task execution in real time while receiving contextual guidance from the virtual assistant. The environment combines interactive user interfaces, robot models, and embedded motion-planning functionalities to make programming concepts directly observable and intuitive.

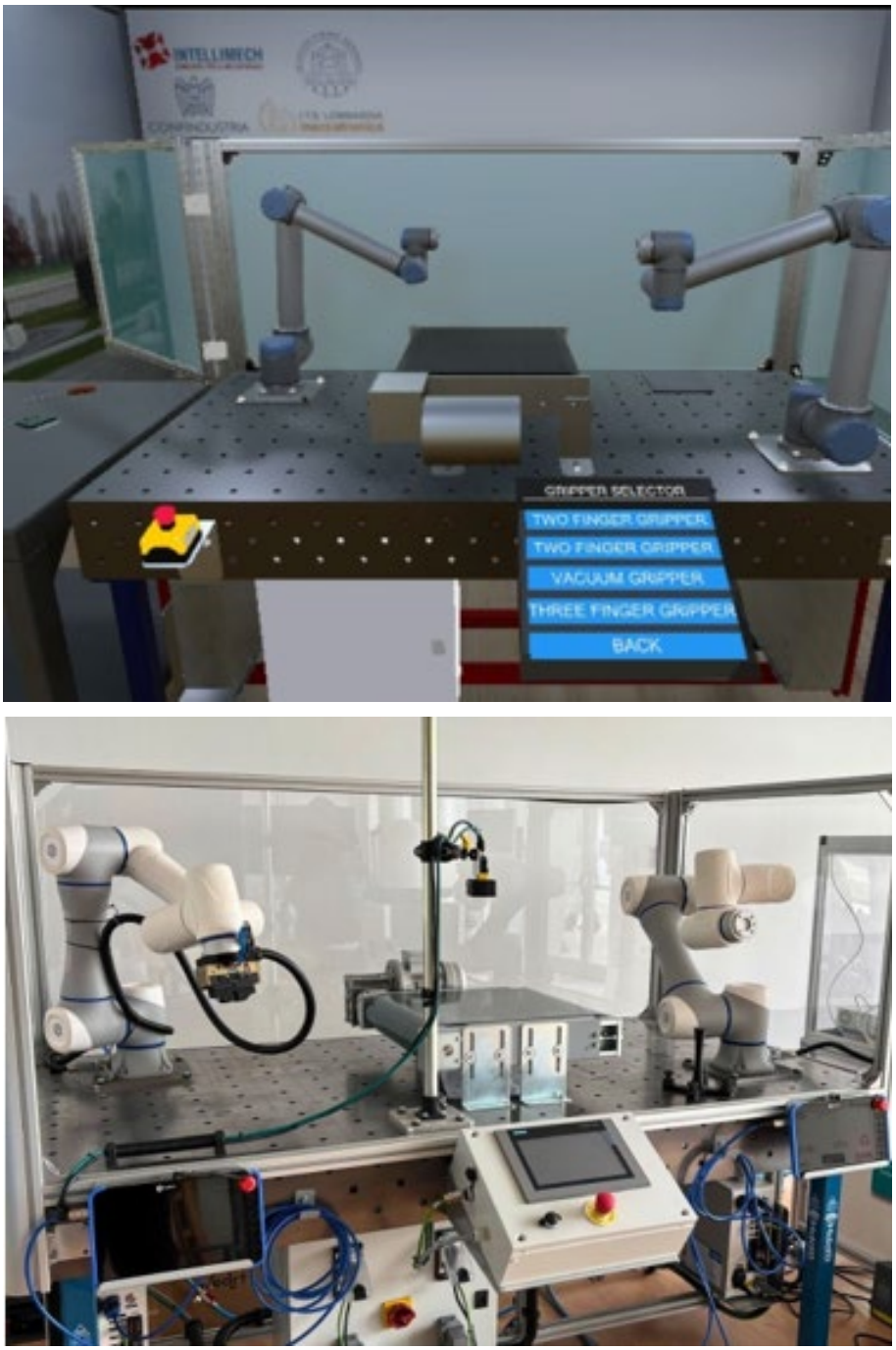


Figure 2. *CoBotXpeRience second scenario: flexible pick-and-place workstation.*

Together, the two scenarios connect contextual understanding of collaborative workstations with hands-on programming practice, providing a comprehensive introduction to industrial collaborative robotics.

Validation and Results

The proposed educational framework was validated with 56 vocational students from different educational backgrounds, 10 teachers specialising in robotics and technical subjects, and 15 industrial workers participating in professional upskilling activities. The validation protocol combined introductory theoretical sessions, guided XR experiences, supervised practical exercises, and post-session feedback collection, which supported the educational value of the proposed framework (Table 1). Participants experienced both immersive educational scenarios. Throughout the activities, the framework progressively guided learners from conceptual understanding towards autonomous robot programming while providing continuous instructional support through a virtual assistant and interactive feedback mechanisms. An additional educational feature emerged during classroom deployment. Although the immersive interaction was performed individually, the headset view was streamed to the classroom, enabling peer observation and collaborative discussion. Students could anticipate mistakes, discuss alternative solutions, and learn from previous users, creating an additional collaborative learning dimension that complemented the individual XR experience. Learners reported high levels of engagement and confidence while interacting with collaborative robots through immersive technologies. Teachers recognised the educational value of the platform as an effective complement to conventional classroom activities, particularly for connecting theoretical concepts with realistic industrial practice. Industrial participants considered the platform relevant for introducing non-expert operators to collaborative robotics and safety procedures before accessing physical robotic systems.

Table 3. *Summary of the Survey Results*

Evaluation Criterion	Rating Scale	Average Score
Level of satisfaction and feedback	0-10	8.1
Achievement of learning objectives	0-10	7.2
Improvement in familiarization with robots	1-5	4.0
Improvement in familiarization with XR	1-5	3.4

Four main educational outcomes emerged from the validation. First, immersive interaction significantly reduced learners' initial uncertainty when approaching collaborative robots. Second, the progressive learning structure facilitated the transition from conceptual understanding to robot programming. Third, simulated errors and immediate visual feedback promoted a deeper understanding of robot behaviour and programming decisions. Finally, the combination of immersive individual practice and classroom observation fostered collaborative knowledge construction, suggesting that XR can effectively support both individual and peer-assisted learning.

Lessons Learned and Future Adoption

The CoBotXpeRience experience demonstrates that XR technologies can effectively complement conventional robotics education. Rather than replacing physical laboratories, immersive digital twins allow learners to acquire safety awareness, confidence, and programming skills before interacting with real collaborative robots, while intuitive interaction lowers the entry barrier for learners with limited previous experience. Beyond the educational outcomes, the project demonstrates the transferability of XR technologies to vocational education and industrial training, and its integration within the VIROO platform provides a reusable and scalable foundation for future XR applications in manufacturing education.

Despite the results, the study presents some limitations. The educational experience was evaluated on a limited set of collaborative robotics scenarios, and no long-term assessment was conducted to determine whether the acquired competencies are retained over time and effectively transferred to real industrial workstations. In addition, interaction was primarily based on standard XR input devices, without investigating the impact of advanced multisensory interfaces. Future work will focus on extending the platform to a broader range of manufacturing processes and collaborative

programming tasks, as well as conducting long-term studies to evaluate knowledge retention and skill transfer from virtual training environments to physical industrial settings. Furthermore, the integration of haptic technologies, including MAGOS haptic gloves, will be investigated to assess their contribution to realism, user embodiment, and engagement during robot programming, component manipulation, and human–robot interaction activities.

References

Barz, M., Karagiannis, P., Kildal, J., Rivera Pinto, A., de Munain, J. R., Rosel, J., ... & Sonntag, D. (2024). MASTER-XR: Mixed Reality Ecosystem for Teaching Robotics in Manufacturing. In Integrated Systems: Data Driven Engineering (pp. 167-182). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53652-6_10

Breque, M., De Nul, L., & Petridis, A. (2021). Industry 5.0: towards a sustainable, human-centric and resilient European industry (No. KI-BD-20-021-EN-N). Directorate General for Research and Innovation (DG RTD) of the European Commission. https://doi.org/10.2777/308407

Ipsita, A., Kaki, R., Liu, Z., Patel, M., Duan, R., Deshpande, L., ... & Ramani, K. (2025). Virtual reality in manufacturing education: A scoping review indicating state-of-the-art, benefits, and challenges across domains, levels, and entities. *arXiv preprint arXiv:2503.18805*. https://doi.org/10.48550/arXiv.2503.18805

Mulero-Pérez, D., Zambrano-Serrano, B., Ruiz Zúñiga, E., Fernández-Vega, M., & Garcia-Rodriguez, J. (2026). X-RAPT: A Virtual Reality Platform for Adaptive Training in Robotic Programming. In International Conference on Soft Computing Models in Industrial and Environmental Applications (pp. 623-632). Springer, Cham. https://doi.org/10.1007/978-3-032-19763-4_58

Mulero-Pérez, D., Zambrano-Serrano, B., Ruiz Zúñiga, E., Fernandez-Vega, M., & Garcia-Rodriguez, J. (2025). Enhancing robotics education through XR simulation: insights from the X-RAPT training framework. *Applied Sciences*, *15*(18), 10020. https://doi.org/10.3390/app151810020

Overcoming Workforce Shortages with a VR Digital Twin: The Liege Airport Case Study

Camille Bouvier, and Michaël Schyns

HEC Liège – Management School of the University of Liège, LabXR, Belgium

Corresponding author: M.Schyns@uliege.be

Keywords: Virtual Reality training, job promotion, shortage occupation, digital twin, airport operations

Introduction

Talent shortage is a widespread challenge faced by many industries, but it is particularly acute in the aviation sector. Airports are complex micro-worlds encompassing a vast array of activities. Liege Airport (LGG), one of Europe's leading cargo hubs, illustrates this reality, with daily operations relying on ground professions such as apron marshalling, cargo highloader driving, snow and de-icing operations, and aircraft handling. Some are highly specialized occupations; others are common trades people rarely associate with an airport. As a result, candidates rarely think of applying, making these professions increasingly hard to recruit for. To counter this shortage, companies are prompted to organize promotional campaigns and their own on-site training. However, doing so within an airport environment presents unique challenges. First, it is difficult to organize visits on such an ultra-secure site. Furthermore, training on the real infrastructure is structurally constrained: operations cannot be interrupted, safety margins are tight, and real vehicles and aircraft stands are scarce resources over which logistics operations always take priority.

To overcome these barriers, a virtual reality (VR) digital twin was initiated within the Interreg "Digital Twin Academy" project. Following initial doctoral research by J. Tonka on digital twins (DT), LabXR partnered with Liege Airport Academy (LAA) to transform this DT into an interactive training and job promotion ecosystem using Unity and Quest headsets. Indeed, meta-analyses confirm that VR training meets or exceeds established training methods (Kaplan et al., 2021; Howard et al., 2021), including in safety-critical settings (Scorgie et al., 2024). For promotion, an additional photorealistic tool using an Insta360 Pro camera was deployed. Designing such tools and measuring their real added value raises new research questions. This paper presents the tool we developed, preliminary evidence from three complementary sources, a promotion-tool survey, informal feedback from training-module users, and a quantitative acceptance study of one module, along with the operational lessons learned.

An ecosystem of training and promotion modules

Thanks to its team of developers specialized in game design, our lab developed several operational training modules for LAA, all built on a shared 3D reconstruction of the airport site (Figure 1). The MAN module, by far the most used, trains and evaluates mastery of the traffic rules and safety procedures required to circulate on the manoeuvring area. The highloader module trains operators of the vehicle that lifts unit load devices into the aircraft hold. The pushback and snow/de-

icing modules place trainees in the cab of the corresponding vehicle. One module teaches the basic steps of aircraft maintenance and the pre-departure checks to be carried out. In partnership with the airline ASL Airlines, a 'pre-flight' module allows pilots to practice their procedures and checklists under highly realistic conditions. Finally, a marshalling module teaches learners how to guide an aircraft into its parking position. Learners can choose between a 'tutorial' mode, which guides them step by step, and a 'test' mode, where they are expected to already know the procedures and receive a performance report from the system at the end. Instructors can monitor what the learner is doing at any time and adjust the scenario if they wish: technical incidents, fog, night conditions, and more.



Figure 21. *Highloader, Pilot check-list, Maintenance, Pushback, Snow removal training & Computer Vision DT*

Synchronized with on-site Internet of Things (IoT) sensors and machine learning algorithms (Fig. 1), the digital twin serves operational monitoring while providing trainees with a high-fidelity environment identical to the real tarmac. It places learners in an environment they will recognize rather than in a generic simplified simulator, thereby easing the transfer of skills from virtuality to reality.

Additionally, to showcase airside jobs hidden behind security fences, a second generic tool was developed. It allows the airport, or any company, to easily create photorealistic VR/360° tours with interactive job description hotspots (Figure 2). This makes it possible to bring the airport anywhere, notably to job fairs and schools, to promote available positions and jobs.

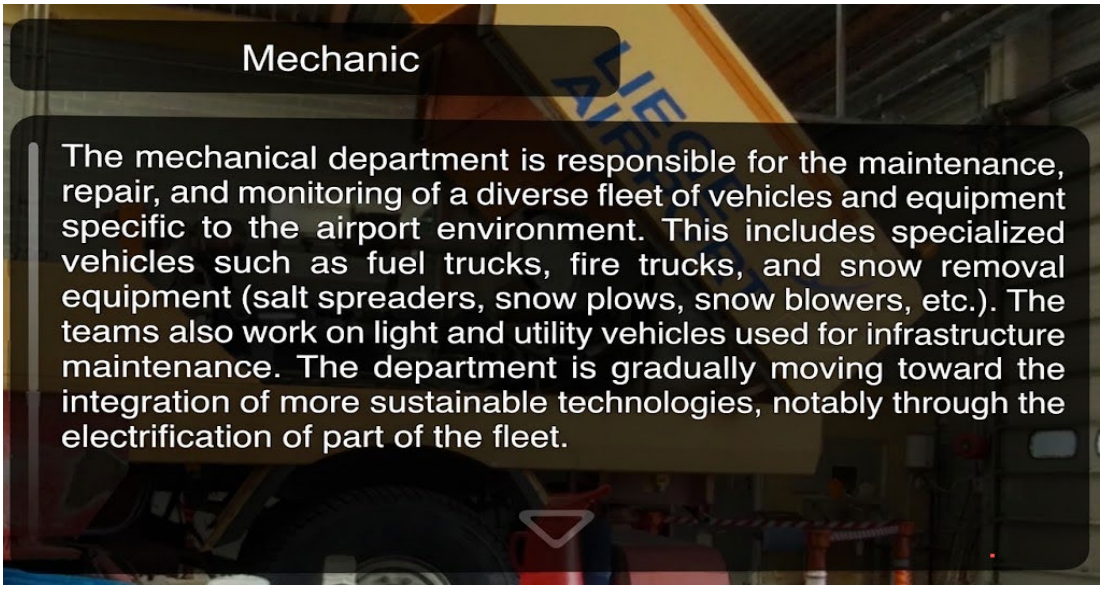


Figure 22. *Job-description panel of the VR/360° promotion module*

Results

A proof-of-concept for the promotion tool was evaluated with 15 participants (13 valid responses). All respondents agreed the tool improved their understanding of airport trades, rating visual realism at 3.9/5 (77%≥4). However, its impact on intent to apply was more modest (54%), confirming that job awareness and career motivation are distinct outcomes. Beyond the initial "wow factor," the VR technology alone is far from sufficient. Human resource and marketing teams play an essential role in defining the core messaging to be embedded in the application. Moreover, the tool cannot stand on its own and must be integrated into broader, traditional recruitment and promotional approaches.

For training modules, informal feedback from dozens of users confirmed high perceived utility and fidelity. Reaching that point, however, required more than building an immersive environment, particularly in a professional context where operators are busy, under operational pressure, not naturally inclined to step outside established routines, and not necessarily comfortable with digital tools. Two lessons stood out:

- Early and Continuous Trainer Involvement: Domain experts are essential for procedural accuracy but often under-specify VR requirements. Conversely, modules designed without their input lack ownership. For example, despite its formal integration, the MAN module was initially underutilized due to unfamiliarity. Dedicated support during early sessions built instructor confidence, leading to enthusiastic adoption; trainers noted that errors made in VR effectively prevent real-world mistakes. An agile, iterative development loop is critical.
- Open, Flexible Game Design: Modules feature three modes: guided tutorial, evaluative test, and free/instructor-led mode. The free mode proved vital, enabling trainers to improvise, such as building custom control panel walkthroughs or integrating external voice-chat to simulate air traffic control. Rigid scenarios would have restricted these spontaneous pedagogical uses.

Quantitative evaluation of the MAN theoretical training module (n = 25, mean age 38) using the Technology Acceptance Model (Davis, 1989) and UQO presence/cybersickness metrics confirmed robust acceptance (Fig. 3). Statistical power analysis (Cohen's d = 0.8, α = 0.1) confirmed sample adequacy (power = 87%). Perceived Ease of Use (79.4%), Perceived Usefulness (74.0%), and Sense of Presence (73.7%) all scored high. Ease of use correlated with usefulness (r = 0.49), and both correlated with presence (r = 0.67 and 0.72). Technical issues significantly reduced ease of use (p = 0.036), while session length and simultaneous users showed negative trends, highlighting cybersickness and duration as key friction points.

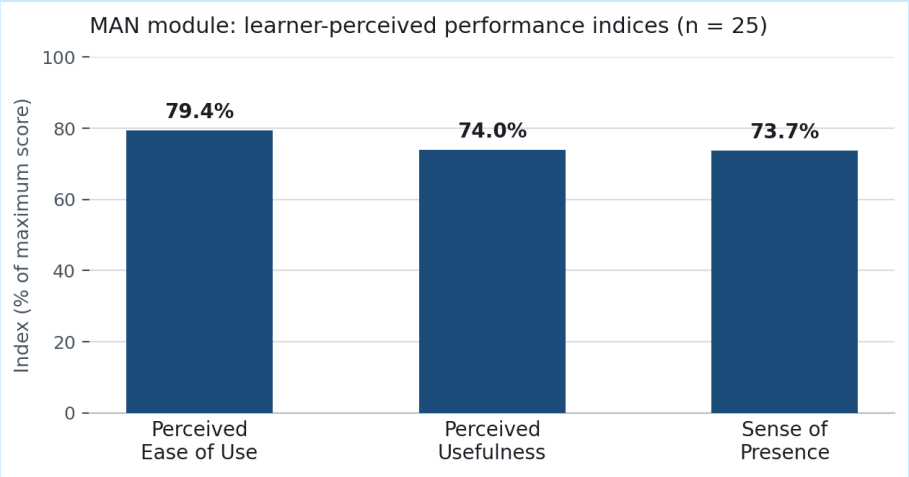


Figure 23. *Learner-perceived performance indices for the MAN module (adapted from Droeven, 2024)*

What comes next

Custom VR development for niche environments remains costly. To ensure long-term viability, tools must stay modular, open, and extensible. Furthermore, Artificial Intelligence is required to

expand the range of possible scenarios. Moreover, early usage revealed, e.g., when instructors adapted the MAN module to simulate air traffic control communications, that collaborative features would be highly valued. This reflects a gap in the literature: most VR training research still centers on individual learners, whereas most of the tasks trained in an airport, as in many other high-risk operational settings, are in practice team tasks. Building on this, a PhD project led by the first author will extend it further, combining career-exploration modules, comparing immersive job previews to conventional ones, with Artificial Intelligence/Operations-Research-driven adaptive orchestration of multi-user collaborative scenarios.

Conclusion

Over several years, a digital twin, a job-promotion tool, and several VR training modules have grown into a coherent training ecosystem for a 24/7, safety-critical, hard-to-recruit-for environment. Learner feedback and preliminary quantitative results are positive and suggest VR could become a powerful tool to fight against job shortages through promotion and training. These encouraging results obtained for the airport case may well be extended to other areas. However, the main lesson so far is organizational rather than purely technical: sustained trainer involvement and a deliberately open, extensible game design mattered at least as much as the immersive experience itself.

Acknowledgements

The authors thank LAA, and in particular I. Perin, for their sustained support throughout this program, and the HEC master's students who each contributed one piece of this ecosystem: C. Ruche, N. Droeven, M. Jiang and A. Fiévet. The digital twin was developed within the Digital Twin Academy project, funded under Interreg V-A Euregio Meuse-Rhin with support from the European Regional Development Fund.

References

Bouvier, C. (2026). Development of a virtual promotion module for Liège Airport Academy. *Master's thesis, HEC Liège*, University of Liège. http://hdl.handle.net/2268.2/25746

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.

Droeven, N. 2024. Design, development, and integration of an effective virtual reality training module for logistics operators in the field of air cargo. Master's thesis, HEC Liège, University of Liège. http://hdl.handle.net/2268.2/20081

Fiévet, A. (2025). Enhancing pilot training through virtual reality: a training module developed at ASL Airlines Belgium. *Master's thesis, HEC Liège*, University of Liège. http://hdl.handle.net/2268.2/22855

Howard, M. C., Gutworth, M. B., & Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior*, 121, 106808. https://doi.org/10.1016/j.chb.2021.106808

Jiang, M. (2024). Development of a virtual reality training application for high loader operators during aircraft loading. *Master's thesis, HEC Liège*, University of Liège. http://hdl.handle.net/2268.2/20147

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, 63(4), 706–726.

Makransky, G., & Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education*, 134, 15–30. https://doi.org/10.1016/j.compedu.2019.02.002

Ruche, C. (2025). Development and optimization of virtual training modules at Liège Airport Academy. *Master's thesis, HEC Liège*, University of Liège. http://hdl.handle.net/2268.2/22770

Scorgie, D., Feng, Z., Paes, D., Parisi, F., Yiu, T. W., & Lovreglio, R. (2024). Virtual reality for safety training: A systematic literature review and meta-analysis. *Safety Science*, 171, 106372.

XR Aircraft Dismantling Assistant for Aerospace Circularity

Gersa Tome, Bastian Fibi, and Richard Revesz

Austrian Institute of Technology, Austria

Corresponding author: gersa.tome@ait.ac.at

Keywords: Mixed Reality, Aircraft Dismantling, Circular Economy, Sustainable Aviation, Industrial Extended Reality, End-of-Life Aircraft Operations.

Introduction

The increasing number of aircraft reaching end-of-life (EoL) is creating a growing need for efficient and sustainable dismantling processes. Aircraft contain substantial quantities of valuable materials, with aluminium alloys accounting for more than 75% of aircraft net mass (Airbus, 2025). However, materials recovered during dismantling are often shredded and downcycled, limiting the potential for high quality material-specific recovery.

Efficient dismantling requires detailed knowledge of the individual aircraft. Throughout their service life, aircraft undergo numerous maintenance, repair, and modification activities. Relevant information is available in technical documentation but is typically distributed across multiple sources. During dismantling, operators must therefore combine this information with practical experience to determine how and where an aircraft should be dismantled.

Mixed Reality (MR) offers an opportunity to provide aircraft-specific information directly within the physical work environment. Rather than consulting separate documents and translating technical information to the physical aircraft, relevant information can be spatially aligned with the aircraft structure, supporting operators during dismantling activities.

Previous XR applications in aviation have primarily focused on maintenance, training, inspection, and procedural support (De Crescenzio et al., 2010; Eschen et al., 2018; Peng et al., 2022; Torrence & Dressel, 2022). XR has also been explored for disassembly planning and human–robot collaboration (Yang et al., 2023; Hoebert et al., 2024). However, the application of Mixed Reality (MR) to aircraft EoL dismantling remains comparatively unexplored, particularly in the context of supporting material recovery and enabling more circular EoL practices.

This work presents an MR-assisted dismantling application that connects aircraft-specific digital information with the physical aircraft. The application is designed to support operators in identifying material recovery areas and documenting the recovered materials during the dismantling process.

XR assisted dismantling solution

To address the challenges identified in aircraft EoL dismantling, the proposed MR-assisted dismantling application was developed in close collaboration with an industrial company specializing in aircraft EoL dismantling and material recovery.

The process begins by aligning the virtual and physical aircraft, leveraging the Magic Leap 2´s robust tracking capability and establishing the spatial reference required to display aircraft-specific information in the MR environment.

The application is developed in Unity, using the Universal Render Pipeline for mobile-optimized rendering and deployed on the Magic Leap 2 headset via the Magic Leap SDK. Interactions are implemented with Unity's XR Interaction Toolkit, and an OpenXR integration supports future portability to other headsets. Aircraft-specific 3D models are loaded at runtime via glTFast, with a WebSocket connection maintaining communication with backend services for aircraft data and recovery-section records.

Key information relevant to the dismantling task can then be presented as AR overlays directly on the physical aircraft, including material information, repair areas, and structural areas. This provides operators with task-relevant information directly within the work environment, addressing information requirements identified by operators as relevant to dismantling decisions but not typically readily available at the point of work.

The recovery request defines the material to be recovered from the aircraft fuselage and the requirements for the recovered section. These include the required surface area and, where applicable, whether windows should be included, as illustrated in Figure 1. By providing these requirements within the MR environment together with the aircraft-specific information, the application supports the operator in identifying suitable areas for material recovery.

Based on the recovery request and the visualized aircraft information, the operator can define cut zones directly on the fuselage. The AR overlays allow structural areas and repair zones to be considered when positioning the cut, while the spatial representation enables the proposed recovery section to be assessed from both the interior and exterior of the aircraft. Predefined dimensions reflect requirements for handling, logistics, and downstream remanufacturing, supporting the definition of recovery sections that can be efficiently processed after dismantling.

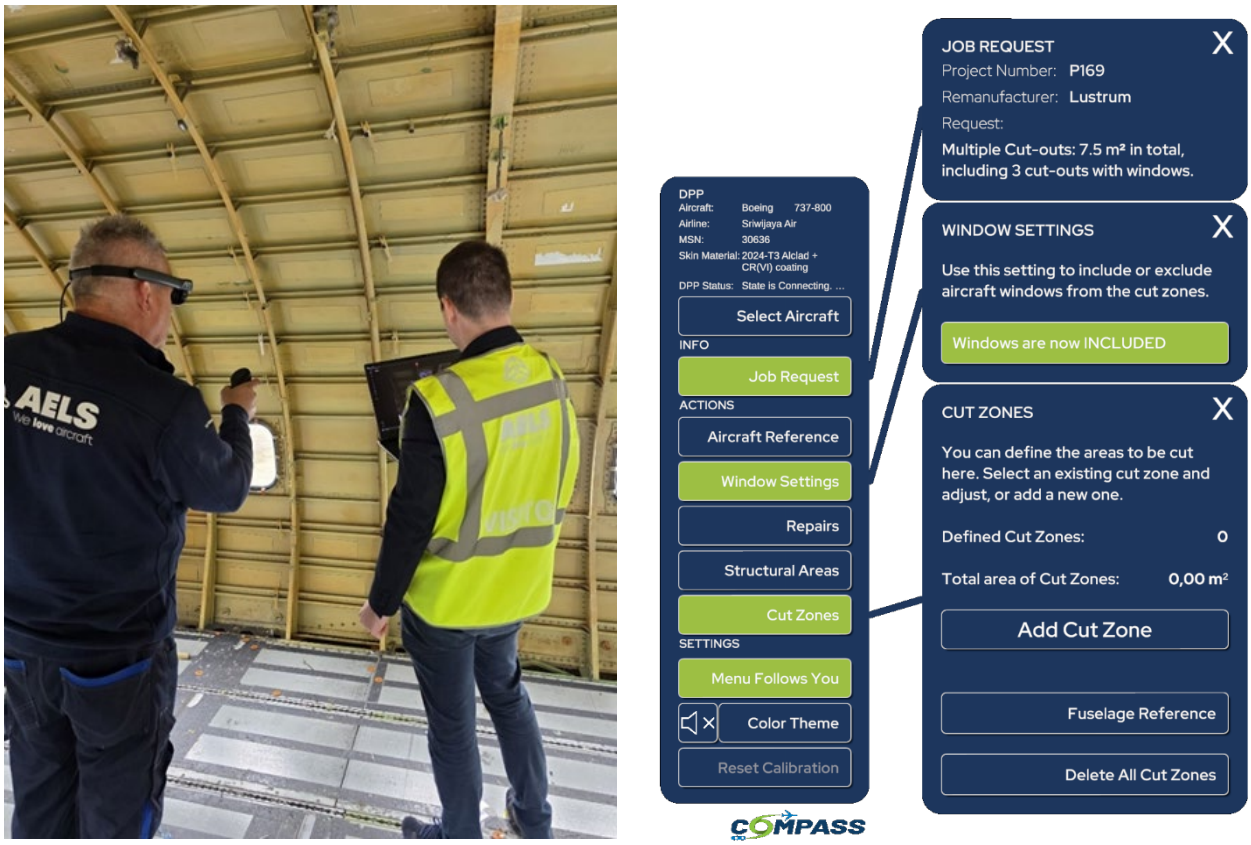


Figure 24. *The operator executing the recovery task while wearing the Magic Leap 2 glasses (left) and the expanded menu view showing the dedicated menus for a specific recovery request (right).*

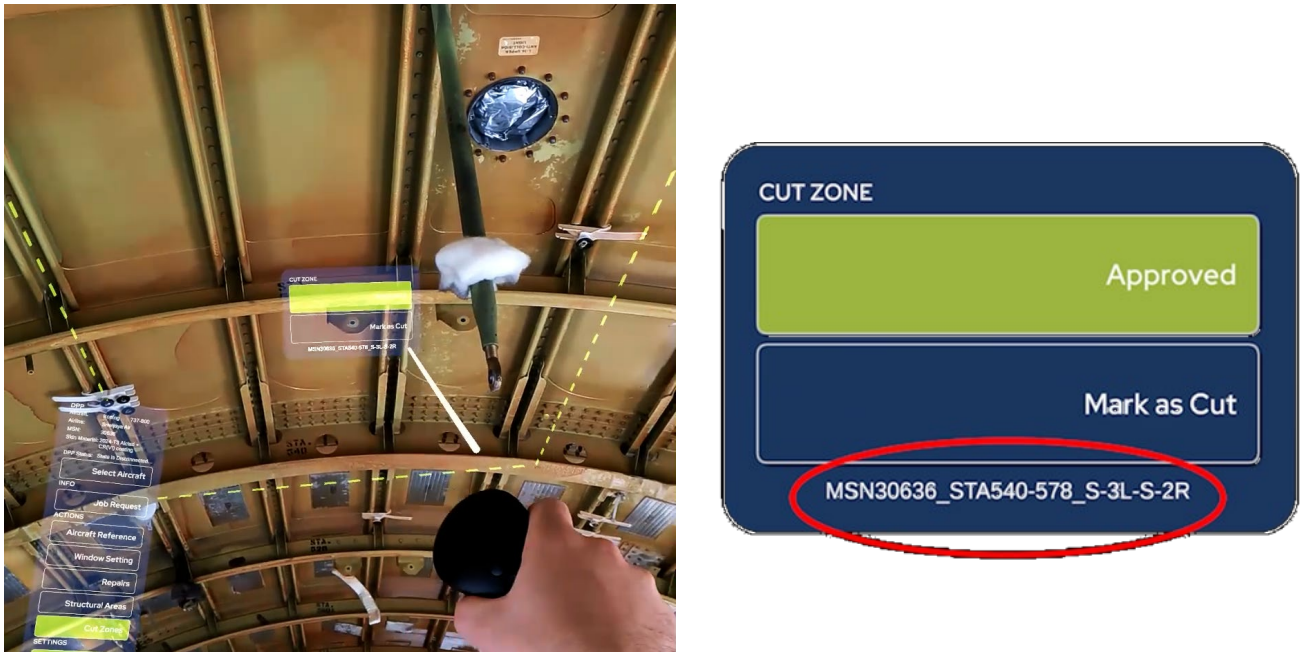


Figure 2. *Demonstration of the identified cut zone on the aircraft fuselage (left) and its associated identifier (right).*

Following the definition of a cut zone, the recovered section is assigned to a unique identifier linking it to its aircraft of origin and location. The identifier and associated information can be transferred to a downstream database, while a physical label can be printed and attached to the recovered panel. This maintains the connection between the physical material and its digital information after removal from the aircraft, enabling the recovered section to remain identifiable throughout subsequent remanufacturing and recovery processes.

Evaluation and Observations

The application was evaluated using an iterative qualitative research approach incorporating direct observation, semi-structured interviews, and demographic questionnaires, with data analyzed via qualitative thematic analysis. The study involved three dismantling experts (2M, 1F) in the first field study, five XR/HCI experts (4M, 1F) for usability, and two dismantling experts (2M) in the second field study. During these sessions, participants performed representative dismantling tasks, such as inspecting aircraft-specific information and managing cut zones (creating, modifying, or deleting) directly on the aircraft structure.

Participants valued the ability to access aircraft-specific information through spatial representation, repair areas, material information, and cut-zone visualization; all considered relevant to dismantling preparation. The traceability functionality was also considered useful because it provides a means of maintaining the identity and origin of recovered aircraft sections.

At the same time, the physical dismantling environment places important constraints on MR interaction. Operators work with protective equipment and heavy cutting tools and must continuously reposition themselves and their equipment. Furthermore, environmental limits such as elevated temperatures inside the fuselage, limited ventilation, and residual components in the workspace affected the operational conditions. Consequently, continuous interaction with the MR system during cutting is not practical. The findings therefore support the use of MR primarily during planning and preparation.

Conclusion

This work presented an MR-assisted aircraft dismantling application that connects aircraft-specific information with the physical aircraft to support material recovery and traceability at EoL. By linking recovery requests, cut-zone definition, and identification of recovered sections, the

application establishes a digital connection between dismantling and downstream remanufacturing. In doing so, the application supports circularity by enabling more targeted and traceable recovery of materials and components for subsequent reuse and remanufacturing.

Acknowledgements

This work was developed within the COMPASS project, co-funded by the European Commission under the Horizon Europe Programme, grant agreement No. 101136940.

References

Airbus. (2025). Global Market Forecast 2025–2044. Airbus.

De Crescenzio, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., & Salti, S. (2011). Augmented reality for aircraft maintenance training and operations support. *IEEE Computer Graphics and Applications*, 31(1), 96–101. https://doi.org/10.1109/MCG.2011.4

Eschen, H., Kötter, T., Rodeck, R., Harnisch, M., & Schüppstuhl, T. (2018). Augmented and virtual reality for inspection and maintenance processes in the aviation industry. *Procedia Manufacturing*, 19, 156–163. https://doi.org/10.1016/j.promfg.2018.01.022

Hoebert, T., Seibel, S., Amersdorfer, M., Vincze, M., Lepuschitz, W., & Merdan, M. (2024). A framework for enhanced human–robot collaboration during disassembly using digital twin and virtual reality. *Robotics,* 13(7), 104. https://doi.org/10.3390/robotics13070104

Peng, C.C., Chang, A.C., & Chu, Y.L. (2022). Application of augmented reality for aviation equipment inspection and maintenance training. *In 2022 8th International Conference on Applied System Innovation*, 58–63. IEEE. https://doi.org/10.1109/ICASI55125.2022.9774490

Torrence, B., & Dressel, J. (2022). c. In International Conference on Human-Computer Interaction, *Springer*. 270–288.

Yang, Y., Keivanpour, S., & Imbeau, D. (2023). Integrating X-reality and lean into end-of-life aircraft parts disassembly sequence planning: a critical review and research agenda. *The International Journal of Advanced Manufacturing Technology*, 127, 2181–2210. https://link.springer.com/article/10.1007/s00170-023-11581-0

AI-GLASS: Physics-Informed AI and Pixel-Streamed Digital Twin for Glass Manufacturing

Yang Zhang[1], Simon Campion[1], Giuseppe Maria Bellantuono[1], Aye Chan Zay-Hta[1], Craig Rothwell[1], Harry Ellis[1], Dzianis Litskevich[1], Hannah Whyatt[1], David Boardman[1], Jim Scotson[2], Aston Fuller[2], and Konstantin Vikhorev[2]

[1]The Virtual Engineering Centre (VEC), University of Liverpool, UK

[2]Glass Futures in partnership with NVIDIA

Corresponding author: Yang.Zhang@liverpool.ac.uk

Keywords: Digital twin; industrial AI; physics-informed AI; NVIDIA Omniverse; OpenUSD; pixel streaming; multi-user VR; sustainable manufacturing.

Introduction

Glass manufacturing is energy intensive, and producers are under pressure to evaluate hydrogen firing, electric boosting and alternative batch compositions without experimenting on a running furnace (Atzori et al., 2023). These changes act on heat transfer, combustion and melt behaviour in ways that are inherently spatial, which makes them natural candidates for immersive inspection.

The obstacle is practical rather than conceptual. A twin of this kind carries high-resolution laser-scanned geometry, CAD and BIM assets, live historian connections and volumetric simulation output. Distributing it to each XR client means moving large asset sets onto devices that cannot render them at framerate, cannot reach the plant historian, and cannot host the inference services the predictions depend on. Earlier EuroXR work has demonstrated the value of multi-user XR for digital-twin information and for collaborative industrial environments while reporting the same deployment difficulty (Aust et al., 2021; Uzun et al., 2023).

AI-GLASS is a collaboration between Glass Futures, the VEC and NVIDIA, centred on an NVIDIA Omniverse-based digital twin of Glass Futures' 165,000 sq. ft. pilot facility in St Helens, UK (Glass Futures, 2026). Its architecture keeps the engineering scene, historian connectivity, rendering and inference services server-side, while streaming rendered frames to browser and VR clients and returning user interactions. The contribution is not remote rendering itself, but the integration of immersive and non-immersive access to a common operational digital twin. The same server-side scene supports browser inspection and VR, with ONNX-served surrogate predictions incorporated into the interaction loop. Model validation is reported below; systematic end-to-end latency and task-level user evaluation remain future work.

Digital Twin Platform

The environment brings three kinds of content into one OpenUSD stage (Figure 1). Plant geometry comes from high-resolution Leica laser scanning registered against existing CAD and BIM

assets. Operational information is drawn from furnace instrumentation and the plant historian, covering temperature, pressure, flow, level and quality indicators across instrumented tags, each bound to the scene element it describes rather than presented on a separate dashboard. Predictive information comes from surrogate models trained on CFD results and operational plant data using NVIDIA PhysicsNeMo, exported to ONNX and served behind a REST interface.

The layered composition of OpenUSD matters here for a reason specific to industrial deployment. Site, building, furnace and equipment assets are revised independently as the facility changes, without rebuilding the scene. A pilot plant under continual research modification is not a static model, and the asset pipeline has to assume that.

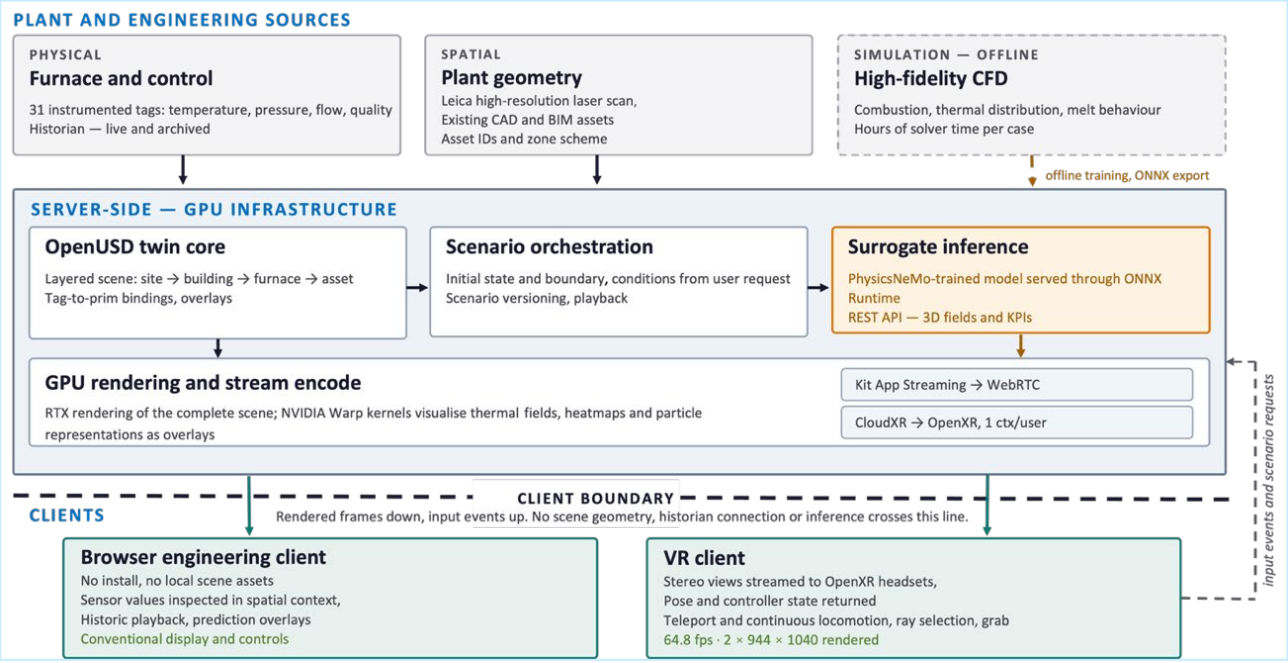


Figure 25. *AI-Glass digital twin architecture, from operational furnace data and spatial assets to CFD-trained surrogate prediction and streamed XR experiences.*

Pixel-Streamed Application and Immersive VR Experience

Two complementary delivery modes give users access to the digital twin without local high-end hardware. A representative engineering task is to inspect a furnace condition before a cross-site review. In the pixel-streamed browser application, an engineer selects a furnace hotspot, reads the associated live sensor value and time history, and compares its spatial location with a model-derived thermal field (Figure 2a). To assess a candidate change, the engineer submits an updated operating parameter, such as increased boost power, and the surrogate-predicted thermal field refreshes in place over the same scenario before a physical trial is considered. Heatmap and particle overlays provide spatial context, while the interface supports navigation and selection. Each session has one controlling user, and additional participants can join as spectators who see the same stream, which supports remote discussion of the scenario. If the controlling user disconnects, control passes to a spectator, so the session continues without restarting. The evaluated configuration used up to three spectator clients. The demonstrator supports inspection and advisory review. It does not issue automatic control commands.

The second delivery mode extends this remote-visualisation capability into an immersive VR experience using NVIDIA CloudXR pixel streaming (Figure 2b). A single user reviews the furnace condition in VR, receiving a rendered view driven by their own pose and controller input. A management agent launches application instances and assigns incoming connections to them. Several VR instances can run in parallel on the same server, limited by available GPU memory. Each instance serves one user, however, and state is not shared between instances. Shared multi-user

VR sessions are therefore outside the scope of this demonstrator, and the evaluated configuration used a single VR instance. Capacity planning assumes approximately 24 GB of GPU memory (VRAM) per streamed instance.

a) b)

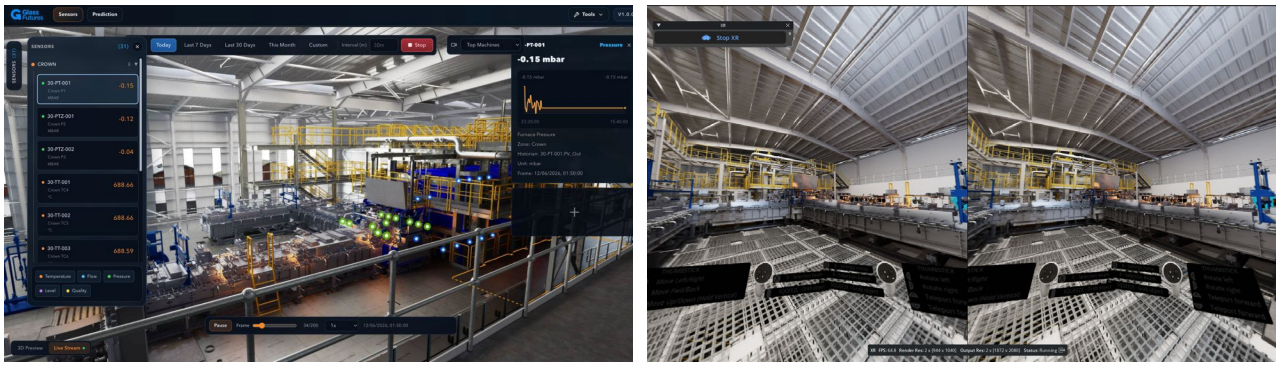


Figure 26. *The two pixel-streamed delivery modes for the AI-GLASS digital twin: (a) a browser application showing live sensor data overlay on the furnace scene; and (b) a stereo VR client view with controller mapping and streaming-status information. Both modes are rendered on the server and streamed to the client.*

Physics-Informed AI

AI-GLASS uses surrogate models trained offline on over 3,000 CFD cases spanning pull rates of 28-32 t/day, natural-gas/hydrogen blends of 0-100% H_2, electric boosting up to 1,200 kW, cullet fractions of 0-1, and Na_2O concentrations of 10-18 wt%. A DeepONet predicts combustion-space fields, POD-MLP models predict glass-melt fields, and an MLP ensemble predicts 15 scalar KPIs. Physics-based constraints improve consistency: a Joule-heating penalty reduced out-of-range rule violations from 79.9% to 7.7%, while mass-conservation correction enforces the specified pull rate. On held-out cases, R^2 reached 0.99 for combustion temperature, 0.92-0.99 for velocity components, and 0.6-0.99 across glass-melt variables. The KPI ensemble achieved a mean R^2 of 0.91 at unseen operating points, supporting use within the validated operating envelope rather than unrestricted extrapolation. GPU-based ONNX Runtime inference updates scenario overlays in under 5 seconds, compared with hours required for individual CFD simulations. Systematic end-to-end latency and user evaluation remain future work.

Results and Benefits

The demonstrator shows that a common industrial digital twin can support desktop and immersive interaction without maintaining separate representations. Centralised assets and rendering reduce client-side requirements and allow the same underlying twin to support users with different devices. However, pixel-streamed XR remains dependent on network conditions and introduces streaming latency. Systematic evaluation of latency, usability and concurrent deployment remains future work.

Conclusion and Future Work

AI-GLASS demonstrates a server-hosted industrial digital twin combining operational data, high-fidelity spatial assets and physics-informed surrogate models, delivered to browser and VR clients without local rendering of the full engineering scene. The current demonstrator supports spatial inspection, shared technical review and rapid scenario evaluation within validated operating conditions.

By reducing individual evaluations from approximately 5-12 hours of CFD computation to second-scale surrogate inference, the platform enables engineers to screen scenarios such as hydrogen firing and electric boosting before detailed simulation or physical trials. Future work will

extend model validation, assess streaming latency and collaborative XR workflows, evaluate decision quality and sustainability benefits, and investigate longer-term closed-loop optimisation

References

Atzori, D., Tiozzo, S., Vellini, M., Gambini, M., & Mazzoni, S. (2023). Industrial technologies for CO2 reduction applicable to glass furnaces. *Thermo*, *3*(4), 682-710. https://doi.org/10.3390/thermo3040039

Aust, M., Kvrgic, J.F., Frohnmayer, J., Otto, M. and Helin, K. (2021). Multi-User-XR for digital twin data in construction. In: Helin, K., Perret, J. and Arlati, S. (eds), EuroXR 2021: *Proceedings of the Virtual EuroXR Conference*. VTT Technology 395, pp.31–35. https://sarjaweb.vtt.fi/pdf/technology/2021/T395.pdf#page=34

Glass Futures. (2026). Glass Futures launches AI-driven digital to reinvent glass manufacturing. Available at: https://www.glass-futures.org/news/glass-futures-launches-ai-driven-digital-twin-to-reinvent-glass-manufacturing/ (Accessed: 3 September 2026).

Tran, K. Q., Leiss, E. L., Tsekos, N. V., & Velazco-Garcia, J. D. (2025, November). A Real-Time Immersive Augmented Reality Interface for Large-Scale USD-Based Digital Twins. In *Virtual Worlds* (Vol. 4, No. 4, p. 50). MDPI. https://doi.org/10.3390/virtualworlds4040050

Uzun, Y., Riedlinger, U., Buchholz, F., & Oppermann, L. (2023). Building Industrial Metaverse: Technical challenges of real-world collaborative XR application development. In: EuroXR 2023: Proceedings of the 20th EuroXR International Conference. VTT Technology 422, pp.56–59. https://www.researchgate.net/profile/Yuecel-Uzun/publication/376618433_Building_Industrial_Metaverse_Technical_Challenges_of_Real-_World_Collaborative_XR_Application_Development/links/658162f96f6e450f1986c00a/Building-Industrial-Metaverse-Technical-Challenges-of-Real-World-Collaborative-XR-Application-Development.pdf

Toward a Digital Gardener: An AI-Enabled XR Cross-Platform for Hydroponic Culture Inspection

Massimo Bortolamei[1], Giacomo Inches[1], Ivan Acampora[2], and Ciro Caputo[2]

[1]*Institute of Information Systems and Networking (ISIN), University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Lugano, Switzerland*

[2]*GreenVerse SA, Lugano, Switzerland*

Corresponding author: massimo.bortolamei@supsi.ch

Keywords: Mixed Reality, smart glasses, plant detection, hydroponics, computer vision, AI.

Introduction and state of the art

Domestic hydroponic systems can make small-scale cultivation more accessible, yet users must still combine visual inspection with data provided by smart sensors to recognize when an individual plant requires attention. We introduce a Digital Gardener Assistant in which a person wearing smart glasses looks at a hydroponic home cultivation tower: “Green Totem” by GreenVerse and immediately sees plant-specific information spatially aligned with the corresponding plant. Such an interface combines the plant's visually detected condition with measurements supplied by the tower, reducing the need to navigate a separate dashboard and helping non-expert users interpret the state of their cultivation.

Computer vision and deep learning have been widely investigated for image-based plant-disease detection and classification (Saleem et al., 2019). Extended Reality (XR) has also been explored in precision farming for data visualization, monitoring, and decision support, often in combination with sensors, computer vision, and localization technologies (Hurst et al., 2021). More broadly, reviews of XR in agriculture, livestock farming, and aquaculture describe the field as promising but still at an early stage of development (Anastasiou et al., 2023). In an XR application, detection is only one part of the problem: the system must access camera frames, run a compatible model, associate a detection with a physical tower slot, and present contextual information through the capabilities exposed by the target device. However, the availability and implementation of these capabilities vary across current headsets and smart glasses.

This paper reports on a prototype-based feasibility study of an AI-enabled XR assistant for Green Totem. Its contributions are: (1) an end-to-end Quest 3 reference prototype combining visual condition detection, spatial slot matching, and contextual Mixed Reality (MR) feedback; (2) exploratory implementations on wearable devices: Spectacles and Meta Ray-Ban Display, together with an assessment of XREAL Air 2 Ultra; and (3) cross-platform lessons for a future consumer smart glasses architecture. The scope of this study is the technical feasibility of the Digital Gardener Assistant concept and its interaction design. Agronomic diagnostic accuracy and user effectiveness will be addressed in subsequent validation work.

Prototyping the Digital Gardener

The implemented workflow processed a camera feed with a detector that classified basil targets as healthy or withered. The detected position was then matched to a known plant slot, allowing plant

identity and tower data to be presented at the corresponding spatial location. In this study, pH, humidity, light, and water-status values were simulated; integration with the existing Green Totem controller was outside the project scope.

Starting from the Segmented Basil Photos dataset (Tomtechnicus, 2023), we curated a new basil image collection and manually annotated 85 images with healthy and withered bounding boxes. Images used in the controlled tower demonstration belonged to the validation set and were excluded from the training samples. The physical Green Totem tower was used as the spatial structure. Manually positioned MR references represented expected plant slots, first enabling interface prototyping without detection and later serving as a ground-truth proxy for qualitatively assessing detection-to-slot matching.

We tested the workflow with four different devices that are state of the art for MR: Quest 3 (VR visor with passthrough capabilities), and three commercial smart glasses: Spectacles (developer edition), Meta Ray-Ban Display (March 2026) and XREAL Air 2 Ultra.

Quest 3 Reference Prototype

The Quest 3 prototype was developed in Unity 6.2 using the Meta XR SDK. A fine-tuned YOLOv8 detector was exported to ONNX and adapted for local GPU inference through Unity Sentis. After adjusting the post-processing pipeline, the application processed passthrough camera frames and displayed bounding boxes, detected condition, and confidence scores (Figure 1).

Two spatial-reference strategies were implemented. In anchor mode, every manually positioned plant reference created an independent spatial anchor that could be saved and restored across sessions. In tower-offset mode, the references were stored relative to a tower origin identified through a QR-code pipeline, allowing the plant layout to move with the tower. Once detection was introduced, each detected basil target was associated with the nearest reference. The interface then displayed expected condition, detected condition, matching distance, and simulated sensor values. This demonstrated the complete interaction concept and helped identify implementation considerations related to lighting, spatial-feature detection, QR-trackable lifecycle, and non-continuous QR pose updates.



***Figure 27. Quest 3 prototype**: (a) manual placement of a spatial plant reference; (b) AI-based plant detection and condition classification; and (c) contextual information with qualitative detection-to-slot matching.*

Wearable Smart Glasses Prototypes

Spectacles offered the closest approximation to the intended wearable MR experience. The prototype used Lens Studio and SnapML to run detection on-device within an optical see-through spatial interface. Since the available pipeline supported YOLOv7 but not the previously trained YOLOv8 model, YOLOv7 was trained with more epochs and a higher input resolution. However, detection remained variable, suggesting that a broader training set and further tuning would be beneficial. Spectacles nevertheless demonstrated the most relevant combination of smart glasses form factor, true spatial MR, and on-device AI inference.

The Meta Ray-Ban Display prototype followed a distributed architecture. An Android/Kotlin application used the Meta Wearables Device Access Toolkit to stream camera frames from the glasses to a connected smartphone. The phone converted the frames for YOLOv8 inference at up to 20 frames per second and displayed bounding-box feedback locally. Detection status was also sent as text to the glasses. Within the SDK workflow evaluated, feedback on the integrated display focused on text, while plant-aligned spatial overlays were not part of the tested configuration due to SDK limitations. The prototype therefore represents a consumer-oriented architecture in which lightweight glasses operate as wearable sensor and display while the phone performs inference and application logic.

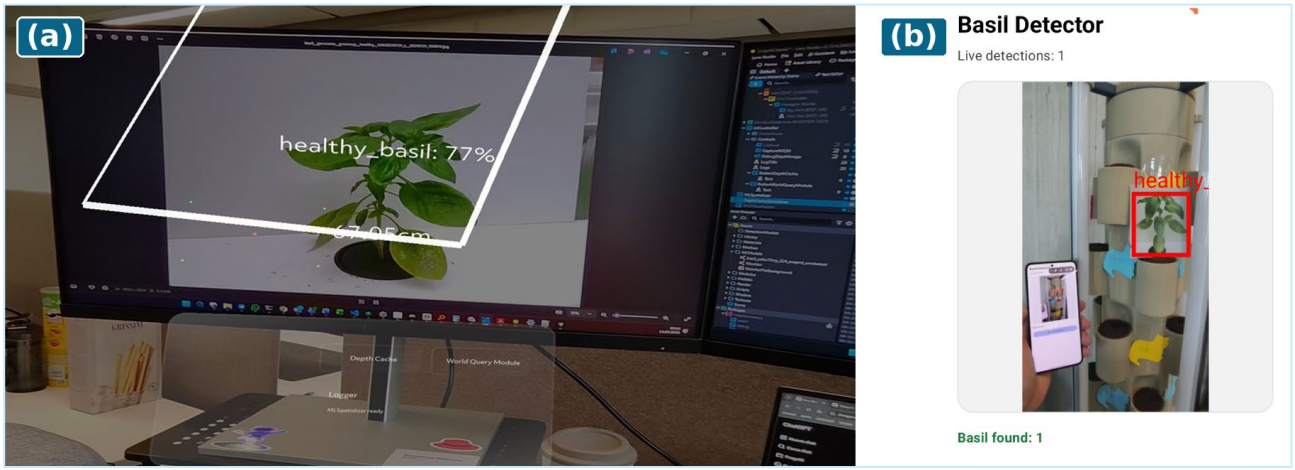


Figure 28. Wearable prototypes: *(a) on-device detection using Spectacles; and (b) the Meta Ray-Ban Display workflow with smartphone-based inference.*

XREAL Air 2 Ultra was assessed but could not be integrated into the evaluated vision pipeline because RGB camera-frame access was not available through the tested SDK. Table 1 summarizes the complementary evidence provided by the four platforms.

Table 4 *Cross-platform prototype comparison.*

Platform	AI execution	XR feedback	Spatial capability	Main findings
Quest 3	On-device GPU; YOLOv8/Sentis	Passthrough MR overlay	Anchors and QR-relative slots	Most complete end-to-end reference prototype
Spectacles	On-device; YOLOv7/SnapML	Optical see-through MR	Device-native spatial XR	Closest to intended wearable MR UX; further tuning needed
Meta Ray-Ban Display	Smartphone; YOLOv8	Text on glasses; bbox on phone	No spatial overlay in tested SDK	Consumer-like distributed architecture; up to 20 fps
XREAL Air 2 Ultra	Not possible	Not evaluated	RGB frames unavailable	Not integrated in the evaluated SDK configuration

Cross-Platform Lessons and Outlook

Across the evaluated configurations, the capabilities required by the Digital Gardener Assistant were distributed among the different platforms. Quest 3 provides the most complete spatial prototype and demonstrates how detection, slot matching, and contextual overlays can operate together. Spectacles are the closest to the intended user experience because they combine a smart glasses form factor, optical see-through MR, and on-device detection; broader training data and additional model tuning would support more consistent detection. Meta Ray-Ban Display suggests a

complementary near-term path based on consumer glasses and smartphone computation, with the evaluated workflow focusing on distributed inference and text-based feedback.

Across platforms, dataset quality and model portability emerged as central factors alongside detector choice. The need for greater image diversity is consistent with known challenges in image-based plant analysis (Upadhyay et al., 2025), while device-specific runtimes required different model generations and deployment pipelines. Camera-frame access, inference location, feedback modality, and SDK maturity should therefore be treated as architectural requirements rather than implementation details. The manually positioned references also showed the value of separating interface development from visual detection: the contextual experience could be designed first and then connected to progressively more capable perception components.

The present work was a technical feasibility study conducted to assess the maturity of the wearable technology towards a full implementation of the Digital Gardener Assistant concept. A controlled physical-tower setup using validation images enabled a repeatable qualitative assessment of the spatial interaction pipeline, while simulated sensor values kept the focus on perception and interaction. Future work should evaluate the system with grown plants and domestic users, expand the dataset, integrate live controller data, and quantify detection and spatial-matching performance, with close attention to the technological evolution of smart glasses and new SDK releases. Such evaluation should also investigate quantitatively whether plant-aligned feedback helps non-expert users understand and maintain their tower.

Acknowledgments

This work was supported by Innosuisse (Swiss Innovation Agency) with grant No. **134.281 INNO-EE.**

References

Anastasiou, E., Balafoutis, A. T., & Fountas, S. (2023). Applications of extended reality (XR) in agriculture, livestock farming, and aquaculture: A review. *Smart Agricultural Technology*, *3*, 100105. . https://doi.org/10.1016/j.atech.2022.100105

Hurst, W., Mendoza, F. R., & Tekinerdogan, B. (2021). Augmented reality in precision farming: Concepts and applications. *Smart Cities*, *4*(4), 1454-1468. https://doi.org/10.3390/smartcities4040077

Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. *Plants*, *8*(11), 468. https://doi.org/10.3390/plants8110468

Tomtechnicus. 2023. Segmented Basil Photos. Kaggle. https://www.kaggle.com/datasets/tomtechnicus/segmented-basil-leafs

Upadhyay, A., Chandel, N. S., Singh, K. P., Chakraborty, S. K., Nandede, B. M., Kumar, M., ... & Elbeltagi, A. (2025). Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture. *Artificial Intelligence Review*, *58*(3), 92. https://doi.org/10.1007/s10462-024-11100-x

Immersive Collaborative VR for Astronaut Training to Mars missions

Mariano Guida, and Paolo Leoncini

CIRA Italian Aerospace Research Center

Corresponding author: m.guida@cira.it, p.leoncini@cira.it

Keywords: Virtual Reality, training, astronauts, Mars, haptic gloves, Collaborative Virtual Environments

Abstract

Virtual Reality has evolved into a mature interface technology capable of creating realistic, synthetic reconstructions of operational settings. Today, it can place astronauts inside highly accurate simulations of planetary environments like Mars. As part of a space exploration technology initiative, the Italian Aerospace Research Center (CIRA) VR team developed an immersive, multi-user VR simulator. By utilizing cutting-edge VR technology, this system enables astronauts to acclimate to extraterrestrial surfaces and begin training routines.

Ultimately, this work highlights how a collaborative, haptics-enabled VR platform can assume a growing role in comprehensive astronaut training (Bosch Bruguera et al. 2019, Garg et al. 2024, Olbrich et al. 2018, Miele et al. 2015, Gagg et al. 2018, Harris et al. 2017, Smith et al. 2019, Rosenthal et al. 2021, Kennedy et al. 2022).

Introduction and Objectives

Preparing astronauts for missions to Mars presents unique challenges that align perfectly with the **DICE** framework, scenarios that are **D**angerous, **I**mpossible, **C**omplex, or **E**xpensive to recreate physically (Bailenson et al. 2024). While Space Agencies construct specialized analog facilities, these environments are costly and difficult to reconfigure as space hardware designs evolve. To address this, the VR team at the **Italian Aerospace Research Center (CIRA)** developed an immersive, multi-user, haptic-assisted Virtual Reality (VR) simulator. The system is designed for the preliminary stages of astronaut training, helping trainees:

- Familiarize themselves with the harsh, exotic environments of alien planets.
- Acquire critical manual skills through collaborative, physics-based interaction.
- Test habitability and reachability based on realistic body sizes.

This paper details the practical experiences and key design decisions made by our virtual reality research laboratory while developing the Mars simulation environment intended to help astronauts train and adapt to extraterrestrial conditions.

System Architecture and Multiplayer Infrastructure

Recent research highlights the strong value of multi-user collaboration in **Collaborative Virtual Environments (CVEs)**, particularly for familiarizing crews with exotic surroundings, building situational awareness, and practicing procedures together through shared feedback (Doroudian 2025).

The simulator is built using Epic Games' **Unreal Engine 5**. It allows two co-located users to interact seamlessly in the same physical and virtual space without physical collisions or cable constraints (**Figure 1**).



Figure 29. *Collaborative sessions in the CIRA VR Lab.*

For realistic **astronaut training**, standard generic puppets are insufficient; credible male and female astronaut avatars are required to maintain high immersion. Achieving full self and reciprocal **embodiment** requires precise **full-body tracking** so avatars match participants' real physical movements accurately. Furthermore, this tracking must be synchronized and coherent across all users sharing the same physical space, preventing collisions and enabling seamless coexistence. Our solutions to all these aspects are provided below.

- **Co-registered Tracking (Map Sharing):** To prevent users from bumping into each other, the system utilizes **HTC VIVE's Location Based Entertainment (LBE)** "map sharing" feature. Headsets map the room via inside-out Visual **SLAM** and share a single spatial map, ensuring virtual avatars exactly match the players' real-world physical boundaries.
- **Full-body inside-out tracking:** By standardizing both the VR headsets and body trackers (feet, hips) on the same inside-out tracking technology and a shared spatial map, the system eliminates the need for co-registration, ensuring high data coherence and fast, reliable setup for single and multi-user sessions.
- **Wireless PCVR (Wi-Fi 6e):** To maintain maximum freedom of movement while preserving high graphical fidelity, the simulator streams data wirelessly from dedicated VR-ready PCs to the headsets via Wi-Fi 6e.
- **Haptic feedback gloves:** Besides the baseline interaction through handheld controllers, the manual interaction of a user with the planet surface environment is based on one pair of **WEART TouchDIVER Pro gloves** per participant, providing threefold haptic feedback, i.e. tactile, grasp haptic force, and temperature.

Hand-object interaction is generally ruled by the Unreal built-in rigid-body dynamics simulation.

The developed multiplayer VR Simulator

The developed immersive multiplayer simulator of a human installation on Mars provides a **high-fidelity graphics** experience considering realistic lighting parameters and environmental characteristics (**Figure 2**).



Figure 30. *Third person view of the developed Mars VE with two astronaut users.*

The VR simulator is a PC VR setup for an Unreal-based application, leveraging SteamVR middleware interface to VR devices and the VIVE Business Streaming middleware application to remotely control the VIVE Focus 3 or XR Elite headsets in wireless streaming mode. The system is designed to be scalable on available hardware (Table 1):

Table 5. *Operating modes for the CIRA Mars VR Simulator.*

Feature	Basic Configuration	Advanced Configuration
Primary Inputs	Handheld controllers	Haptic gloves + Wrist/Body trackers + Motorized shoes
Leg Tracking	None (Procedurally simulated)	Direct (3x VIVE Ultimate Trackers)
Avatar Movement	Blended poses inferred from HMD speed/height	Real-time Inverse Kinematics (IK)
Interaction Style	Trigger button / proximity sensors	High-fidelity, physics-based grasping

A configuration for two players includes a server running on one of the computers on the LAN, and a client instance per player, each running on a dedicated VR-ready PC.

A male and a female avatar allows users to see their own bodies in the virtual environment. Users can move freely around the environment, interact with objects, perform reachability and feasibility tests, and perform habitability tests using their own body size.

In basic mode, when full-body or detailed hand tracking is unavailable, the system uses an Unreal Engine tool to bilinearly blend pre-computed mannequin poses, synthesizing natural-looking intermediate movements from minimal tracking data.

The VR simulator can be used with a variety of sensor configurations including handheld controllers, full-body tracking, and haptic gloves with wrist trackers.

In advanced configuration, three VIVE Ultimate Trackers, which use inside-out technology for 6-DoF peripheral tracking, are attached to the feet and to the hip of a user for full-body tracking, allowing his/her avatar to directly reproduce the user's walking movements by Unreal's internal IK on the avatar standard skeleton.

Both configurations can be set to use one or two TouchDIVER gloves whereas tactile, grasp and/or thermal feedback are required and/or a motorized pair of VR shoes, developed by the British start-up Freeaim, for walking across infinite virtual distances.

Haptic Integration

A key differentiator of CIRA's simulator is the integration of WEART TouchDIVER Pro haptic gloves. Whilst the movement of the hands in the base configuration is limited to grasping and pointing with the index finger, using the headset controllers, through the integration of the haptic gloves, it is possible to calculate the movements of each phalanx, transfer them to their virtual counterparts in the 3D scene and replicate them on the server, allowing users to see the movements of their own hands in relation to one another.

Advanced Haptic and Thermal Feedback

The Weart TouchDIVER Pro gloves consist of a series of finger caps, one for each finger. These caps incorporate actuators that provide tactile sensations, such as vibration and pressure, as well as a Peltier cell that provides rapid thermal feedback within a temperature range of 18°C to 42°C (Aurilio et al. 2019). The five finger caps are connected to a hub on the back of the hand via flexible corrugated strips. A further disc-shaped element extends from this hub to provide vibro-tactile stimulation to the palm (Figure 3).

Wireless Range Extension: The gloves' default Bluetooth connection struggled with the lab's physical dimensions. The team solved this by plugging the Bluetooth dongles into a USB-C hub attached directly to the standalone VR headset. Using the VirtualHere USB-over-WiFi client-server utility, raw Bluetooth data is streamed to the host PC, ensuring the transmitter remains within centimeters from the user's hands.



***Figure 31**. The Weart TouchDIVER Pro gloves.*

Grasp Capabilities

The WEART TouchDIVER Pro is a second-generation haptic glove. Instead of traditional exoskeletons that use physical mechanical brakes or pulling forces to stop fingers when grasping a virtual object, it relies on a bio-inspired approach developed by researchers at the University of Siena: skin indentation (Spagnoletti et al. 2018).

When a user grasps a virtual object, the glove compresses the user's fingertips in proportion to how deeply the virtual fingers penetrate the object's surface. This pressure tricks the brain into feeling mechanical resistance. The mild discomfort naturally encourages the user to loosen their grip, preventing them from over-tightening without risking real physical harm (**Figure 4**).



Figure 32. *Correspondence between the hand wearing the Weart TouchDIVER Pro glove and the virtual avatar's hand*

The integration of thermal feedback is particularly novel. In extra-habitat, extra-vehicular activities (EVAs) on Mars, temperature changes indicate safety hazards or the operating status of machinery. Providing realistic thermal cues directly enhances an astronaut's environmental awareness (Figure 5).



Figure 33. *The Weart TouchDIVER Pro gloves grasping capabilities.*

Locomotion in Place

In both base and advanced configurations, the VR app can allow users to walk across infinite virtual distances while remaining stationary. Developed by the British start-up Freeaim, the motorized VR shoes implemented ensure that, when a user steps forward and shifts their weight, motorized rollers pull them back to their starting position. To ensure safety and prevent early sensations of slipping, the system includes a circular floor mat and a fold-out walking frame for stability (Figure 6).



Figure 34. *The app whilst using the Freeaim VR Shoes and the user inside the walking safety frame.*

From a technical perspective, the shoes track translational movement and yaw (heading), combining this data with the VR headset's 6-DoF tracking. Integration into custom Unreal Engine projects is streamlined because Freeaim developed a custom SteamVR driver to translate physical walking data (speed, stride, direction) into standard OpenXR joypad inputs. The hardware communicates wirelessly via Bluetooth to a PC, mapping physical steps directly onto standard smooth locomotion axes, effectively mirroring a traditional thumbstick input. Within the Unreal Engine Character Blueprint, this input directly drives the standard Add Movement Input node to move the virtual avatar.

Finally, user comfort and movement behavior can be dynamically adjusted externally using either the OpenVR/SteamVR overlay menu or the Freeaim Mobile App. Key configuration options include Speed Scaling, which adjusts the ratio between physical steps and virtual speed, and Speed Smoothing & Deceleration, which controls how gradually the character stops to eliminate jarring camera movements and reduce simulator sickness (Figure 7).

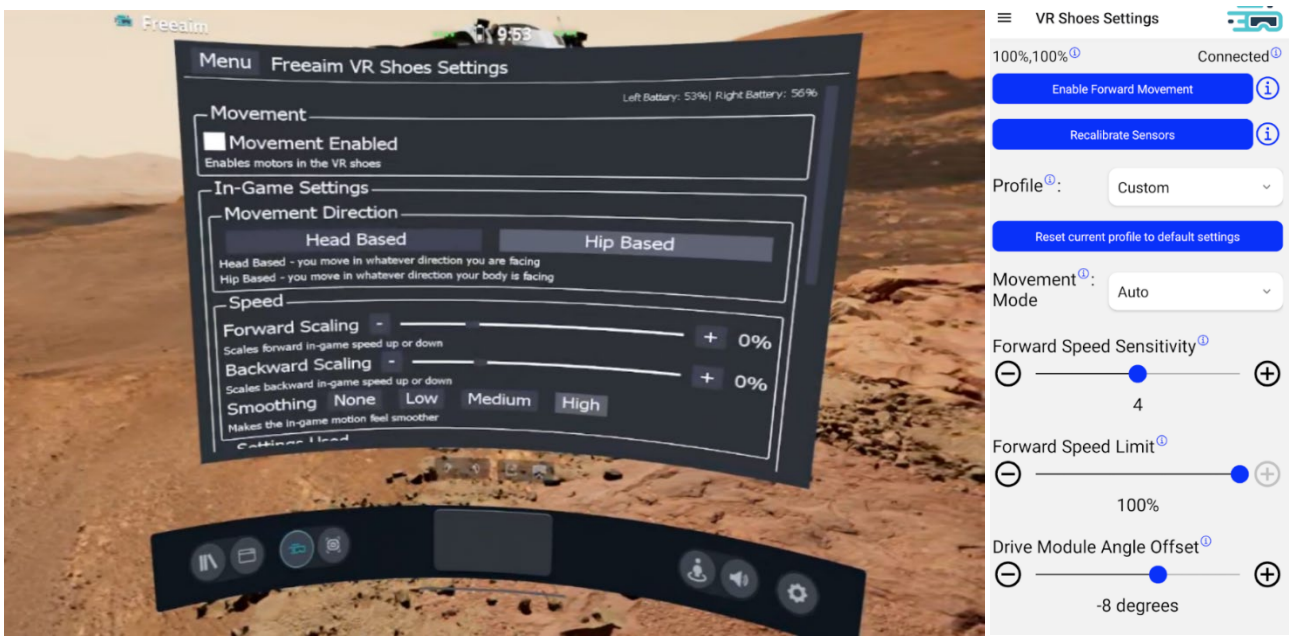


Figure 35. *The OpenVR/SteamVR overlay menu and the Freeaim Mobile App.*

Innovation and results

The main innovation of the proposed VR simulator lies in combining haptic and thermal feedback within a physics-based, multi-user Collaborative Virtual Environment (CVE) designed for astronaut training. Our system allows co-located, fully immersed users to move untethered while maintaining the high performance of a PC-VR setup.

Development of the simulator for planetary exploration and astronaut familiarization is progressing rapidly. It is nearly ready for human subject evaluations with real astronauts, which we aim to conduct in partnership with space agencies responsible for astronaut training.

Thermal sensation remains one of the least explored sensory channels in virtual reality, primarily due to a shortage of practical hardware. By integrating WEART gloves into our platform, we can now conduct user trials and subjective assessments to measure the concrete benefits of thermal feedback for astronaut training. Building on foundational research [19] from the developers of the TouchDIVER gloves, we anticipate that hand-perceived temperature cues will provide vital safety indicators, particularly during Extravehicular Activities (EVAs), where thermal management is critical, as well as immediate tactile confirmation of equipment operational status during handling.

Limitations and Future Work

While the simulator is ready for getting feedback from real astronauts, several areas of development remain:

- **Sensory Enhancements:** Incorporating whole-body collision feedback, ground texture feel under boots, and positional audio.
- **Suit Simulation:** Integrating realistic (or mock-up) weighted space suits to simulate the restrictive mass and inertia astronauts experience during extra-habitat activities.
- **Reduced Gravity Walk:** To simulate Mars gravity, the team plans to combine a passive/active **gravity offloading suspension system** with the already implemented **Freeaim VR Shoes**. This setup will allow astronauts to walk indefinitely in place with reduced body weight, preventing falls through the harness safety stops.

Acknowledgment

This work was performed within the CIRA TEDS research program included in the Italian National Aerospace Research Program (PRO.R.A.), funded by the Italian Ministry of Research (MUR). Authors thank the project manager Nunzia Favaloro for her continued support.

References

Aurilio, M., Peccerillo, B., Spagnoletti, G., & Prattichizzo, D. (2019). Using Wearable Haptics for Thermal Discrimination in Virtual Reality Scenarios. *Haptic Interaction, Kajimoto*, H., Lee, D., Kim, S-Y., Konyo, M., Kyung, K-U., (eds.), pp. 144-148, Springer SingaporeBailenson, J. N., & Lesher, M. (2024). Virtual reality and its opportunities and risks. In *OECD Digital Economy Outlook* (Vol. 1). Embracing the Technology Frontier, OECD Publishing.

Bosch Bruguera, M., Ilk, V., Ruber, S., & Ewald, R. (2019). Use of Virtual Reality for astronaut training in future space missions – Spacecraft piloting for the Lunar Orbital Platform - Gateway (LOP-G). 70th International Astronautical Congress, Washington D.C., United States.

Doroudian, S. (2025). Collaboration in Immersive Environments: Challenges and Solutions. https://doi.org/10.48550/arXiv.2311.00689 [cs.HC].

Freeaim VR Shoes product website [www.freeaim.com].

Gagg, L., & Karr, T. (2018). Virtual Reality in Astronaut Training: A Review of Current Applications and Future Prospects. *International Journal of Aerospace Engineering*, 6(2), 25-34.

Garg, V., Singh, V. & Soni, L. (2024). Preparing for Space: How Virtual Reality is revolutionizing Astronaut Training". *IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE)*. Bangalore, India.

Harris, C., & Pritchard, D. 2017. Immersive Training Systems for Space Missions: A Virtual Reality Approach. *Simulation in Aerospace*, 14(1), 59-68.

Kennedy, B. & Dawson, P. (2022). From Earth to Space: Virtual Reality Training for Astronauts. *Aerospace Research Review*, 15(1), 78-91.

Miele, A., & Scolaro, A. (2015). Virtual Reality as a Tool for Astronaut Training: Enhancements and Future Directions. *Journal of Space Safety Engineering*, 2(3), 115-123.

Olbrich, M., Graf, H., Keil, J., Gad, R., Bamfaste, S. & Nicolini, F. (2018). Virtual Reality Based Space Operations – A Study of ESA's Potential for VR Based Training and Simulation. VAMR 2018, LNCS 10909, pp. 438–451.

Rosenthal, L. & Stewart, R. (2021). Virtual Reality and Haptic Feedback in Astronaut Simulation Training. *Journal of Space Technology and Applications*, 5(3), 40-52.

Smith, J., & Green, H. (2019). Application of Virtual Reality to Astronaut Space Mission Training. *Space and Human Factors Journal*, 12(4), 203-214.

Spagnoletti, G., Meli, L., Lisini Baldi, T., Gioioso, G., Pacchierotti, C., & Prattichizzo, D. (2018). Rendering of Pressure and Textures Using Wearable Haptics in Immersive VR Environments. IEEE Conference on Virtual Reality and 3D User Interfaces, Pages 691-692, Reutlingen, Germany.

VirtualHere USB over IP/WiFi product website [www.virtualhere.com].

VIVE Ultimate Trackers product website [www.vive.com/us/accessory/vive-ultimate-tracker].

VIVE Wrist Tracker product website [www.vive.com/us/accessory/vive-wrist-tracker].

WEART TouchDIVER VR Gloves product website [weart.it].

Mobile XR-based C2 for better situation awareness – Field trials in Bulgaria

Kaj Helin, Jaakko Karjalainen, Andrea Alesani, and Laura Salmela

VTT Technical Research Centre of Finland Ltd

Corresponding author: kaj.helin@vtt.fi

Keywords: Extended Reality (XR), Command and Control (C2), Situational Awareness, Border Surveillance, Human-Machine Interaction, Field Trials

Introductions

Border surveillance operations increasingly rely on distributed sensing, mobile command-and-control, and rapid interpretation of heterogeneous information. The BorderForce system addresses these challenges by combining multimodal data fusion, reactive risk assessment, resilient communications, XR-enhanced situational awareness, and collaborative mission planning, integrating sensor observations, OSINT information, risk indicators, and operational visualisations into a common operational picture. This abstract presents the development and field trial of a Mobile XR-Based Command and Control (C2) tool that enables commanders and field personnel to monitor border surveillance operations through an interactive 3D map environment, visualising assets, streaming, detections, tracks, and other mission-relevant information received from the BorderForce backend in real time to support improved situational awareness and decision-making.



Figure 36. *Field trial set-up in Bulgaria with XR C2, several assets, cameras and drones*

The field trials in Bulgaria were organised as part of the BorderForce operational testing week in Stara Zagora, at the Central Artillery Technical Testing Range of the Bulgarian Ministry of Defence, with activities executed from 29 June to 3 July 2026. The test programme included hardware and software setup, integration, and system testing with a use case scenario (see Figure 1).

Mobile XR-Based C2 Tool

The BorderForce Mobile XR-Based C2 Tool was developed to provide field personnel with an immersive situational awareness capability for border surveillance operations (see Figure 2). The application enables users to observe and interact with operational information through a spatial 3D representation of the mission area, allowing complex information to be understood more intuitively than through conventional 2D interfaces.

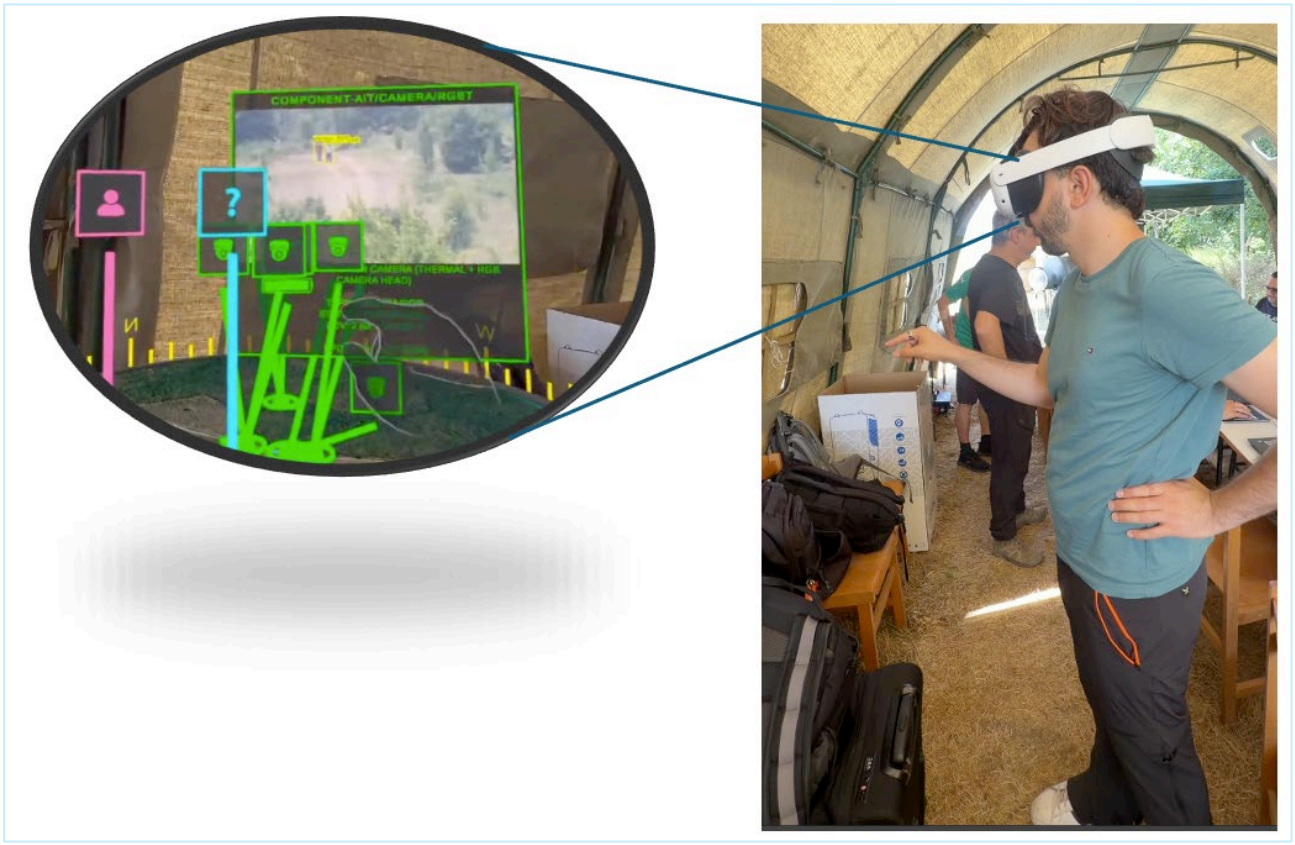


Figure 37. *The Mobile XR-Based C2 Tool during trial*

The XR application operates as a client connected to the BorderForce backend ecosystem. Operational information originating from sensors, intelligence services, situational awareness modules, risk assessment components, tracking systems, and command-and-control services is distributed through the BorderForce communication infrastructure and delivered to the XR application using the MQTT messaging protocol. This architecture allows the XR tool to receive near real-time updates from multiple BorderForce subsystems and present a unified operational picture to the user. In addition, it can visualise recorded data for playback and analysis of past scenarios, as well as synthetic data that simulates realistic movements and behaviours of different operational assets.

The solution was implemented using the Unity game engine and OpenXR framework, enabling hardware-independent deployment and future portability across different XR devices. The hardware configuration used during the field trials consisted of a standalone Meta Quest 3 XR headset connected wirelessly to the BorderForce backend services. The standalone headset configuration also supports mobile deployment in operational environments without the need for external computers or wired connections.

The main functionality of the Mobile XR-Based C2 Tool is the visualisation of the operational area through an interactive 3D map interface. Users can navigate the environment naturally by changing scale, perspective, and viewing position while maintaining awareness of the geographical context of ongoing operations. Information received from the backend is represented as visual objects within the 3D environment, including operational assets, camera streaming (UxVs, various camera types), sensor detections, tracks, alerts, and other mission-relevant entities. The application continuously updates these objects as new information is received through MQTT communication channels, enabling users to monitor events and developments in near real time (see Figure 3).

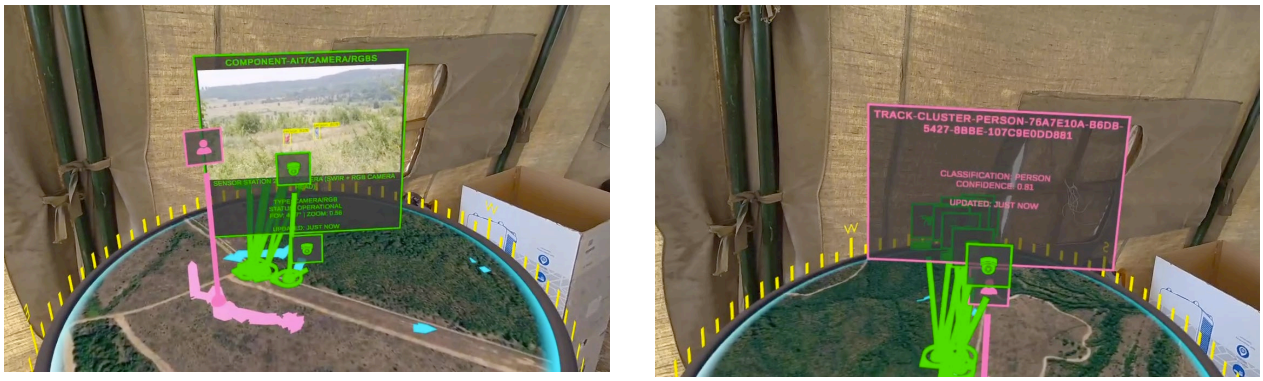


Figure 38. *Left: 3D map showing tracks, detections, assets and video streaming*
Right: Showing additional data when icon clicked

Field Trial and Use Case Scenario

The Mobile XR-Based C2 Tool was evaluated during the BorderForce operational testing week held in Stara Zagora, Bulgaria, from 29 June to 3 July 2026, where BorderForce technologies were integrated and tested in realistic field conditions.

The demonstration scenario started from OSINT-based intelligence indicating possible illegal border crossing activity in a monitored area. Based on this information, unmanned aerial systems were deployed to investigate the potential threat and provide aerial surveillance of the area. The drones searched for suspicious activity and transmitted situational information to the BorderForce backend system. Additional detections were generated by ground-based surveillance cameras, which identified and tracked several individuals moving within the operational area. The information from drones and cameras was fused within the BorderForce platform and distributed to operational users.

The Mobile XR-Based C2 Tool received the real operational data and visualised scenarios, detected persons, tracks, assets, and events within an interactive 3D map environment. This allowed operators to monitor the evolving situation in real time, analyse the location and movement of detected individuals, and maintain awareness of the operational picture from different perspectives. Based on the detections and tracking information, command-and-control actions could be planned and coordinated using a common spatial view of the situation.

Evaluation and Outcomes

The evaluation of the Mobile XR-Based C2 Tool was conducted through direct observation and expert assessment. An end user operated the system during the operational scenario, while other end users, domain experts, and developers followed the demonstration through live observation and via recorded video. Feedback was collected during and after the exercise to assess usability, situational awareness support, technical performance, and overall user experience.

Overall, the system performed reliably throughout the trial and all planned functionalities operated successfully. The evaluators considered the concept intuitive and easy to understand, allowing users to quickly navigate the 3D operational environment and interpret information from

multiple surveillance sources. The immersive visualisation was regarded as a useful way to improve understanding of the operational situation and to support command-and-control activities.

The evaluation also identified several usability improvements. Users suggested reducing the visibility of connection lines when zooming in, as they can become distracting. Navigation controls should follow the current map orientation rather than fixed directions for more natural interaction. Evaluators also recommended displaying the sensor source of each detection when available and using clearer visual differentiation between drones and cameras to improve interpretation of detection information.

Conclusion and next steps

The evaluation results indicated that the system was reliable, intuitive to use, and capable of providing a clear and shared operational picture for end users. The immersive visualisation helped users better understand the spatial relationships between sensors, detected targets, and operational assets, supporting decision-making during the demonstration scenario. Expert feedback also identified several usability improvements, providing valuable guidance for future development iterations. These updates will be implemented before the upcoming large-scale operational trials at real border environments in Bulgaria (Nov -26) and Lithuania (Feb -27), where the system will be further validated with end users under realistic operational conditions.

Overall, the study demonstrates that mobile XR technology can effectively complement traditional command-and-control systems by improving the visualisation and interpretation of complex operational information. The results highlight the potential of XR-based situational awareness tools for future border security and surveillance applications (Dennison et al, 2020 & Conceição et al, 2022).

Acknowledgment

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101168316 (BorderForce).

References

Conceição, V., Tsigopoulos, A., Galanis, G. & Toivonen, S. (2022). ARESIBO – Augmented Reality Enriched Situation Awareness for Border Security: Evaluation Framework and Results of Large-Scale Trials. 4.º Encontro de Investigação e Desenvolvimento em Ciências Militares, 1–7.

Dennison, M., Reardon, C., Gregory, J., Trout, T. & Rogers, J.G. (2020). Creating a mixed reality common operating picture across C2 echelons for human-autonomy teams. Proceedings of SPIE, Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations, 11426, 1–12. https://doi.org/10.1117/12.2556642

Training the Mission, Not Only the Aircraft: A Co-Designed VR Simulator for UAV-Assisted Avalanche Search and Rescue

Maëlan Poyer, Iris David Du Mutel de Pierrepont Franzetti, Laura Battistel, Lorenzo Stefani, Matteo Mirafiori, and Riccardo Parin

terraXcube, Eurac Research, 39100 Bolzano, Italy

Corresponding author: riccardo.parin@eurac.edu

Keywords: Virtual reality, UAV, avalanche, search and rescue, performance assessment

From flight control to mission competence

Uncrewed aerial vehicles (UAVs) support search and rescue (SAR) by extending visual and thermal observation over terrain that is difficult or unsafe for ground teams (Lyu et al., 2023). In avalanche response, however, the remote pilot does more than control an aircraft. The pilot must interpret an incomplete briefing, plan and document a search, select sensors, maintain communication, and coordinate with crewed aviation under time pressure. Field exercises remain indispensable, but weather, cost, logistics and safety restrict how often rare critical events can be reproduced.

Current UAV simulators mainly address flight proficiency. Structured simulator pre-training has improved subsequent real-flight accuracy, including a 32% reduction in mean final displacement in a controlled study (Somerville et al., 2024). Yet incident analysis shows that RPAS safety occurrences also involve skill-based, perceptual and decision errors (Murray, et al.., 2025). The resulting gap is not another manoeuvre trainer, but a way to elicit and observe mission behaviour without exposing people or aircraft to failure. Recent avalanche-rescue research found that VR can support knowledge acquisition comparable to field instruction, while confirming that physical equipment practice remains necessary Knierim, et al., 2025). Presence and agency can facilitate immersive learning (Makransky & Petersen, 2021), but neither perceived realism nor immersion alone demonstrates operational competence.

We therefore developed a co-designed VR simulator for an end-to-end UAV-assisted avalanche SAR mission. The contribution is an assessment-ready scenario that couples task-specific digital models, familiar controls, controllable environmental demands and standardized safety events. Its first formal use is planned for the Pilot Day on 25 September 2026, where experienced rescue pilots will complete the mission without being trained to a prescribed solution. The aim is to characterize current operational skills and identify needs from which targeted training can subsequently be built.

Co-designing with the rescue community

Requirements were collected during two half-day workshops at terraXcube and NOI Techpark on 3 and 16 April 2025, followed by the 2025 Pilot Day in Belluno. Approximately 35 participants represented Croce Bianca, Bergrettung BRD, CNSAS Südtirol, the Alpine Training Centre of the

Carabinieri, the Alpine Troops Command, fire brigades, Civil Protection, SABES and research partners. Participants rejected generic flight exercises in favour of complete missions grounded in their own procedures. They requested familiar equipment and interfaces, together with controllable wind gusts, darkness, cold, degraded communication, system failures and the approach of a helicopter. These inputs defined three design priorities: functional fidelity at decision-critical points, repeatable variation of operational demands, and continuous observability of actions rather than a single game score.

Simulator design and avalanche mission

The simulator reproduces a DJI Matrice 4T and its remote-controller interface. A data-driven flight model was parameterized from controlled motion-capture recordings, while a high-fidelity digital reconstruction of Val Gardena provides recognizable terrain. The virtual environment can vary visibility, snow, fog, wind and lighting and can inject communication or aircraft faults. A tracked, 3D-printed controller preserves the geometry and tactile layout of the operational device. The system is deployed in a terraXcube climatic chamber so that selected physical demands, particularly cold and wind, remain real while the mission stays virtual. This selective hybrid fidelity follows evidence that decision-relevant physical cues, rather than visual detail alone, can change behaviour in VR safety training (Monteiro et al., 2021).

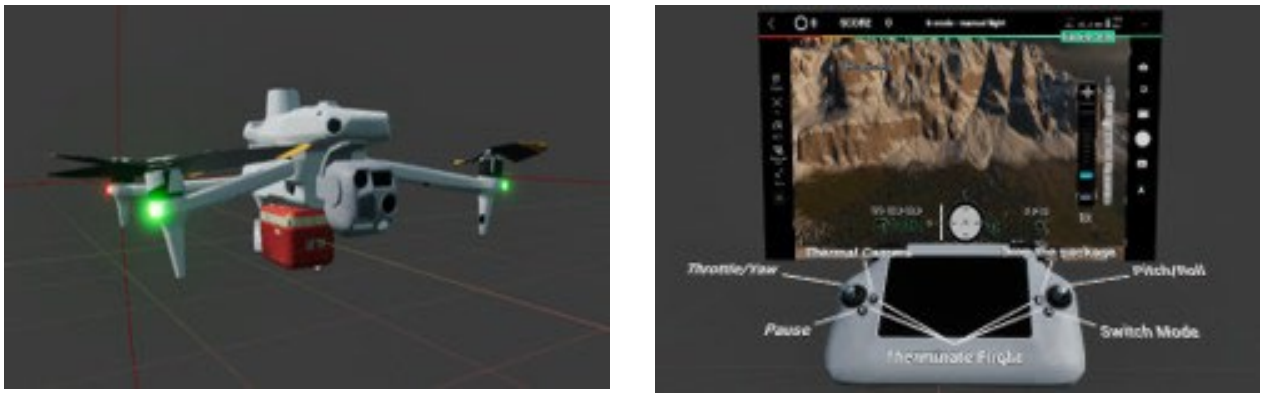


Figure 1. *Task-specific aircraft model and replicated controller interface used in the VR simulator.*



Figure 2*. Real and virtual Val Gardena terrain (top) and configurable clear, fog and snow conditions (bottom).*

Each session follows the same mission backbone. A neutral tutorial first isolates interface learning from later performance. The candidate receives an operational map briefing with the last known position of the victim, assigned sector, terrain, weather and other air assets, and chooses how to prepare the search. During flight, checkpoints, search coverage, sensor selections and detections are logged. A rescue helicopter subsequently enters the sector at a fixed scenario event. The simulator records whether and how the pilot prioritizes separation, communicates and terminates or redirects the UAV, as well as the latency and sequence of those actions. Environmental and technical variations can change between sessions; critical injections remain anchored to equivalent mission states. This combination responds to evidence that repeated VR stress scenarios must introduce unpredictable but context-specific demands if their effect is to persist beyond novelty (Binsch et al., 2021).

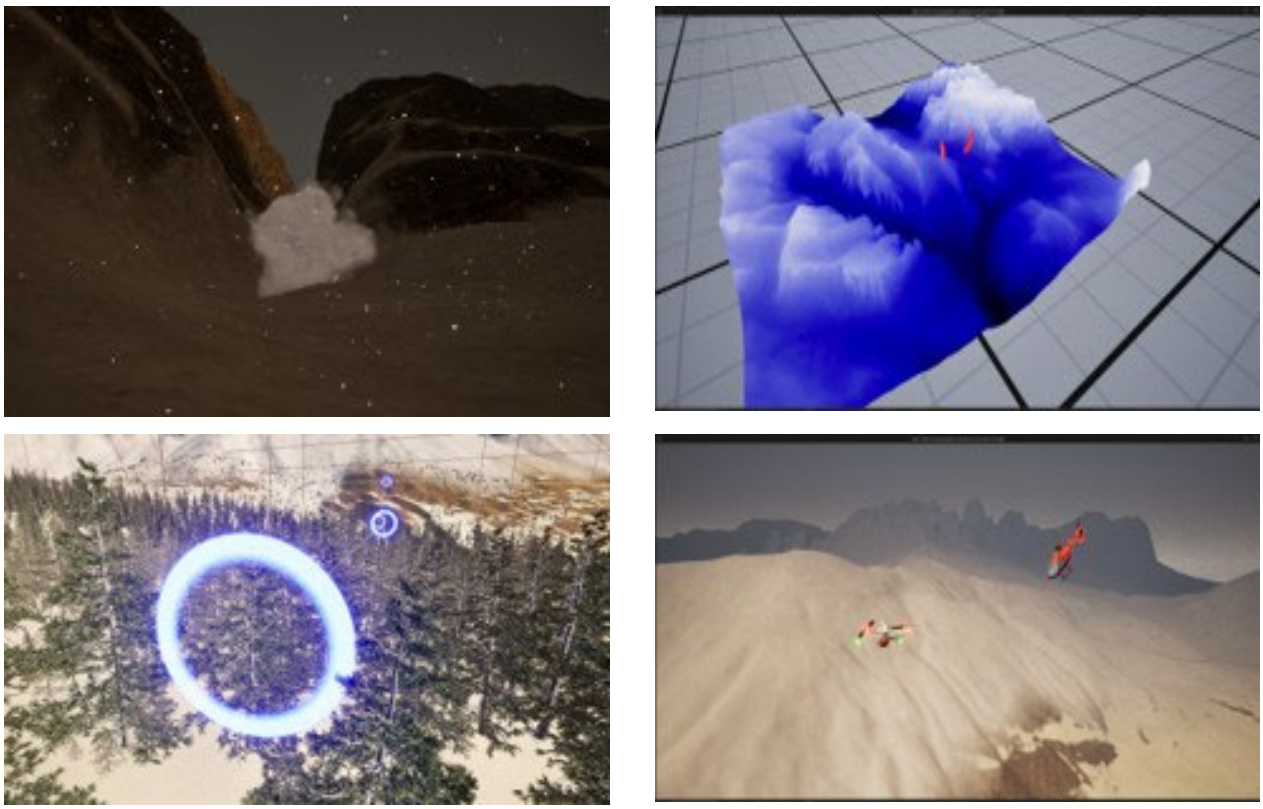


***Figure 3**. Standardized mission sequence: avalanche trigger, operational briefing, checkpoint search and helicopter encounter.*

Pilot Day assessment protocol

The 25 September 2026 Pilot Day is designed as a feasibility and baseline assessment, not yet as a test of training effectiveness. Rescue practitioners will complete familiarization, receive the briefing, conduct the mission and join an immediate debrief. During the assessed run, the simulator explains controls and safety boundaries but does not prescribe a search strategy. Performance will be interpreted against practitioner-approved procedural criteria. Consistent with mission-based rescue research, the analysis separates situation assessment, plan formulation and execution instead of collapsing competence into aircraft-control accuracy (Cohen-Hatton & Honey, 2015).

Table 1. *Assessment dimensions and synchronized evidence.*

Mission phase	Logged evidence	Interpretation
Preparation	Briefing queries, route and sector plan, pre-flight actions	Situation assessment and plan formulation
Search	Flight path, coverage, checkpoints, sensor use and detections	Aircraft handling and systematic search
Helicopter event	Response latency, UAV command, radio call and action sequence	Safety prioritization and airspace coordination
Attention and load	Head-controller pose, errors and NASA-TLX	Orientation between scene and screen; perceived workload

All simulator events and user actions are timestamped on a common clock. Before data collection, each scripted injection will be linked to an expected response, an observation window and one or more evidence sources, following event-based assessment design (Nguyen et al., 2016). Head pose and controller pose provide a coarse indicator of orientation between the external scene and the screen; they are not treated as eye tracking. Objective records will be triangulated with a five-point post-session questionnaire covering engagement and pacing, presence, agency, functional fidelity and utility, and stress and multisensory coherence, plus two open questions. Perceived workload will be captured with NASA-TLX (Hart & Staveland, 1988) and adverse symptoms with the Simulator Sickness Questionnaire (Kennedy et al., 1993). These experience measures will be reported alongside, not used as substitutes for, mission performance.

The initial analysis will remain descriptive and dimension-specific: event-aligned timelines, coverage maps, omissions, critical actions and response categories will be reviewed together with the practitioner debrief. We will not calculate a single compensatory score, because efficient flight should not mask a safety-critical failure. The study will establish whether the protocol can be completed consistently, whether the logs are sufficiently complete for reconstruction, and whether experts judge the derived indicators to be operationally meaningful.

From assessment to targeted training

The Pilot Day will establish whether the scenario is recognizable and usable for professionals and whether the logged indicators expose meaningful differences in how a real mission is planned and conducted. The design does not support a claim that one exposure changes behaviour: there is currently no pre-test, comparator or retention measure. Its immediate output is a baseline profile of strengths and recurrent procedural gaps. Those gaps can then be converted into focused modules, for example search planning, sensor discipline, communication under degraded links or rapid deconfliction with a helicopter, and practised across controlled variations. Validation can then progress from content validity and feasibility to known-groups comparisons, sensitivity to repeated practice, retention and transfer to field performance (Harris et al., 2020). In this way, the simulator becomes a structured path from assessment to individualized, evidence-based practice rather than a generic flight experience.

References

Binsch, O., Bottenheft, C., Landman, H.M., Roijendijk, L., & Vermetten, E. (2021). Testing the applicability of a VR simulation platform for stress training of first responders. *Military Psychology* 33, 182-196 https://doi.org/10.1080/08995605.2021.1897494

Cohen-Hatton, S.R., & Honey, R.C. (2015). Goal-oriented training affects decision-making processes in virtual and simulated fire and rescue environments. *Journal of Experimental Psychology*: Applied 21, 395-406https://doi.org/10.1037/xap0000061

Harris, D.J., Bird, J.M., Smart, P.A., Wilson, M.R., & Vine, S.J. (2020). A framework for the testing and validation of simulated environments in experimentation and training. Frontiers in Psychology 11, 605 https://doi.org/10.3389/fpsyg.2020.00

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX: Results of empirical and theoretical research. In: Human Mental Workload, pp. 139-183. North-Holland https://doi.org/10.1016/S0166-4115(08)62386-9

Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. International Journal of Aviation Psychology 3, 203-220 https://doi.org/10.1207/s15327108ijap0303_3

Knierim, P., Hetzenauer, A.-L., & Sitar, C. (2025). Transforming avalanche rescue training: Evaluating the effectiveness of virtual reality training in high-stakes environments. In*: IEEE ISMAR 2025*, pp. 1344-1353 https://doi.org/10.1109/ISMAR67309.2025.00139

Lyu, M., Zhao, Y., Huang, C., & Huang, H. (2023). Unmanned aerial vehicles for search and rescue: A survey. *Remote Sensing* 15, 3266. https://doi.org/10.3390/rs15133266

Makransky, G., & Petersen, G.B. (2021). The cognitive affective model of immersive learning (CAMIL). *Educational Psychology Review* 33, 937-958 https://doi.org/10.1007/s10648-020-09586-2

Monteiro, P., Melo, M., Valente, A., Vasconcelos-Raposo, J., & Bessa, M. (2021). Delivering critical stimuli for decision making in VR training. *IEEE Transactions on Human-Machine Systems* 51, 65-74 https://doi.org/10.1109/THMS.2020.3030746

Murray, J., Richardson, S., Joiner, K., & Wild, G. (2025). Identifying human factor causes of remotely piloted aircraft system safety occurrences in Australia. *Aerospace* 12, 206 https://doi.org/10.3390/aerospace12030206

Nguyen, N., Watson, W.D., & Dominguez, E. (2016). An event-based approach to design a teamwork training scenario and assessment tool in surgery. Journal of Surgical Education 73, 197-207 https://doi.org/10.1016/j.jsurg.2015.10.005

Somerville, A., Lynar, T., Joiner, K., & Wild, G. (2024). Use of simulation for pre-training of drone pilots. *Drones* 8, 640. https://doi.org/10.3390/drones8110640

GeoBoardXR: A WebXR-Based Immersive Terrain and Geospatial Data Viewer

Costas Boletsis, Maria Emine Nylund, and Ophelia Prillard

SINTEF Digital, Norway

Corresponding author: konstantinos.boletsis@sintef.no

Keywords: digital twin, GeoJSON, open source, virtual reality, WebXR

Introduction

Climate change is reshaping ecosystems, economies, and coastal communities worldwide. Nature-based solutions (NbS) have emerged as important interventions for strengthening social-ecological resilience and supporting climate adaptation (Turner et al., 2022). Effectively communicating such interventions by translating complex climate data and geospatial models into comprehensible representations is important for stakeholder engagement and planning. Visualization can support this process by communicating spatial information, potential impacts, and alternative scenarios (Wissen et al., 2008; Bohman et al., 2015).

Three-dimensional (3D) terrain visualization can complement traditional 2D cartography by providing spatially explicit representations of landscape characteristics and relationships (Lovett et al., 2015). Extending these capabilities, fully immersive digital twins further enable interactive exploration of complex 3D geospatial datasets within spatially organized virtual environments (Zhao et al., 2019). The development and application of such systems can further benefit from open-source approaches, which promote accessibility, transparency, reproducibility, and adaptability across different contexts (Schlögl et al., 2026).

In this context, GeoBoardXR[1] is introduced as an interactive, tabletop-scale virtual reality (VR) terrain viewer built on a WebXR-compatible framework. Being a foundational component of the EU RESIST project's Ocean Model platform[2], GeoBoardXR was developed as a fully immersive tool for communicating climate adaptation scenarios and NbS in the coastal and marine landscapes of Norway's Vesterålen region. More broadly, GeoBoardXR enables researchers, practitioners, and stakeholders to develop and explore high-resolution digital terrain models and layered geospatial datasets within an immersive 3D environment. By combining open-source development practices with accessible WebXR deployment, GeoBoardXR aims to support evidence-based decision-making through immersive, spatially grounded visualization.

[1] Code repository: https://github.com/SINTEF/GeoBoardXR, video: https://xrlab.no/geoboardxr/video/
[2] RESIST Ocean Model: https://github.com/SINTEF/oceanwebxrmodel, https://doi.org/10.5281/zenodo.22041952

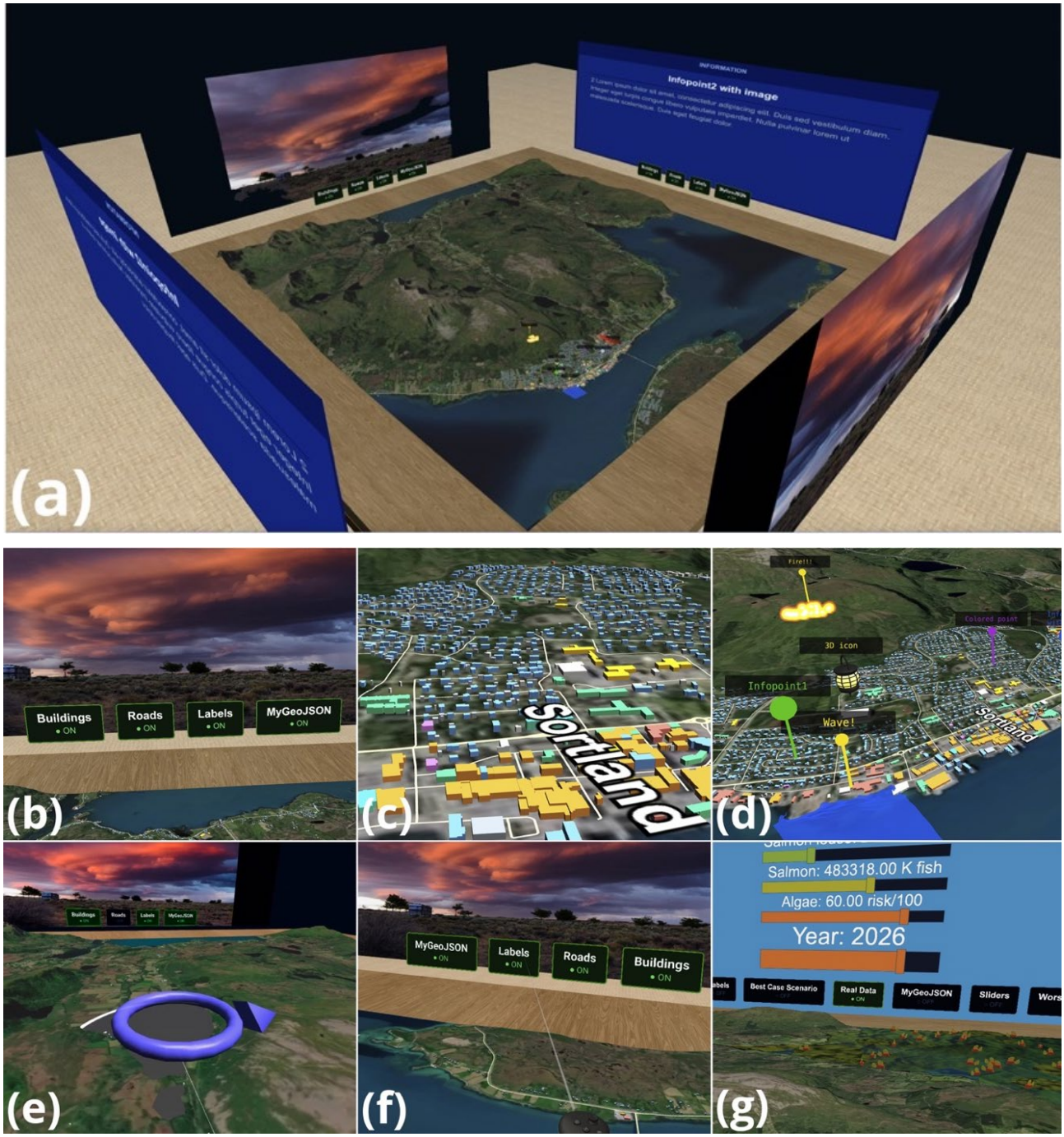


Figure 39. *(a) Overview of the GeoBoardXR virtual scene. (b) Toggle buttons managing geospatial content. (c) Overview of the OpenStreetMap Overpass API content with 3D buildings, roads and place labels. (d) Visualized user-created GeoJSON dataset. (e) GeoBoardXR's teleportation locomotion technique. (f) GeoBoardXR's raycasting-based "aim-and-shoot" input technique. (g) Visualization of a fuzzy cognitive map.*

Development Framework

GeoBoardXR is implemented as an open-source, browser-based XR application using TypeScript and Babylon.js, with Vite providing the development and build environment. Immersive functionality is implemented through the WebXR API and Babylon.js WebXR components, allowing the same application to be accessed directly through a URL without installation, while using the same codebase across desktop browsers and supported VR headsets. The application is served over HTTPS and requires no backend infrastructure beyond static file hosting. The current implementation has been tested using the Meta Quest 2 and Quest 3 headsets and Chromium-based web browsers.

Virtual Scene

GeoBoardXR organizes geographic content around a three-dimensional terrain surface. Terrain geometry is generated from Mapbox Terrain-RGB elevation tiles. The raw elevation grid is simplified using Martini, a real-time implementation of the Right-Triangulated Irregular Network (RTIN) method. The algorithm preserves geometric detail in areas where elevation varies while simplifying comparatively uniform regions. Mapbox satellite imagery is subsequently applied to the resulting terrain mesh as a surface texture.

From a design perspective, the terrain is positioned on a virtual table, which users can directly navigate by moving across its surface. Along the edges of the tabletop, interactive toggle buttons control the visibility of associated geospatial content, while “projection screens” around the tabletop display supplementary information (Fig. 1a). The buttons (Fig. 1b) and projection screens are instantiated on multiple sides of the table to remain accessible as users change position around the model.

The tabletop and projection screens interaction metaphor was inspired by the physical installation of the Gaia Projection Mapping System, which was created in a previous project for the Vesterålen region (Boletsis & Prillard, 2025). The tabletop provides a clearly defined spatial boundary for the virtual environment while limiting the geographic extent that must be rendered, thereby reducing computational demands. Moreover, users can move across the tabletop to inspect the terrain and the geospatial content at closer range, supporting spatial understanding of local features and their broader geographic context. The projection screens provide dedicated surfaces for presenting supplementary textual and visual information without occluding the terrain or its content.

Geospatial Content

Geospatial content is provided in two forms. Upon loading the terrain, API-sourced contextual data are obtained through the OpenStreetMap (OSM) Overpass API with a sequential mirror-fallback strategy. Separate loaders retrieve buildings, roads, and place labels (names). For buildings, their footprints are extruded from 2D outlines into 3D objects and they are color-coded according to their type (e.g., residential, commercial, office, and garage). Roads are represented using ribbon geometry, while place names are rendered using billboarded labels (Fig. 1c). Retrieved OSM responses are cached in the browser’s localStorage for seven days based on the geographic tile coordinates, minimizing redundant network requests. These three data layers can be activated or deactivated using the corresponding toggle buttons on the tabletop.

GeoBoardXR accepts user-defined geospatial data as GeoJSON FeatureCollections stored in the application’s public asset directory. Point, Polygon, and LineString geometries are parsed at load time and rendered as interactive 3D scene objects, with user-defined colors and labels (Fig. 1d). Points are positioned by geographic coordinates and displayed as pins or optional GLB 3D models. Each point can also function as an “infopoint”, enriched with text, images, and videos shown on the projection screens surrounding the tabletop. Polygon coordinates are transformed into scene coordinates, with color and opacity defined by GeoJSON properties, and can also serve as spatial boundaries for procedural fire and wave animations developed to communicate hazards such as forest fires and sea-level rise. LineStrings follow the supplied geographic coordinate sequence, while width and height properties control whether they appear as flat ribbons or vertical walls. All visual and informational attributes are defined through GeoJSON Feature properties, and each FeatureCollection can be activated or deactivated using a tabletop toggle button.

Together, these geospatial layers can support applied planning scenarios. For example, a local planner can combine terrain, buildings, roads, and a user-provided flood or sea-level-rise layer to identify potentially exposed infrastructure and inspect affected locations across spatial scales. Associated infopoints can provide contextual information on potential adaptation measures, supporting discussion of alternative interventions within the shared spatial representation.

User Interaction

In GeoBoardXR, users navigate the terrain by combining on-table teleportation locomotion (Fig. 1e) with real-walking within their available physical space. Teleportation locations are selected using the controller's joystick. Using a single VR controller, users interact with scene elements through a raycasting-based "aim-and-shoot" technique (Fig. 1f). This technique is used for triggering actions, e.g., to activate toggle buttons, activate infopoints by "shooting" at the pin heads or the 3D models attached to the point data, and play or pause videos. WebXR input profiles support consistent interaction across different VR controller configurations.

Conclusion: Limitations & Future Work

The current implementation has certain limitations related to its architectural design. WebXR performance depends on browser and headset support and may decrease with higher terrain resolutions or when large numbers of complex geospatial objects are rendered simultaneously. The accuracy and completeness of contextual layers also depend on external API sources and user-provided datasets, thereby retaining the uncertainties inherent in those original data sources. In addition, the current interaction design primarily targets controller-based VR and has not yet been systematically evaluated for accessibility or for users requiring alternative interaction modalities.

GeoBoardXR is under continuous development, with an emphasis on extending its geospatial capabilities for different use cases. Future work will focus on applying the tool across diverse scenarios to evaluate its effects on spatial understanding, communication, and decision-support activities, while identifying gaps and opportunities for improvement, particularly regarding usability and user experience, API integrations, and user-provided datasets. Another direction is the integration of fuzzy cognitive maps (FCMs), with nodes and weighted edges visualized as an interactive 3D graph layer on the terrain. Initial work on this visualization has been completed in a new fork (Fig. 1g). Planned developments also include GeoJSON-driven animations; dedicated layers for bathymetry, ocean currents, and water-column data; AI-assisted data curation; and improved AR calibration for table-scale placement.

Given its open-source nature and ongoing development, researchers and practitioners within and beyond the XR community are encouraged to fork GeoBoardXR, apply it to their own use cases, and provide feedback on potential features and software issues.

Acknowledgements

The work on GeoBoardXR presented herein was funded by the EU Horizon RESIST project (Grant Agreement No. 101093968). GeoBoardXR represents an early version of the RESIST Ocean Model platform, developed under WP3 and LSDT4 (Vesterålen). The Ocean Model was developed by: Maria Emine Nylund (SINTEF), Ophelia Prillard (SINTEF), and Costas Boletsis (SINTEF) with the valuable contributions of Jan Håvard Skjetne (SINTEF), Jiaxin Li (SINTEF), and the Vesterålen regional partners: Lofotr Næringsdrift AS, Andfjord Salmon AS, and Vesterålen Regionråd. Since June 2026, the continued development of GeoBoardXR and the Ocean Model platform has been supported by SINTEF funding.

References

Bohman, A., Neset, T.-S., Opach, T. & Rød, J.K. (2015). Decision support for adaptive action – assessing the potential of geographic visualization. *Journal of Environmental Planning and Management* 58(12), 2193–2211. https://doi.org/10.1080/09640568.2014.973937

Boletsis, C. & Prillard, O. (2025). The Gaia System: Revolutionizing Museum Storytelling with Projection Mapping. *Virtual Worlds* 4(4), 49. https://doi.org/10.3390/virtualworlds4040049

Lovett, A., Appleton, K., Warren-Kretzschmar, B. & von Haaren, C. (2015). Using 3D visualization methods in landscape planning: An evaluation of options and practical issues. *Landscape and Urban Planning* 142, 85–94. https://doi.org/10.1016/j.landurbplan.2015.02.021

Schlögl, M., Waltersdorfer, L., Regner, P., Siposova, A. & Brenning, A. (2026). Overcoming barriers to reproducibility in geoscientific data analysis: Challenges and practical implementation strategies. *Environmental Modelling & Software* 200, 106962. https://doi.org/10.1016/j.envsoft.2026.106962

Turner, B., Devisscher, T., Chabaneix, N., Woroniecki, S., Messier, C. & Seddon, N. (2022). The role of nature-based solutions in supporting social-ecological resilience for climate change adaptation. *Annual Review of Environment and Resources* 47, 123–148. https://doi.org/10.1146/annurev-environ-012220-010017

Wissen, U., Schroth, O., Lange, E. & Schmid, W.A. (2008). Approaches to integrating indicators into 3D landscape visualisations and their benefits for participative planning situations. *Journal of Environmental Management* 89(3), 184–196. https://doi.org/10.1016/j.jenvman.2007.01.062

Zhao, J., Wallgrün, J.O., LaFemina, P.C., Normandeau, J. & Klippel, A. (2019). Harnessing the power of immersive virtual reality – visualization and analysis of 3D earth science data sets. *Geo-Spatial Information Science* 22(4), 237–250. https://doi.org/10.1080/10095020.2019.1621544

SPACE+: Prototyping a Spatial AI Ecosystem for Participatory Spatial Management in XR

Hermenegildo Solheiro[1] and Keisuke Toyoda[1]

[1]*Institute of Industrial Science, The University of Tokyo, Tokyo, Japan*

Corresponding author: hermenegildosolheiro@g.ecc.u-tokyo.ac.jp

Keywords: Spatial AI, XR, Digital Twins, Localization, Semantic Integration, Participatory Spatial Management, Smart City

Abstract

SPACE+ is a Spatial AI framework for participatory spatial management in XR. Its prototype, built on consumer devices and interchangeable AI components, lets users georeference their indoor sessions through SLAM maps that inherit fixes from nearby VPS locations. Users then interact with their environment assisted by a spatially aware AI voice agent that answers their questions and reasons with them about how to improve their space. Their insights and suggestions are shared as georeferenced XR content with other users and managers through a governed data-exchange framework. We describe the system and its preliminary deployment on the University of Tokyo's Komaba campus.

The Prototype

Localization. Using commodity devices — phones and tablets — users point at prerecorded localization spots to join a common coordinate system. This ensures that spatial data and semantics remain spatially consistent across sessions, allowing integrated use and analysis. Users can also save new localization spots, expanding the localization network for further sessions that, adding geometry and semantic labels, progressively build a spatially consistent semantic-rich digital twin. Under the hood, stored indoor SLAM maps (e.g. ARKit ARWorldMap) pre-georeferenced against outdoor commercial VPS fixes (e.g. Niantic Lightship) allow a device that localizes indoors to inherit georeference. Keeping the session area limited (~50 m radius) and localizing against the nearest localization spot for each session keeps XR content consistently positioned at session level (Kim et al. 2022), while expanding the relocalization chain to connect multiple VPS fixes allows correcting multi-session drift and keeping consistency at ecosystem level.

Semantic orchestration. Users hold their devices as cameras and interact with the environment through spatially anchored semantic labels visible on screen; tapping an element opens a spatially aware conversational interface with a voice agent that informs them about the space and its elements and reasons with the user about spatial problems, reaching solutions and insights that can be shared with other users, as well as with managers and planners at ecosystem level.

Under the hood, scanned 3D geometry is segmented in real time by multiple models (OneFormer, SAM, and YOLO; Solheiro et al. 2025; Kent et al. 2025) and enriched by vision–language models (Claude, Gemini, ChatGPT), location-based live web content, and users through voice interaction. Task-specific knowledge graphs and serializers expose the relevant elements to an LLM, which

reasons conversationally over the scene. Knowledge graphs and serializers can be re-defined, so new tasks can be configured without re-engineering.

Governed sharing. Users choose from the outset whether and how to share their data, and different user profiles experience the environment differently according to their roles and permissions — guests see usage instructions and attach opinions, while staff see task schedules and add suggestions, both contributing to internal management decisions. Underneath, governance attributes are embedded in the generated data from the start. Every runtime view — a map, a 3D reconstruction, an information card — is instantiated only from elements the requesting user, device, or application is authorized to see. The same attributes protect contributions, personalize consumption, and keep every observation — including user reports and suggestions — attributable to its author, making it possible to reward participants or give them feedback on their data and its usage (Li et al. 2023), further stimulating participation.

Prototype Deployment

The prototype is implemented as an iOS client for in-situ capture and interaction, a desktop client for ecosystem-level views, and a backend that persists all spatial knowledge. It was exercised by a small group of users on consumer devices, across multiple sessions in connected indoor and outdoor environments of the Komaba campus. Three scenarios tested its capabilities.

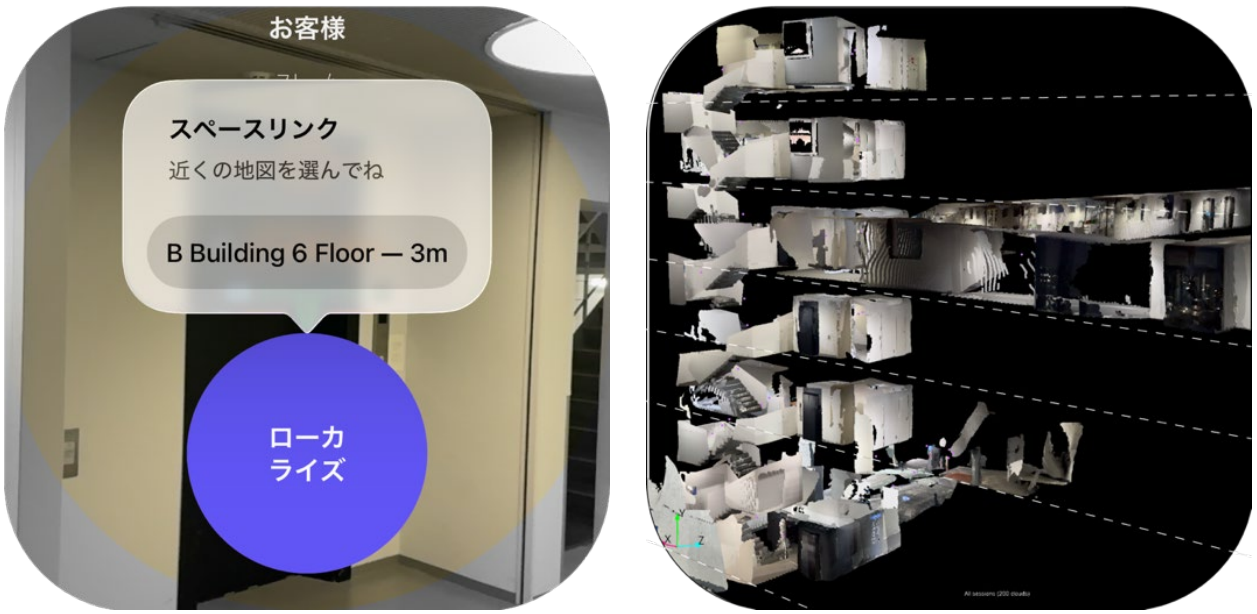


Figure 40. *Georeference inheritance and ecosystem-level integration.* ***Left****: indoor localization spots extend georeferencing across floors and subsequent sessions.* ***Right****: the desktop client aggregates persistent georeferenced geometry and semantics.*

Georeference inheritance across sessions and devices

A first *staff* user, using a smartphone, acquires an outdoor VPS fix, walks indoors, and saves and shares the device-local map next to the elevator access area on the building's entrance floor, then takes the stairway and saves a fresh localization spot next to the elevators on each floor. From then on, a second *guest* user on a different device (a tablet) conveniently relocalizes new sessions at their floor's access area and saves their own localization spots — in their lab and in a favorite common area — extending the network for easy relocalization and allowing further indoor sessions to inherit georeferencing together with annotated spatial anchors and labels. The app recommends the nearest spot for localization and suggests saving a new one when none exists within a 50 m radius, to minimize in-session drift. In the deployment tests, object positions remained consistent across chains up to three hops deep over four sessions and two devices, with drift between under 5

cm and 0.4 m across sessions spanning 10–60 m — sufficient for shared content placement and in-situ information delivery at room-to-building scale.

Geometry-aware AI reasoning: disaster prevention as use case

Next, inside the Lab, the *guest* user runs the iOS client configured for a disaster-prevention task, generating an annotated scene in which each detected object carries a semantic label and earthquake/fire hazard assessments. The user reasons together with the voice agent: tapping an object opens its risk card, and the agent conversationally proposes concrete mitigations — e.g. relocating a flammable box, identified as an earthquake fall hazard, to a lower shelf (Figure 2). The insights were persisted and shared with other *guest* and *staff* users. The guest user shared session contents selectively: risk assessment data was shared at ecosystem level, contributing to building-level risk assessment, while geometry and image data were kept private. In a human evaluation of the generated annotations, semantic labels reached ~80–85% accuracy and earthquake and fire risk reports 94.3% and 91.4%, respectively, supporting coherent conversational reasoning and actionable proposals (a systematic evaluation of risk-assessment accuracy and reliability is beyond the scope of this paper).

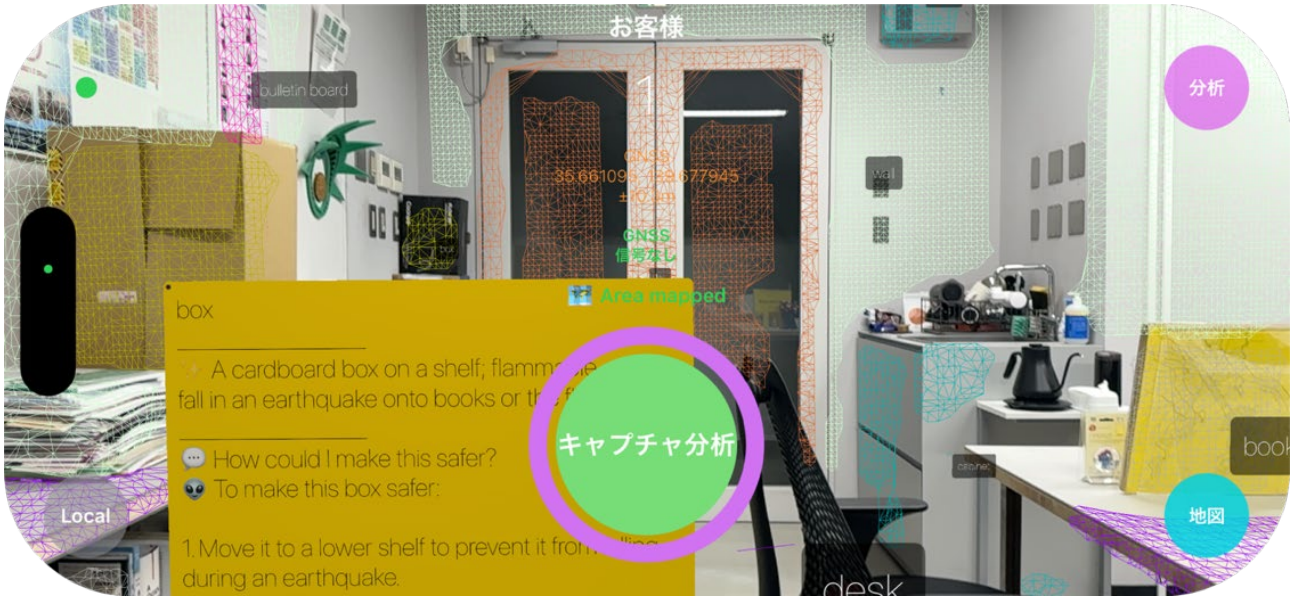


Figure 41. *Geometry-aware disaster-risk reasoning in the SPACE+ UI. A mesh-segmented laboratory is annotated with per-object semantics; selecting an object opens its risk card, while the voice agent proposes a spatial mitigation based on the scene context. Data is selectively shared: semantic data contributes to building-level risk assessment, while geometry and image data are gated for privacy.*

One substrate, many governed views

Finally, both guest and staff users walked to a common area, relocalized and inspected XR labels that showed different information according to user role. Figure 3 shows the same sitting area viewed under two roles: a guest is shown usage guidance for the table (light use only, hours of use), while a staff member is shown maintenance instructions (cleaning schedule, stability check, evacuation clearance) for the same object. The same attributes used to define roles and permissions and select the delivered semantics answer to core federated data-space building blocks (Gaia-X 2024; IDSA 2022) — identities, roles, persistent IDs, object-bound permissions, policy-checked retrieval. This shows that, rather than being an obstacle, governance attributes in spatial data and XR support functionality, while structurally enabling feedback and incentive attribution, to stimulate participation.



Figure 42. *Role-scoped semantics on a shared scene-graph node: the same table resolves to usage guidance for a guest (left) and maintenance and evacuation-clearance instructions for staff (right). The attributes that gate aggregation, define roles, and select the delivered semantics correspond to core federated data-space building blocks.*

Conclusion and Outlook

The prototype shows that a progressive, persistent, user-contributed Spatial AI ecosystem for XR can be assembled from consumer devices and interchangeable AI components: indoor georeferenced XR without exhaustive VPS coverage, configurable geometry-aware AI reasoning, and a shared 3D geometric and semantic substrate that many participants can safely enrich and differently experience. Going forward, additional localization methods, e.g. Open AR Cloud's OSCP (Open AR Cloud 2026), and devices (Android smartphones, smart glasses, etc.) can support vendor independence, while usage accounting can be activated to reward participants for the spatial knowledge they generate, helping drive the development of a fully participatory Spatial AI ecosystem. Next, the deployment envisions expansion toward campus scale. We believe these mechanisms can extend to the surrounding city areas and, ultimately, to Tokyo — supporting citizen participation in architectural and urban interventions over a shared, governed spatial AI.

References

Gaia-X. (2024). Gaia-X Architecture Document, Release 24.04. *Gaia-X European Association for Data and Cloud. docs.gaia-x.eu*

IDSA. (2022). IDS Reference Architecture Model (IDS-RAM), Version 4. *International Data Spaces Association*. docs.internationaldataspaces.org

Kent, L., Solheiro, H. & Toyoda, K. (2025). Multiple multi-modal AI for semantic annotations of 3D spatial data. In: *Proc. GRAPP/VISIGRAPP 2025*, 308–316. https://www.scitepress.org/Papers/2025/132353/132353.pdf

Kim, J., Song, M., Lee, Y., Jung, M. & Kim, P. (2022). A benchmark comparison of four off-the-shelf proprietary visual–inertial odometry systems. *Sensors* 22(24), 9873. https://doi.org/10.3390/s22249873

Li, H., Vincent, N., Chancellor, S. & Hecht, B. (2023). The dimensions of data labor. In: *Proc. ACM FAccT 2023*, 1151–1161. https://dl.acm.org/doi/abs/10.1145/3593013.3594070

Open AR Cloud. (2026). Open Spatial Computing Platform (OSCP). *openarcloud.org/oscp*

Solheiro, H., Kent, L. & Toyoda, K. (2025). 3-2-3 Multi-AI Segmentation Framework: LoD-based, incremental segmentation of 3D scan data using any 2D AI. In: *Virtual Reality and Mixed Reality (EuroXR 2024)*, LNCS 15445, 120–133.

ITI MIND: Translating Five Years of Mixed Reality Research for Mental Health and Wellbeing into a Platform for Real-World Care Settings

Patricia Pons, José Luis Soler-Dominguez, and Samuel Navas-Medrano

HCI research group, Instituto Tecnológico de Informática (ITI), Valencia, Spain

Corresponding author: ppons@iti.es

Keywords: Mixed Reality, Mental Health, Wellbeing, Gamification, User-Centred Design.

Introduction

Extended Reality (XR) has created new opportunities for technology-mediated mental health and wellbeing interventions (Omisore et al., 2024; Bell et al., 2024). However, many XR solutions remain designed around a single therapeutic strategy, target population, or application scenario, which creates challenges when moving from research prototypes to practice (Torous et al., 2025). Mental health professionals routinely work with heterogeneous users, diagnoses, cognitive abilities, therapeutic objectives, and intervention strategies. This mismatch is a central challenge when designing XR technologies for mental health: rather than requiring professionals to adopt independent technological solutions for different needs, XR platforms should provide sufficient flexibility to accommodate different therapeutic practices and individual users.

Furthermore, the field has historically been dominated by Virtual Reality (VR), particularly for exposure-based interventions (Bell et al., 2024; Pedram & Piatkowski, 2025), leaving Mixed Reality (MR) comparatively underexplored (Omisore et al., 2024; Navas-Medrano et al., 2024). MR offers a distinct design space rather than simply a less immersive form of VR: digital content can be spatially integrated into the user's real environment while preserving awareness of the physical setting and visual contact with professionals. These properties create opportunities for co-present guidance, in-situ practice, and may improve accessibility for users who find fully immersive VR difficult to tolerate (Pons et al., 2022; Navas-Medrano et al., 2024).

Over the last five years, we have explored these challenges together with mental health professionals and users. This research trajectory evolved from isolated explorations of how MR could support emotional regulation, self-compassion, and physiological awareness, into a comprehensive MR platform. Rather than resulting in independent prototypes, these successive research and design iterations gradually converged into a pioneering MR platform featuring a persistent self-care environment, physiological biofeedback, and selectable modules tailored for mental health, rehabilitation, and social care. This paper presents this research-to-application trajectory, synthesizes the main design and evaluation insights, and describes the current transition towards longitudinal validation in real-world settings.

From Co-Design to an Integrated Mixed Reality Platform

We adopted a human-centred and participatory co-design approach, involving care professionals from different backgrounds (psychology, occupational therapy, physiotherapy, and social workers) to identify processes that are applicable across different mental health conditions and user profiles. Initial designs focused on physiological manifestations of emotions, mindfulness-based regulation, and thought externalization (Soler-Dominguez, 2024). Physiological biofeedback also emerged as a relevant design opportunity, as it provides users with an additional resource for reflection and self-awareness. To sustain motivation and engagement over repeated therapeutic sessions, gamified elements and a coherent narrative were also incorporated. Professionals repeatedly emphasized the need for flexibility across users and care practices, so rather than defining a fixed therapeutic sequence, the system was conceived as a modular environment in which professionals could combine activities according to individual needs and therapeutic objectives.

Iterative user studies with both care professionals and service users helped identify usability barriers, interaction breakdowns, and opportunities for extending the therapeutic experience. These evaluations directly informed subsequent design decisions (Soler-Dominguez, 2024). It also allowed to redefine compound interactions, such as drawing in the air with your hand for externalizing an intrusive though, that relied too much on augmented panels. Instead, based on feedback, this interaction was redesigned based only on hand tracking and gesture recognition, obtaining better usability results in successive evaluations (García-Ballesteros, 2024). In addition, modifications of existing activities and the need for new ones emerged naturally during conversations facilitated by the evaluations and direct testing of professionals and users. Consequently, in the last couple of years, the research expanded beyond affective regulation to include occupational and social care, physical rehabilitation, and cognitive stimulation, culminating in a unified MR platform, ITI MIND.

ITI MIND: A Modular Mixed Reality Platform for Care and Wellbeing

ITI MIND is an MR platform featuring interactive and gamified therapeutic and wellbeing activities, designed to support user motivation while allowing professionals to adapt the available content to different objectives and profiles. The experience is organised around a self-care garden that metaphorically represents the user's wellbeing, provides access to the different activities, and incorporates visual progression mechanisms intended to support motivation over repeated use. A second cross-cutting component is the emotional avatar, a digital character that externalises physiological signals such as breathing and heart-related measures captured through wearables. The avatar does not classify emotions; it makes otherwise internal processes visible in a non-evaluative form intended to support awareness and self-compassion. The platform covers three application areas: 1) Affective wellbeing and mental health**:** Modules targeting physiological self-awareness, mindfulness, intrusive-thought externalization and emotion identification; 2) Occupational and social care**:** Digital animal interactions that support fine-motor skills and facilitate the learning and internalization of care and self-care routines; 3) Physical and cognitive rehabilitation**:** Movement-based active gaming with adjustable difficulty levels, alongside exercises targeting attention, memory and impulse control. Additionally, a companion app allows professionals to monitor user interaction live, generate biometric reports from sessions, and select activity combinations based on intervention goals.

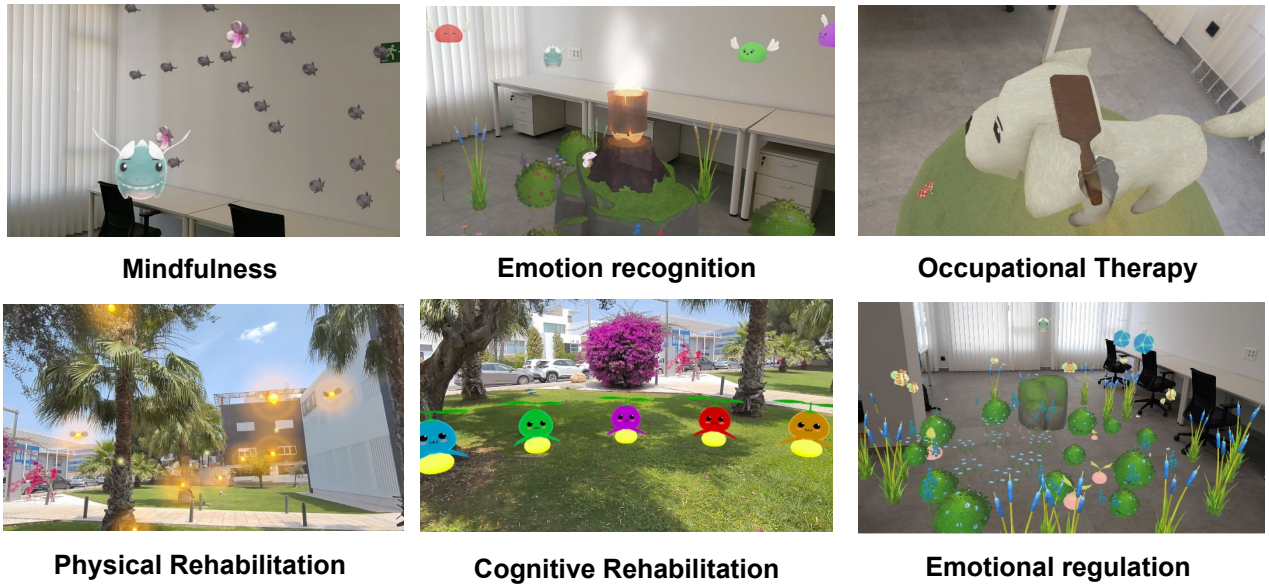


Figure 43. *Several activities available at ITI MIND.*

Validation evidence and transferable design lessons

Across published studies, evaluations have involved 30 therapists and 32 users aged 20-67, spanning heterogeneous clinical and cognitive profiles. Mixed methods yielded encouraging feasibility data: almost all professionals stated that they would use the system with patients and would recommend it to a colleague; 86% users reported feeling calmer after the session (SAM measures also showed significant increases in valence and dominance); >95% considered that the experience could help them feel better when distressed and wanted to use it again (Soler-Domínguez, 2024; García-Ballesteros, 2024). Additional formal validations (10 professionals, 7 users) and user feedback gathered during demos in national and international events (pending publication) also corroborate this tendency. Overall, the system has generally been perceived as motivating, accessible, and suitable for further use.

Several findings have remained consistent across evaluations. First, MR interaction presents relatively low learning barriers, even for participants without previous XR experience. Users report that gesture interaction feels natural, no need for prior training, and easy to understand metaphors and visual representations. Second, both users and professionals showed high acceptance levels and willingness to reuse the technology. Professionals considered interactions adequate and useful across varied mental health situations and for users with diverse cognitive and physical abilities. Finally, users and professionals expressed a need for greater personalisation and interaction possibilities. The cumulative design and evaluation process has led to four practical principles that have guided the consolidation of the platform:

Keep interactions simple, but meaningful. Care applications may involve users with heterogeneous cognitive and physical abilities, making unnecessary interaction complexity particularly problematic. At the same time, interactions that are excessively passive may become repetitive during continued use. Gamified feedback and narrative context can therefore increase engagement without requiring mechanically complex interactions.

Balance autonomy with guidance. MR naturally supports exploratory interaction, but many care activities rely on meaningful sequences or specific objectives. Guidance from therapists remains essential, while spatial cues and contextual feedback can help preserve this balance.

Design for the physical environment. Unlike fully virtual environments, MR experiences depend directly on the physical characteristics of the deployment space. Field-of-view limitations, furniture, occlusion, movement requirements, and interaction distance need to be considered when deciding where and how digital content is placed.

Design for heterogeneity and personalisation. Differences in cognitive abilities, therapeutic goals, professional practices, physical capabilities, and user preferences make rigid one-size-fits-all experiences difficult to transfer into care settings. Modularity, adjustable difficulty, selectable activities, and customizable content therefore constitute core requirements for adoption rather than optional enhancements.

From Short-Term Validation to Real-World Adoption

The accumulated evidence currently supports the usability, acceptability, positive immediate user experience, and perceived application potential. However, this evidence remains predominantly based on relatively short individual sessions, typically around 15–20 minutes. Such evaluations cannot determine whether the initial positive experience persists after repeated exposure, whether novelty effects disappear, or whether the platform produces meaningful changes in wellbeing over longer periods. To address this limitation, our platform is currently moving into this next stage through a longitudinal validation in real-world settings within an Urban Sandbox. This phase shifts the focus from controlled testing to sustained routine clinical integration, investigating whether the platform remains viable, engaging, and impactful when embedded in everyday care practices.

The evolution of ITI MIND illustrates how a sustained human-centric and participatory research trajectory can progressively move from exploratory MR interactions towards an integrated application platform. Our platform can thus be understood as the result of cumulative research-to-practice translation: successive HCI findings have informed interaction redesign, system architecture, modularity, and new application domains, ultimately converging into a platform intended to support mental health, wellbeing, rehabilitation, and social care.

Acknowledgments

We are grateful to the mental health professionals and patients who participated in the codesign and evaluation of the system for their invaluable feedback. This work was funded by the Valencian Institute for the Enterprise Competitiveness-Generalitat Valenciana (IMAMCA/2026/11) and IVACE/FEDER (IMDEEA/2022/22, IMDEEA/2023/40, IMDEEA/2022/22).

References

Bell, I. H., Pot-Kolder, R., Rizzo, A., Rus-Calafell, M., Cardi, V., Cella, M., ... & Valmaggia, L. (2024). Advances in the use of virtual reality to treat mental health conditions. *Nature Reviews Psychology*, *3*(8), 552-567.

García-Ballesteros, M., Navas-Medrano, S., Soler-Dominguez, J.L. & Pons, P. (2024). The Garden Within: Using Mixed Reality to Foster Wellbeing Through Emotional Regulation. *ISMAR Adjunct 2024*. https://doi.org/10.1109/ISMAR-Adjunct64951.2024.00071

Navas-Medrano, S., Soler-Dominguez, J.L. & Pons, P. (2024). Mixed Reality for a collective and adaptive mental health metaverse. *Frontiers in Psychiatry* 14:1272783. https://doi.org/10.3389/fpsyt.2023.1272783

Omisore, O.M., Odenigbo, I., Orji, J., Hernandez Beltran, A.I., Meier, S., Baghaei, N. & Orji, R. (2024). Extended Reality for Mental Health Evaluation: Scoping Review. *JMIR Serious Games* 12:e38413 https://doi.org/10.2196/38413

Pedram, S. & Piatkowski, T. (2025). Exploring the potential of virtual reality (VR) in mental healthcare: a systematic literature review. *Virtual Reality* 29:135

Pons, P., Navas-Medrano, S. & Soler-Dominguez, J.L. (2022). Extended reality for mental health: Current trends and future challenges. *Frontiers in Computer Science* 4:1034307. https://doi.org/10.3389/fcomp.2022.1034307

Soler-Dominguez, J.L., Navas-Medrano, S. & Pons, P. (2024). ARCADIA: A gamified mixed reality system for emotional regulation and self-compassion. CHI 2024. https://doi.org/10.1145/3613904.3642123

Torous, J., Linardon, J., Goldberg, S. B., Sun, S., Bell, I., Nicholas, J., ... & Firth, J. (2025). The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry, 24(2), 156-174.

Evaluating a Co-Designed Immersive Virtual Reality Prototype for Social Anxiety Support in Young People

Ivan Phelan, MSc; Alicia Carrion-Plaza, PhD

Impact VR Lab, Sheffield Creative Industries Institute, College of Social Sciences and Arts, Sheffield Hallam University, Sheffield S1 1WB, UK

Corresponding author: i.phelan@shu.ac.uk

Keywords: Immersive Virtual Reality; Social Anxiety; Young People; Virtual Exposure; Breathing Techniques

Background

Social Anxiety Social Anxiety Disorder (SAD) is characterised by a persistent fear of social situations and commonly develops during adolescence (Salari et al., 2024; Holmberg et al., 2020). A recent global systematic review and meta-analysis estimated that SAD affects approximately 8.3% of adolescents, although prevalence estimates vary substantially according to population, setting and assessment criteria (Salari et al., 2024). Cognitive Behavioural Therapy (CBT) remains the recommended psychological treatment for children and young people, with graduated exposure to feared or avoided social situations as a core component (Tan et al., 2025).

Immersive Virtual Reality (IVR) can provide controlled and repeatable social situations in which users experience simulated interactions and practise coping strategies (Wong et al., 2023). Recent systematic reviews and meta-analyses, conducted predominantly with adults, indicate that Virtual Reality Exposure Therapy (VRET) can reduce social-anxiety symptoms and may produce outcomes comparable to other psychological interventions (Shahid et al., 2024; Beidel et al., 2021; Boyle et al., 2024).

Evidence involving children and adolescents is smaller but growing (Azimisefat et al., 2026; Ramsey et al., 2023; Geraets et al., 2019), with studies suggesting that IVR interventions can be feasible, acceptable and capable of eliciting clinically relevant anxiety. Further research is needed to identify which virtual social situations are most relevant to young people, how they respond to different scenarios and coping activities, and how these systems should be designed for future therapist-supported use.

This paper presents IVR-SAD, a clinician co-designed prototype combining graded virtual social situations with breathing activities. This laboratory playtest explored participants' responses, usability, acceptability, perceived realism and real-world relevance, alongside practitioner feedback to inform further prototype development.

IVR-SAD Prototype

IVR-SAD was developed as an immersive virtual environment incorporating graded social situations and coping activities. The prototype contained three scenarios intended to expose users to different potentially challenging social situations (Figure 1). A clinical psychologist informed the design, advising that young people with SAD may feel anxious when encountering peers or believing that peers are judging them. This input informed the design of the scenarios. Participants were not

recruited based on a clinical diagnosis, and their social-anxiety symptoms were not formally assessed.



Figure 1. *IVR-SAD prototype scenarios and coping activities. (A) dog barking and peer interaction; (B) tree breathing; (C) virtual human interaction and square breathing; (D) coffee-shop interaction.*

In the first scenario, participants encountered peers across the street while walking. A nearby dog barked unexpectedly, after which the peers laughed at the participant's startled response. The dog was therefore not intended as an isolated fear stimulus, but as part of a broader scenario involving peer observation and perceived ridicule. This was followed by a tree-breathing coping activities, in which the virtual environment was visually simplified to direct attention towards breathing. In the second scenario, a young avatar approached the participant and, at a later exposure level, unexpectedly shoved them. This scenario was informed by practitioner feedback that simply encountering an approaching peer, even from a distance, can provoke anxiety. The shove was added to increase the intensity of the exposure.

The third scenario simulated ordering a drink in a coffee shop. Participants interacted verbally with a virtual staff avatar using voice recognition. The staff avatar presented three interaction styles, friendly, neutral, and passive-aggressive, to vary the social tone of the interaction. Participants also encountered social elements, including another customer observing them and a person waiting behind them. No breathing technique was prompted, allowing participants to decide independently how to manage their response.

Method

A laboratory-based playtest was conducted with eight young people aged 13–16 years (M = 14.75). Seven participants were recruited through Sheffield Futures Charity and one through Sheffield Children's Hospital. Five participants had access to Sheffield Futures wellbeing café supporting group, two had a diagnosis of Autism Spectrum Disorder (ASD) and two of Attention Deficit Hyperactivity Disorder (ADHD). Ethical approval was granted by Sheffield Hallam University (Ethics Review ID: ER53542568). Written parental or guardian consent and participant assent or consent were obtained as appropriate. Participants received age-appropriate information about the potentially distressing scenarios, voluntary participation and their right to withdraw. Sessions were supervised by the Sheffield Futures Health and Wellbeing Coordinator.

Participants experienced the three IVR scenarios and associated coping activities. Semi-structured interviews were conducted face-to-face by two members of the research team, with background in psychology and user experience research. Interviews explored participants' experiences of the IVR system, including their responses to the virtual scenarios and breathing activities, perceived realism, usability, acceptability, and perceived relevance to real-world. Interviews were audio-recorded, transcribed and anonymised. An individual interview was conducted with a Sheffield Futures Health and Wellbeing Coordinator to capture a practitioner perspective on the prototype and its further development.

As part of the interviews, participants provided ad hoc self-reported ratings of anxiety associated with each virtual scenario and immediately before and after the tree and square breathing activities. Additional ratings explored perceived realism, perceived potential of the coping activities to prepare someone for difficult real-world situations, and ease of use of the IVR system. Quantitative ratings were summarised using descriptive statistics.

Qualitative interview data were analysed using directed content analysis. Initial areas of analysis reflected the aims of the playtest, particularly participants' perceptions of the prototype's usability, acceptability, virtual scenarios and coping activities.

Results

The virtual scenarios elicited varying levels of self-reported anxiety (Table 1). Mean ratings were 5.19/10 for dog barking/peer laughter, 5.86/10 for the virtual avatar shoving interaction, and 4.69/10 for the coffee-shop scenario. Although the coffee shop produced the lowest mean rating, interviews identified socially relevant sources of discomfort, including speaking to virtual staff, being observed, changing an order, and feeling pressure from another person waiting.

Self-reported anxiety ratings were lower following both breathing activities (Table 1), from 5.31 to 2.13 following tree breathing and from 5.25 to 1.88 following square breathing. Participants described the activities as helping them focus and regulate their breathing; five of eight preferred square breathing, particularly because it offered greater control and was easy to remember. During the subsequent unprompted coffee-shop scenario, two participants independently applied a previously practised breathing strategy, while others reported forgetting to use it.

Participants rated the virtual experiences highly for realism (M=8.44/10) and perceived potential of the coping activities to prepare someone for difficult real-world situations (M=7.63/10). Controls were generally manageable following familiarisation, although some participants experienced difficulties with orientation, doors/gates, and slow locomotion. Six of eight would use the system again, and all eight would recommend it to a friend with social anxiety.

Table 1. *Self-reported ratings during the IVR-SAD playtest.*

Domain	Measure	n	Mean rating (0-10)
Anxiety response	Dog barking/peers laughing	8	5.19
	Virtual human shoving	7	5.86
	Coffee-shop interaction	8	4.69
Coping activity	Tree breathing: before → after	8	5.31 → 2.13
	Square breathing: before → after	8	5.25 → 1.88
Perceived experience	Realism of the virtual experiences	8	8.44
	Potential of coping techniques to prepare for difficult real-world situations	8	7.63
Usability	Ease/difficulty of using the IVR system	8	4.94

Note. Anxiety ratings ranged from 0 (not at all anxious) to 10 (extremely anxious). Realism and perceived real-world relevance were also rated on 0–10 scales. Ease/difficulty was rated from 0 (too easy) to 10 (impossible). Pre/post breathing ratings are descriptive and should not be interpreted as evidence of clinical effectiveness.

Discussion and Future Development

The playtest provided initial insights into the usability, acceptability and perceived realism of IVR-SAD, while highlighting substantial individual differences in responses to social scenarios and coping activities. This supports developing more diverse and personalised scenarios, particularly peer-focused and conversational interactions, alongside multiple coping options. Lower post-breathing anxiety ratings and spontaneous within-VR use by two participants support retaining these activities, although these descriptive findings do not demonstrate clinical effectiveness or real-world transfer. Usability challenges, including orientation, navigation and locomotion, also highlight the importance of minimising system-related difficulties that may interfere with the intended social experience. Future development will therefore focus on progressive virtual avatar interactions, personalisation, improved onboarding and navigation, and additional sensory features. Further evaluation with the intended clinical population is required to assess suitability for future therapist-supported use.

References

Azimisefat, P., Ravanbod, S., de Jongh, A., et al. (2026). Virtual reality exposure-enhanced cognitive behavioral therapy versus CBT with imaginal exposure for adolescent patients with social anxiety disorder symptoms: A pilot randomized controlled trial. *Depression and Anxiety*, 9956436. https://doi.org/10.1155/da/9956436.

Beidel, D.C., Tuerk, P.W., Spitalnick, J., et al. (2021). Treating childhood social anxiety disorder with virtual environments and serious games: A randomized trial. *Behavior Therapy* 52(6), 1351–1363. https://doi.org/10.1016/j.beth.2021.03.003.

Boyle, D., Morrin, N., Scargill, J., et al. (2024). Co-designing a virtual reality exposure therapy with students experiencing school anxiety: A proof-of-concept study. *Irish Journal of Psychological Medicine* 41(4), 469–477. https://doi.org/10.1017/ipm.2024.22.

Geraets, C.N.W., Veling, W., Witlox, M., et al. (2019). Virtual reality-based cognitive behavioural therapy for patients with generalised social anxiety disorder: A pilot study. *Behavioural and Cognitive Psychotherapy* 47(6), 745–750. https://doi.org/10.1017/S1352465819000225.

Holmberg, T.T., Eriksen, T.L., Petersen, R., et al. (2020). Social anxiety can be triggered by 360-degree videos in virtual reality: A pilot study exploring fear of shopping. *Cyberpsychology, Behavior, and Social Networking* 23(7), 495–499. https://doi.org/10.1089/cyber.2019.0295.

Jefferies, P. & Ungar, M. (2020). Social anxiety in young people: A prevalence study in seven countries. *PLOS ONE* 15(9), e0239133. https://doi.org/10.1371/journal.pone.0239133 .

Kim, M.-K., Eom, H., Kwon, J.H., et al. (2021). Neural effects of a short-term virtual reality self-training program to reduce social anxiety. *Psychological Medicine*. https://doi.org/10.1017/S0033291720003098.

National Institute for Health and Care Excellence (NICE). 2013. Social anxiety disorder: Recognition, assessment and treatment. Clinical guideline CG159. Available at: https://www.nice.org.uk/guidance/cg159.

Ramsey, K.A., Essoe, J.K.-Y., Boyle, N., et al. (2023). Immersive virtual reality exposures for the treatment of childhood anxiety. *Child Psychiatry & Human Development*.

Salari, N., Heidarian, P., Hassanabadi, M., et al. (2024). Global prevalence of social anxiety disorder in children, adolescents and youth: A systematic review and meta-analysis. *Journal of Prevention* 45(5), 795–813. https://doi.org/10.1007/s10935-024-00789-9.

Shahid, S., Kelson, J.N. & Saliba, A.J. (2024). Effectiveness and user experience of virtual reality for social anxiety disorder: Systematic review. *JMIR Mental Health* 11, e48916. https://doi.org/10.2196/48916 .

Tan, Y.L., Chang, V.Y.X., Ang, W.H.D., et al. (2025). Virtual reality exposure therapy for social anxiety disorders: A meta-analysis and meta-regression of randomized controlled trials. *Anxiety, Stress, & Coping* 38(2), 141–160. https://doi.org/10.1080/10615806.2024.2392195.

Wong, K.P., Lai, C.Y.Y. & Qin, J. (2023). Systematic review and meta-analysis of randomised controlled trials for evaluating the effectiveness of virtual reality therapy for social anxiety disorder. *Journal of Affective Disorders* 333, 353–364. https://doi.org/10.1016/j.jad.2023.04.043.

From Virtual Operating Room to Training Tool: User Experience and Acceptance of an Immersive VR Anesthesia Simulation

Krista Hoek, Elbert Waller, Asmanie Thakoer, Monique van Velzen, Elise Sarton, and Christina Jaschinksi

LUMC – Leiden University Medical Center, The Netherlands

Saxion University, The Netherlands

Corresponding author: k.b.hoek@lumc.nl

Keywords: Virtual reality; anesthesiology; simulation-based training; user experience; embodiment; presence

Introduction

Simulation is an established component of anesthesiology education, allowing clinicians to practice technical and non-technical skills without exposing patients to unnecessary risk (Yunoki, K. & Sakai, 2018). Manikin-based simulation provides realistic hands-on training but requires equipment, facilities, instructors and preparation time. Immersive virtual reality (VR) offers a complementary approach in which clinical environments and scenarios can be reproduced digitally (Huang et al., 2020; Fleet et al., 2025).

In immersive VR, user experience is influenced not only by visual realism but also by *presence*, the subjective sense of "being there", and *embodiment*, including body ownership and agency (Makransky et al., 2017; Roth & Latoschik, 2020). We developed an immersive VR anesthesia simulation and evaluated its deployment among anesthesiologists and anesthesiology residents in a medical conference setting. This pilot focused on acceptance, perceived training value, presence, embodiment and adverse physical effects.

VR Anesthesia Training Application

The application was delivered using a standalone VR headset and handheld controller and places the learner in a virtual operating room containing a virtual patient, anesthesia workstation, patient monitoring, infusion pumps, medication trolley, operating table and ventilator (Figure 1). During the approximately 15-minute scenario, participants move through the virtual operating room and use the controller to interact with selected interactive clinical elements. The workflow progresses from preparing the patient and equipment, applying monitoring and preparing medication-related components, through induction of general anesthesia to the pre-intubation phase.

A separate instructor interface allows the trainer to modify physiological parameters, including blood pressure, oxygen saturation, hemoglobin level, pupil size and skin appearance while the learner remains immersed in the scenario. This learner–instructor setup was designed to allow the

clinical state to change during a scenario and to support future adaptive training scenarios. The educational value of this functionality was not evaluated in the present pilot.



Figure 44. *Immersive virtual operating room and anesthesia training environment.*

Pilot Deployment and Evaluation

The application was tested during the Anesthesiology Days of the Netherlands Society of Anesthesiology in Maastricht, the Netherlands, on 25–26 May 2023. Anesthesiologists and residents participated voluntarily in a dedicated VR setup. Participants received brief instructions before entering the simulation; no extensive onboarding was provided because of the time-limited conference setting. Following the VR session, participants were invited to complete an online post-simulation questionnaire. Completion of the questionnaire was separate from completion of the VR scenario, and participants could omit individual questions. Reasons for non-completion of the post-simulation questionnaire were not systematically recorded.

Acceptance and perceived training value were assessed using purpose-developed 5-point Likert items. Presence was evaluated using the Multimodal Presence Scale (MPS) [4], embodiment using the Virtual Embodiment Questionnaire (VEQ) [5], and nausea, dizziness and disorientation using additional 5-point Likert items. Analyses were descriptive, with Likert outcomes reported as medians and interquartile ranges (IQRs). No inferential statistical testing was performed.

The protocol was approved by the institutional science committee and received a certificate of no objection from the institutional review ethics committee. Participants provided informed consent before participation.

Results

Thirty-four participants completed the VR simulation, of whom 13 (38%) provided post-simulation questionnaire data for analysis as shown in Table 1. Acceptance was highest for VR as a complementary training modality (median 5/5, IQR 4–5). VR as an alternative to manikin-based simulation was rated 4/5 (IQR 4–4), while preference for VR over manikin simulation was 3/5 (IQR 2.5–3).

Presence was moderate: the virtual environment appearing real was rated 3/5 (IQR 3–4), while the sense of acting within the virtual environment rather than operating it from outside was rated 4/5 (IQR 4–4). Movement-related embodiment received relatively high ratings, with virtual movements experienced as one's own rated 6/7 (IQR 3.5–6) and synchrony between real and virtual movements 6/7 (IQR 5–6).

Adverse effects were low: nausea and dizziness were both rated 1/5 (IQR 1–2), and disorientation 2/5 (IQR 2–4). No participant discontinued the simulation because of these symptoms.

Table 6. *Selected outcomes from the pilot evaluation*

Domain	Questionnaire Source	Item / Measure	Median (IQR)
Acceptance	ACC (custom)	VR as complementary training mode	5 (4–5)
	ACC (custom)	Preference for VR over traditional simulation	3 (2.5–3)
Presence	MPS	Interaction felt like interacting with a person	2 (1.5–3)
Embodiment	VEQ	Synchrony between real and virtual movements	6 (5–6)

ACC = Acceptance items. MPS = Multimodal Presence Scale. VEQ = Virtual Embodiment Questionnaire. All values represent median (interquartile range). N = 13.

Discussion and Application Lessons

This pilot demonstrates the deployment of an immersive anesthesia simulation among its intended clinical users in a time-limited professional setting. Acceptance was favorable, adverse effects were low, movement-related embodiment received relatively high ratings, and presence was moderate.

The highest acceptance rating concerned VR as a complementary training modality, whereas preference for VR over manikin simulation was neutral. These items were not statistically compared and should not be interpreted as evidence that participants preferred one implementation strategy over another. The descriptive pattern nevertheless supports considering VR alongside established simulation modalities rather than assuming it should replace them.

The relatively high ratings for movement-related embodiment suggest that correspondence between users' actions and the virtual body worked well, while more moderate environmental and social presence identifies interaction with virtual characters as an area for further development. This is consistent with our previous user-centered evaluation of multiplayer VR anesthesia simulation, in which unfamiliar controls, abstract avatars and limited feedback were identified as barriers to interaction and social presence (Hoek et al., 2025).

Participants completed the short scenario after only brief instructions, demonstrating that completion was feasible without extensive onboarding in this specific, time-limited setting. However, usability and onboarding requirements were not formally evaluated. Our relative work on this matter showed that unfamiliar controls can hinder VR interaction and that familiarization may become increasingly important as scenarios become more complex (Hoek et al., 2025).

This exploratory pilot is limited by its small convenience sample and conference setting. Only 13 of 34 participants provided post-simulation questionnaire data, introducing a risk of nonresponse bias. Consequently, the reported user-experience and acceptance ratings may not represent all users who completed the VR scenario.

Furthermore, learning, clinical performance and transfer to practice were not assessed. Future studies should integrate evaluation more closely into the training workflow and combine user-experience measures with objective learning or performance outcomes in formal training settings.

Conclusion

An immersive VR anesthesia simulation was successfully deployed among anesthesiologists and residents in a professional conference setting. Participants reported favorable acceptance, relatively high movement-related embodiment, moderate presence and few adverse physical effects. These findings support further evaluation of VR as a complementary simulation modality, with

intuitive interaction, onboarding and social responsiveness as relevant considerations for future development.

Conflict of Interest

K. Hoek owns ScenariX, an academic spin-off linked to the VR simulation described in this work. ScenariX did not exist at the time the study was conducted. This relationship is disclosed in the interest of transparency. The other authors declare no competing interests.

Funding and Acknowledgements

This work was supported by Tech For Future, Centre of Expertise for Applied Research in High Tech Systems and Materials.

Use of generative AI: OpenAI ChatGPT (GPT-5.6 Sol) was used to assist with manuscript restructuring and language editing. It was not used for data generation or statistical analysis. All text was reviewed and revised by the authors.

References

Fleet, A., Kaustov, L., Belfiore, E., Kapralos, B., Matava, C., Wiegelmann, J., Giacobbe, P. & Alam, F. (2025). Current clinical and educational uses of immersive reality in anesthesia: Narrative review. *Journal of Medical Internet Research* 27, e62785. https://doi.org/10.2196/62785

Hoek, K., van Velzen, M., Sarton, E. & Jaschinski, C. (2025). User evaluation of a virtual reality simulation for anaesthesiologists-intensivists for non-technical skills. PEC Innovation 7, 100441. https://doi.org/10.1016/j.pecinn.2025.100441

Huang, V.W., Jones, C.B. & Gomez, E.D. (2020). State of the art of virtual reality simulation in anesthesia. *International Anesthesiology Clinics* 58(4), 31–35. https://doi.org/10.1097/AIA.0000000000000298

Makransky, G., Lilleholt, L. & Aaby, A. (2017). Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior* 72, 276–285. https://doi.org/10.1016/j.chb.2017.02.066

Roth, D. & Latoschik, M.E. (2020). Construction of the Virtual Embodiment Questionnaire (VEQ). *IEEE Transactions on Visualization and Computer Graphics* 26(12), 3546–3556. https://doi.org/10.1109/TVCG.2020.3023603

Yunoki, K. & Sakai, T. (2018). The role of simulation training in anesthesiology resident education. *Journal of Anesthesia* 32, 425–433. https://link.springer.com/article/10.1007/s00540-018-2483-y

A Virtual Reality-augmented treadmill platform to improve walking patterns in patients with vestibular diseases

Giacomo Giovanni Moroni[1], Giulia Maria Castiglioni[2,3], Marco Sacco[3], Mario Covarrubias Rodriguez[1], and Sara Arlati[3]

[1]Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy

[2]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

[3]Istituto di Sistemi e Tecnologie Industriali Intelligenti per il Manifatturiero Avanzato (CNR STIIMA), Lecco, Italy

Corresponding author: sara.arlati@cnr.it

Keywords: Virtual Reality; Treadmill; Rehabilitation; Vestibular Diseases;

Introduction

Vestibular disorders comprise a group of conditions that affect the vestibular system, disrupting the perception of motion and body orientation, leading to impaired balance and postural control. These impairments significantly increase the risk of instability and falls, particularly in older adults, and patients frequently develop compensatory strategies such as cautious gait patterns or reduced mobility to avoid provoking symptoms (Cleveland Clinic, 2024). Virtual Reality (VR)-augmented treadmill training represents a rapidly evolving and increasingly well-supported approach to gait rehabilitation across a range of neurological conditions (E. De Keersmaecker et al., 2019). Despite growing use of VR in rehabilitation research, several gaps limit clinical translation: existing VR-treadmill platforms lack intuitive interfaces and real-time control; treadmill, motion-capture and VR management are rarely integrated in one architecture; VR games libraries rarely target vestibular-specific mechanisms, and validated protocols for vestibular disorders remain underrepresented.

This work presents an integrated VR-treadmill platform designed to fill these gaps. The platform combines a real-time bidirectional treadmill interface, a head-mounted display (HMD) equipped with trackers to capture lower body movements, and a library of six VR exercises specifically designed to target vestibular rehabilitation mechanisms, all coordinated through a dual-display architecture that pairs an immersive patient view with a mirrored clinician interface, also displaying exercises' controllable parameters.

System

System overview

The core of the platform is an h-p-cosmos 150/50 treadmill (h-p-cosmos sports & medical GmbH), a Risk Class IIb medical device), selected for its possibility to be controlled via a 3rd party software and dedicated safety infrastructure (fall-prevention harness and emergency-stop system). The treadmill is controlled in real time through the proprietary coscom v4 serial protocol over an RS232 interface, enabling closed-loop, checksum-verified commands for speed and incline.

Patient visualization and motion tracking are provided by an HTC Vive Pro 2 headset, together with three Vive Trackers placed on the feet and waist, enabling full-body inverse-kinematic reconstruction of the patient's motion. The PC-based architecture was chosen to support a dual-interface configuration.

All components are integrated within a Unity-based (Unity 6) VR application, acting as the single hub for rendering, clinician control, and treadmill control.



Figure 45. *Example of the treadmill-based VR platform in use, showing the HMD and motion trackers.*

VR application

The application fulfils three roles: it renders the immersive environment presented to the patient through the HMD; it provides a clinician-facing control interface on an external monitor, mirroring the patient's point of view and allowing real-time adjustment of speed, incline, and VR exercise selection; and it acts as the communication bridge with the treadmill via a dedicated serial manager, maintaining the continuous handshake required by the device. This dual-display, single-hub design

allows session setup, therapeutic supervision, and data acquisition to be managed from one integrated system, all written to a common time base.

The patient-side environment uses a path-pooling system to sustain an effectively infinite, low-latency walking path. The patient's avatar is driven by the tracker and headset data through an inverse-kinematic model, reinforcing the sense of embodiment during walking.

Six VR exercises were implemented, each grounded in a specific, evidence-based vestibular rehabilitation mechanism:

1. *Frontal Occlusion* blocks central vision to promote reliance on proprioceptive input (Cousins et al., 2014).
2. *Eye Batting* introduces periodic full-field blackouts for stroboscopic visual training (Kim et al., 2017).
3. *Visual Feedback Distortion* applies shader-based perturbations to induce graded visual-vestibular conflict (Akiduki et al., 2003).
4. *Visual Cluttering* surrounds the patient with a dynamic particle field, combining habituation with dual-task load (Wibble & Pansell, 2023).
5. *Gaze Following* requires smooth-pursuit tracking of a moving balloon to train gaze stabilization (Hall et al., 2022).
6. *Gaze Stabilisation* requires fixation on spawned balloons to train VOR-gain correction (Herdman et al., 2003).

For each exercise, clinicians can adjust key parameters in real time (e.g. occlusion angle, blink frequency, distortion intensity, particle density, trajectory complexity, spawn rate and dwell time), allowing individualized, progressively graded rehabilitation.



Figure 46. *The clinician control panel for the Visual Cluttering exercise, with adjustable parameters and a live view of the patient's VR environment.*

Limitations

A few limitations should be noted. The system relies on tracker-based motion reconstruction rather than a marker-based setup, which may be less precise for fine gait metrics. Visually intense exercises like Visual Cluttering could also affect tracking performance, something the evaluation will assess. From a clinical standpoint, the planned validation targets healthy participants rather than the target population, so its findings will not directly generalize to patients with vestibular impairment.

Conclusions and future work

The integrated design of the proposed platform, combining treadmill control, VR rendering, motion tracking, and clinician supervision within a single synchronized system, aims to close the gap between research-oriented VR-treadmill setups and tools suited for real clinical use. The platform is fully implemented and operational; the evaluation outlined below constitutes the planned next step and has not yet been carried out.

The platform will be first assessed for usability and user-experience on a cohort of healthy young adults in a single-session protocol, in which a subset of the VR exercises will be administered against unaltered walking baselines, in both real-world and VR conditions, . The validation will assess which gait parameters the platform can reliably measure and influence via visual distortion. The evaluation will be complemented by standardized questionnaires on usability, cybersickness and user experiences.

Future work will extend this evaluation to clinical populations with vestibular impairment. Functional tests with final end-users will help refine acquisition rates to improve the reliability of variability-based gait measures and further explore the platform's dual-task and habituation-based VR exercises as tools for progressive, individualized rehabilitation protocols.

References

Akiduki, H., Nishiike, S., Watanabe, H., Matsuoka, K., Kubo, T., & Takeda, N. (2003). Visual-vestibular conflict induced by virtual reality in humans. *Neuroscience letters*, *340*(3), 197-200. https://doi.org/10.1016/S0304-3940(03)00098-3

Cleveland Clinic. (2024). Vestibular disorders. https://my.clevelandclinic.org/health/diseases/vestibular-disorders. Accessed: 2026.

Cousins, S., Cutfield, N. J., Kaski, D., Palla, A., Seemungal, B. M., Golding, J. F., ... & Bronstein, A. M. (2014). Visual dependency and dizziness after vestibular neuritis. *PloS one*, *9*(9), e105426. https://doi.org/10.1371/journal.pone.0105426

De Keersmaecker, E., Lefeber, N., Geys, M., Jespers, E., Kerckhofs, E., & Swinnen, E. (2019). Virtual reality during gait training: does it improve gait function in persons with central nervous system movement disorders? A systematic review and meta-analysis. *NeuroRehabilitation*, *44*(1), 43-66. https://doi.org/10.3233/NRE-1825

Hall, C. D., Herdman, S. J., Whitney, S. L., Anson, E. R., Carender, W. J., Hoppes, C. W., ... & Woodward, N. J. (2021). Vestibular rehabilitation for peripheral vestibular hypofunction: an updated clinical practice guideline from the Academy of Neurologic Physical Therapy of the American Physical Therapy Association. *Journal of neurologic physical therapy*, *46*(2), 118. https://doi.org/10.1097/NPT.0000000000000382.

Herdman, S. J., Schubert, M. C., Das, V. E. & Tusa, R. J. (2003). Recovery of dynamic visual acuity in unilateral vestibular hypofunction. Archives of Otolaryngology–Head & Neck Surgery, 129(8), 819–824. https://doi.org/10.1001/archotol.129.8.819

Kim, K.-M., Kim, J.-S. & Grooms, D. R. (2017). Stroboscopic vision to induce sensory reweighting during postural control. *Journal of Sport Rehabilitation*, 26(5). https://doi.org/10.1123/jsr.2017-0035

Wibble, T., & Pansell, T. (2023). Clinical characteristics of visual motion hypersensitivity: A systematic review. Experimental Brain Research, 241(7), 1707–1719. https://link.springer.com/article/10.1007/s00221-023-06652-3

A Virtual Reality Application for Enhancing BCI Motor Imagery Training through Multisensory Stimulation

Alessandro Nardi[1], Fortuna Galdieri[2], Antonio Esposito[2,3], Marco Sacco[1,3]
Pasquale Arpaia[2,3,4], and Sara Arlati[c]

[1]*Istituto di Sistemi e Tecnologie Industriali Intelligenti per il Manifatturiero Avanzato (CNR STIIMA), Lecco, Italy*

[2]*Department of Electrical Engineering and Information Technology (DIETI), Università degli Studi di Napoli Federico II, Naples, Italy*

[3]*Human-Machine Interfaces for Measurements in Living Environments (HIMaLivE), Naples, Italy*

[4]*Centro Interdipartimentale di Ricerca in Management Sanitario e Innovazione in Sanità(CIRMIS), Naples, Italy*

Corresponding author: alessandronardi1@cnr.it

Keywords: Extended reality; haptic feedback; brain-computer interface; electroencephalography; healthcare/wellness; immersive experience; embodiment

Introduction

In Motor Imagery (MI)-based brain–computer interfaces (BCIs), the user mentally rehearses a movement, recruiting sensorimotor areas that are also engaged during execution (Halder et al., 2011). This makes BCIs attractive for neurorehabilitation, as they could promote brain plasticity even in patients with severe motor disability (Cervera et al., 2018). Most of these interfaces are based on EEG: being non-invasive, they hold potential for alternative ways of communication, translating neural signals into text or commands. The limiting factor is the need for repeated sessions to train the EEG signal decoding algorithms, resulting in long, unmotivating experiences for the user. Another issue is that the sensorimotor system is engaged only in part (Škola & Liarokapis, 2023): conventional training renders an abstract indicator (e.g., an arrow) not resembling how the imagined movement would feel and makes it difficult to imagine a proper kinesthetic simulation that would ease the system's tuning (Alimardani et al., 2016). Instead, current haptic feedback in BCIs has been explored only to a limited extent (Fleury et al., 2020), and no platform delivers three channels (visual, pressure, and temperature) online and is driven by the decoder.

To address these issues, this paper presents an immersive Virtual Reality (VR)-based training platform for MI that delivers visual, pressure-tactile, and thermal feedback concurrently and in real time. The aim is to exploit the potential of multisensory stimulation to improve embodiment and, in turn, the quality of MI signals produced. The expectations are to reduce training time and make the whole process more motivating and engaging for the users.

System

System overview

The platform closes a loop between a wearable EEG-based mental decoder and an immersive VR environment. EEG signals are acquired, classified online, and converted into synchronized visual and haptic stimuli, which are returned to the user via a head-mounted display and a pair of haptic gloves.

EEG is recorded with a Unicorn Hybrid Black cap (g.tec, Austria) with eight hybrid electrodes used in wet mode. The default layout was rearranged into a sensorimotor-centered montage — Fz, FC3, FC4, C5, Cz, C6, CP3, and CP4 — compliant with the 10–10 system and reused from an earlier online pipeline (Arpaia et al., 2023). Samples stream over Bluetooth to a Windows workstation (Intel Core i7-12700H, 16 GB RAM, NVIDIA RTX 3070 Ti), on which acquisition, calibration, and online classification run in MATLAB/Simulink. The processing algorithm is based on a filter bank common spatial pattern extracting frequency-spatial features, then classified by a linear discriminant analysis. It is trained with calibration data acquired in controlled conditions with a synchronous BCI protocol.

The visual scene is presented on a Meta Quest 3, driven over a wired Quest Link for stable throughput, while pressure and thermal cues are rendered by a WeArt TouchDiver Pro (WeArt, 2026) that actuates the five fingertips and the palm across a thermal range of 18–42 °C, connected wirelessly through its USB transceiver.

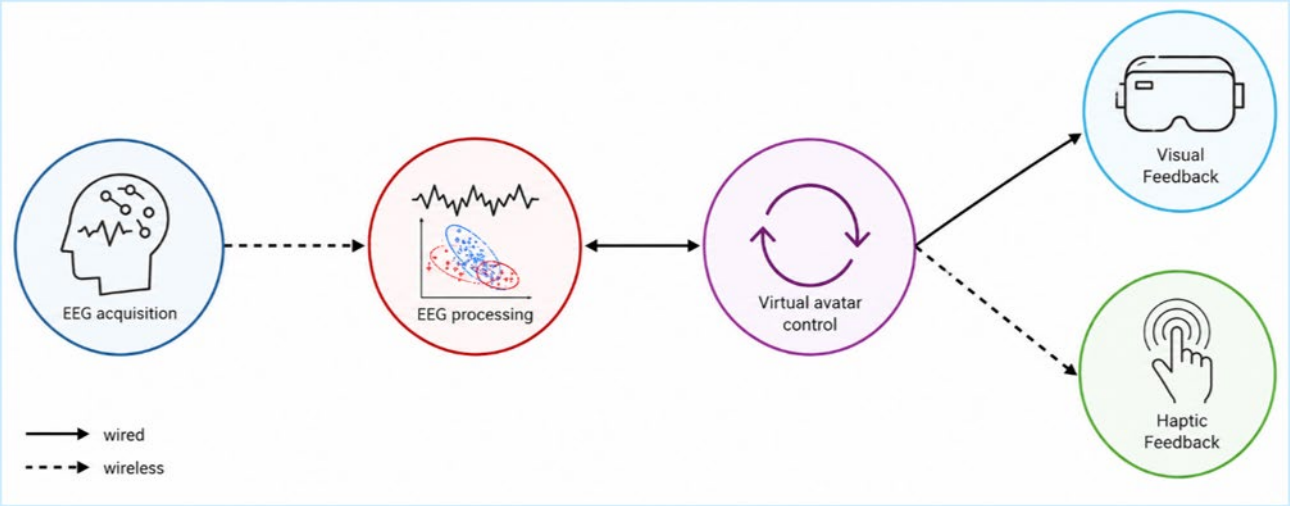


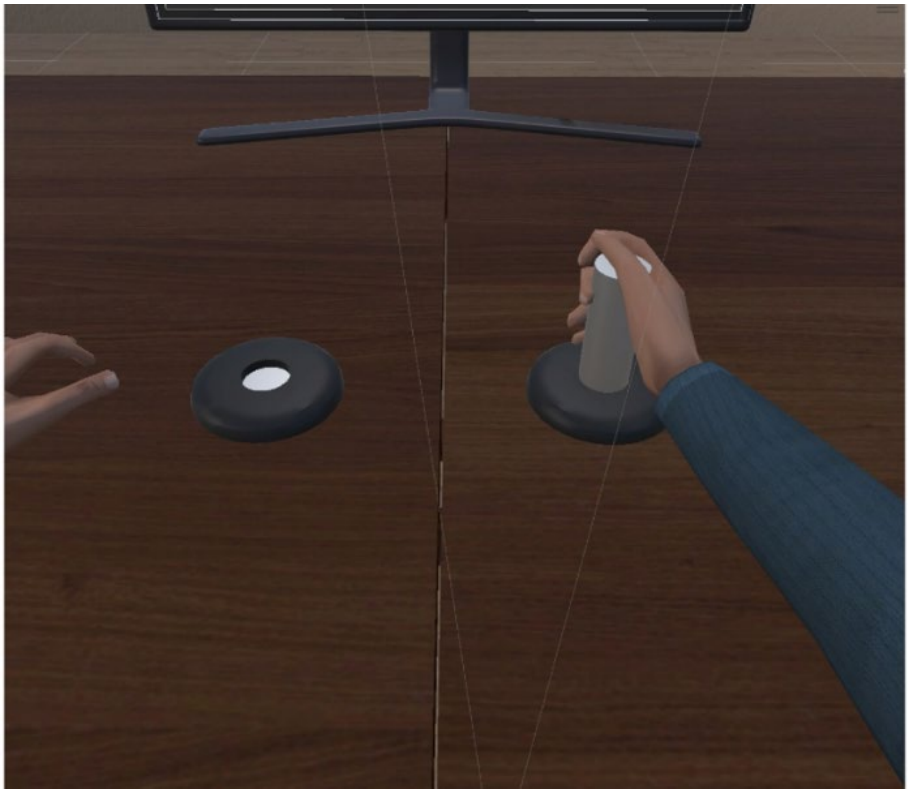
Figure 47. *System overview.*

Signal processing and the VR application communicate over UDP. The link is bidirectional: the decoder sends the predicted MI class and its confidence to Unity to drive the avatar and glove commands, and the application returns trial-event timestamps and the online performance metrics, which are logged in a register with the EEG stream.

Virtual environment and feedback design

The environment was built in Unity 6 and deliberately kept calm and uncluttered to limit distractions throughout the session. A first-person avatar is used to induce embodiment, the subjective sense that the virtual body is one's own, under one's control, and located where one is (Guy et al., 2023), as seeing that body carry out the imagined action has been associated with faster and more accurate generation of the target cortical patterns (Choi et al., 2020). A virtual monitor in front of the avatar displays task text. The desk is split into two symmetric halves, and a metallic bar can rise from an opening on either side. The user is instructed to perform MI according to the side indicated by the bar rising, i.e., using the right or the left hand.

a)



b)

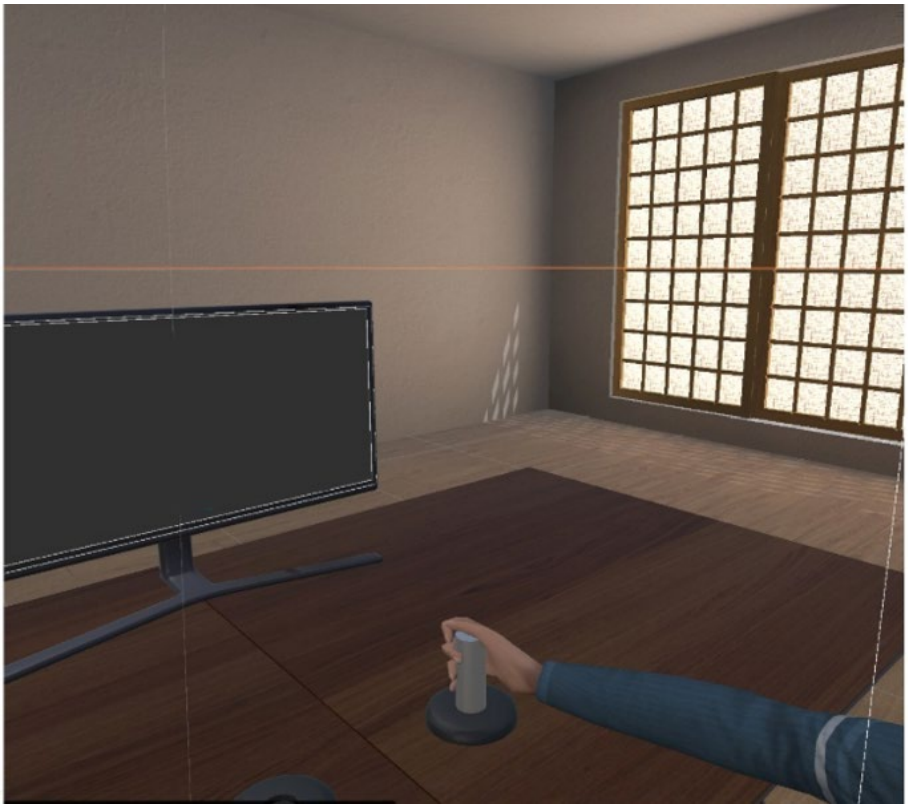


Figure 48. *a. view of the grab action b. view of the room*

The MI is detected in agreement with what was reported by Arpaia et al. (2023). When the MI is detected, the avatar hand executes the imagined grasp, while the haptic glove applies pressure and cold to the corresponding fingertips at the moment the avatar closes on the object. In detail, the target arm moves only when the decoded class matches the cued hand, and confidence sets its speed. To keep early trials rewarding while the user is still learning, the first N = 10 trials are assisted. For these trials (t = 1, …, N), a decaying assistance term h = 1 − t/N is added: the boosted confidence

is computed as confidence + h(1 − confidence), truncated to [0, 1], and applied only when the decoded class is correct. Assistance is strong at first and decreases linearly to zero at t = N; no assistance is applied after the N-th trial. The command is smoothed by a five-sample moving average; when the decoded class is wrong, the raw confidence enters the buffer as a negative value, which drives the average down rather than moving the arm backward to keep realistic feedback and not weaken the embodiment.

Conclusions and Future Work

This work presents a VR-based system to enhance MI training in the context of BCI. The coherent combination of different types of feedback represents the most innovative aspect. The system is expected to increase embodiment and, therefore, enhance the user's ability to perform MI, thereby improving EEG signal quality and reducing training time.

In the near future, studies with healthy participants and patients with motor disabilities will be conducted to assess the effectiveness of such a system in improving BCI performance and inducing embodiment at a subjective level.

References

Alimardani, M., Nishio, S., & Ishiguro, H. (2016). The Importance of Visual Feedback Design in BCIs; from Embodiment to Motor Imagery Learning. PLOS ONE, 11(9), e0161945. https://doi.org/10.1371/journal.pone.0161945

Arpaia, P., Esposito, A., Moccaldi, N., Natalizio, A., & Parvis, M. (2023). Online processing for motor imagery-based brain-computer interfaces relying on EEG. 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 01–06. https://doi.org/10.1109/I2MTC53148.2023.10176052

Cervera, M. A., Soekadar, S. R., Ushiba, J., Millán, J. del R., Liu, M., Birbaumer, N., & Garipelli, G. (2018). Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. Annals of Clinical and Translational Neurology, 5(5), 651–663. https://doi.org/10.1002/acn3.544

Choi, J. W., Huh, S., & Jo, S. (2020). Improving performance in motor imagery BCI-based control applications via virtually embodied feedback. Computers in Biology and Medicine, 127. https://doi.org/10.1016/j.compbiomed.2020.104079

Fleury, M., Lioi, G., Barillot, C., & Lécuyer, A. (2020). A Survey on the Use of Haptic Feedback for Brain-Computer Interfaces and Neurofeedback. Frontiers in Neuroscience, 14. https://doi.org/10.3389/fnins.2020.00528

Guy, M., Normand, J.-M., Jeunet-Kelway, C., & Moreau, G. (2023). The sense of embodiment in Virtual Reality and its assessment methods. Frontiers in Virtual Reality, 4. https://doi.org/10.3389/frvir.2023.1141683

Halder, S., Agorastos, D., Veit, R., Hammer, E. M., Lee, S., Varkuti, B., Bogdan, M., Rosenstiel, W., Birbaumer, N., & Kübler, A. (2011). Neural mechanisms of brain–computer interface control. NeuroImage, 55(4), 1779–1790. https://doi.org/10.1016/j.neuroimage.2011.01.021

Škola, F., & Liarokapis, F. (2023). Virtual Reality Embodiment in Motor Imagery Brain–Computer Interface Training. SN Computer Science, 4(1). https://doi.org/10.1007/s42979-022-01402-8

WeArt. (2026). Touch Diver PRO. https://weart.it/

ACTA-VR: Iterative Design of Ecological VR Tasks for Cognitive Flexibility in Older Adults

Chiara Pupillo[1,2], Vera Colombo[3,4], Silvia Bolognini[4], Giuseppe Prencipe[1], and Giuseppe Riva[2,5]

[1] *Department of Computer Science, University of Pisa, Largo B. Pontecorvo, 3, Pisa, Italy*

[2] *Humane Technology Lab, Catholic University of Sacred Heart, Largo Gemelli, 1, Milan, Italy*

[3] *Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council, Lecco, Italy*

[4] *NeMO Lab, ASST GOM Niguarda Cà Granda Hospital, Milan, Italy*

[5] *Applied Technology for Neuro- Psychology Laboratory, IRCCS Istituto Auxologico Italiano, Via Magnasco 2, Milan, Italy*

Corresponding author: chiara.pupillo@phd.unipi.it

Keywords: Cognitive flexibility; virtual reality; ecological assessment; older adults; user experience

Abstract

Instrumental activities of daily living are ecological indicators of executive dysfunction but are difficult to observe in real life under standardized, instrumented conditions. Immersive virtual reality (VR) can combine ecological validity with experimental control. We present ACTA-VR, the immersive application of the VR-FlexAI project, designed to assess cognitive flexibility through everyday activities while preserving meaningful freedom of action. In Task 1 (Adaptive Savory Breakfast), participants choose among multiple valid cooking strategies, handle a medication-prompt interruption, and resume the cooking task. In Task 2 (Shifting Food Categorization), they discover any of three sorting rules for organizing groceries, maintain it, and then shift to another. In the first evaluation with 19 cognitively healthy, VR-naïve older adults, observations motivated two refinements. First, container-based sorting replaced free spatial clustering in Task 2 to reduce motor-spatial demands at rule transitions. Second, limitations of a usability-focused assessment motivated the development of a broader VR-specific user experience questionnaire. Criterion-order patterns further suggested that ecological framing may influence the salience of spontaneous rules. We retrospectively translate these observations into preliminary, transferable considerations for VR-based cognitive assessment: preserve construct-relevant freedom while constraining interaction mechanics, pilot scenarios for unintended salience effects, and test both interfaces and instruments with older target users.

Introduction

Instrumental activities of daily living (IADLs), such as meal preparation, are among the most sensitive real-world indices of executive dysfunction, often revealing impairment well before it is detectable on decontextualized neuropsychological tests (Diamond, 2013). Direct observation of IADLs in real homes could close this gap, but it is costly to standardize, difficult to replicate, and rarely instrumented with the precision required for quantitative behavioral analysis.

Immersive virtual reality (IVR) offers a trade-off solution, integrating participants into naturalistic, goal-directed scenarios while maintaining full control over event timing, stimulus properties, and data acquisition (Parsons, 2015). Because every action in VR can be recorded, parameterized, and timestamped, it also supports analysis with emerging computational and artificial intelligence (AI)-based approaches to neuropsychological assessment, from automated assessment of digitally administered protocols to multimodal digital phenotyping of cognitive decline (Libon et al., 2024).

This is particularly consequential for cognitive flexibility, i.e., the capacity to disengage from an established mental set and reconfigure behavior when goals or rules change (Diamond, 2013; Miyake et al., 2000; Monsell, 2003). Cognitive flexibility is especially vulnerable to age-related decline.

To address the need for an ecologically grounded yet standardized assessment of cognitive flexibility in aging, the VR-FlexAI project combines an IVR application, called ACTA-VR, and neuropsychological, behavioral, and linguistic measures within a multimodal AI approach (Pupillo et al., 2026). The first ACTA-VR deployment was used to establish feasibility and to identify where interface mechanics and evaluation instruments could distort the expression of cognitive flexibility in a sample of cognitively healthy, VR-naïve older adults.

This paper focuses on the iterative development process underlying the feasibility study, showing how deployment with the target population informed the redesign of both the VR interaction environment and the user experience measurement approach.

ACTA-VR: application and design logic

ACTA-VR was developed in Unity 6 (version 6000.0.25f1) and deployed on Meta Quest 3. The interaction relies on hand tracking; a text-to-speech voice guides the user by explaining the tasks. The application includes two tasks based on two cognitive processes underlying cognitive flexibility: task switching (Task 1, Adaptive Savory Breakfast) and set shifting (Task 2, Shifting Food Categorization). Each task is preceded by a brief familiarization phase using the same interaction mechanics. In both tasks, participants are free to choose how to organize their actions, provided they achieve the task goal and follow the rules applicable to the current situation. Task 1 was performed in a standing configuration that allowed participants to move freely within the available interaction space, supporting naturalistic and embodied interaction with the virtual environment. The initial version of Task 2 was instead administered while participants were seated. During the refinement phase of Task 2, internal testing revealed that introducing sorting baskets reduced object visibility and the interaction area when seated. Consequently, the refined version was administered in a standing configuration. Between the two tasks, participants completed a brief seated cognitive decompression phase in a neutral virtual environment, accompanied by soft background piano music, before proceeding to Task 2.

Task 1 - Adaptive savory breakfast

Participants prepare a savory breakfast including coffee, eggs, bread, and salad, guided by a board listing the required actions while leaving their sequence unconstrained. They may start with the moka, the egg, or manage both concurrently, with all strategies considered valid. Forty seconds into the task, a medication prompt interrupts the activity, requiring participants to decide whether to take the medication immediately or after breakfast, and then resume their plan. This decision determines the final plating rule: immediate medication requires avoiding green leafy vegetables; i.e., the final breakfast must not include salad. The task, therefore, places demand on goal switching,

interruption management, resumption, and context-dependent adaptation without scoring adherence to a single procedural order (Altmann & Trafton, 2002; Monsell, 2003). The freedom-preserving task structure is illustrated in **Figure 1**.

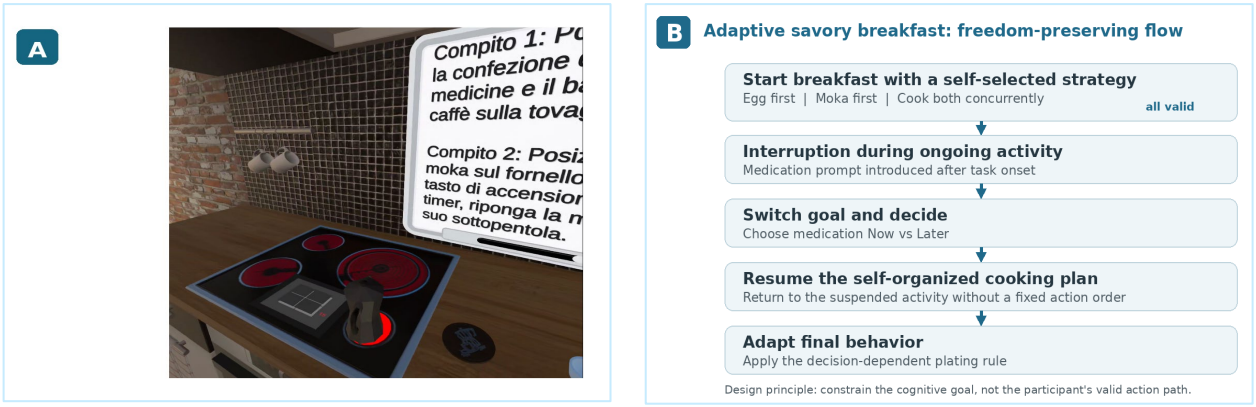


Figure 1. *ACTA-VR Task 1 - Adaptive savory breakfast. (A) Immersive kitchen during moka preparation. (B) Freedom-preserving task structure: participants select among multiple valid cooking strategies before the interruption and resume their own plan after the medication decision.*

Task 2 - Shifting Food Categorization

The second task was inspired by Wisconsin Card Sorting Test (WCST)-like set-shifting principles, retaining color as a grouping dimension while introducing ecologically meaningful criteria. Participants unload groceries by interacting with nine food objects that must be grouped according to three possible grouping rules. Participants must identify three valid rules: color (white, red, orange), storage location (pantry, fridge, freezer), and semantic category (fish, cereals, dairy products). The initial rule is unconstrained: participants may start with whichever criterion they identify first, but once selected, that criterion must be maintained throughout the round. After completing the first group, they must disengage from that mindset and identify a different valid rule, in line with set-shifting models of executive control (Miyake et al., 2000). Since the criterion order is not prescribed, the discovery sequence itself is exploratory behavioral information.

First evaluation and measures

The study was approved by the University of Pisa Bioethics Committee (No. 46/2025). All participants provided informed consent through an electronically signed online consent form. Eligibility was assessed through a telephone pre-screening that included the Itel-MMSE (Quaranta et al., 2025) and structured questions regarding inclusion and exclusion criteria. The first phase of the project assessed the feasibility, more specifically, the usability of the ACTA-VR application and the adequacy of the evaluation instruments with cognitively healthy participants. 20 cognitively healthy older adults were enrolled in this evaluation, ten in each phase. One participant in the refined phase did not complete the VR assessment and was excluded from the analyses because no usable VR data was available. The analytical sample, therefore, comprised 19 participants (age 74.2 ± 4.8 years, range 67–83), almost all of whom were first-time VR users, with ten participants in the initial phase and nine in the refined phase, which included the revised Task 2 and an ad hoc questionnaire replacing the System Usability Scale. Because the redesign was introduced iteratively rather than experimentally assigned, and the two Task 2 versions also differed in interaction configuration, between-version comparisons are descriptive.

Subjective experience was assessed with the System Usability Scale (SUS; Brooke, 1996) and the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993). Some SUS items required clarification for this older sample, echoing known design challenges for IVR with older users. Since spontaneous reactions also referred to enjoyment, immersion, and novelty in addition to usability, and no short, validated VR-specific instrument, to our knowledge, offered similar coverage, a

different questionnaire was included in the refined phase. The questionnaire is a VR-specific user experience questionnaire (VR-UXQ) developed specifically for this study, with six subscales and two summary indices scored from 1 to 5. After the VR experience, 18 participants completed a task-related object knowledge questionnaire, and informal notes on interpretation or interaction issues were recorded. In the initial phase, SUS was 61.0 ± 19.1 out of 100 (n=10), indicating substantial variability and room for refinement. SSQ scores were low for most participants in both phases, with a median of 0 in each (initial: 4.87 ± 7.07, n=10; refined: 10.39 ± 16.49, n=9). In the refined phase, VR-UXQ ratings were highest for Engagement (4.11 ± 0.47) and Usability (4.00 ± 0.36), while the overall User Experience index was 3.89 ± 0.54 and Flow was comparatively lower (3.17 ± 0.83; n=9).

While Task 1 required no substantial changes, observations of user performance informed the development of a refined version of Task 2.

Performance-informed refinement of Task 2

In the initial Task 2 version, nine food items were arranged on a table, and each solution meant physically constructing three free spatial groups (panel A, **Figure 2**). At each rule change, the configuration had to be dismantled and rebuilt in limited space, so moments meant to reveal set shifting also imposed motor-spatial and virtual-physics demands that could be confused with the cognitive transition itself. The initial and refined response formats are shown in **Figure 2**.

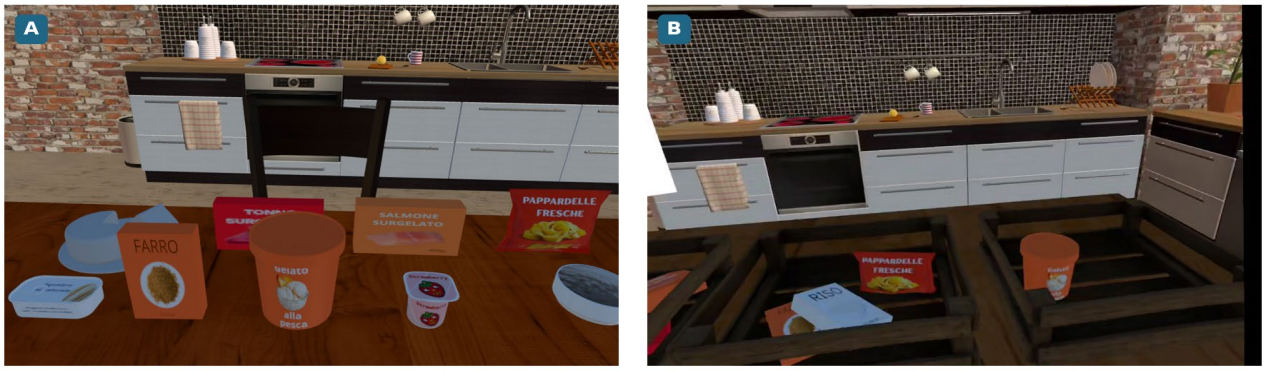


***Figure 2**. Task 2 response-format redesign. (A) Initial free spatial clustering. (B) Refined container-based sorting. Strategic freedom was preserved, while the physical response format was structured to reduce motor-spatial reorganization at rule transitions.*

The refined version introduced a common source box and three baskets: at a rule change, baskets could be emptied and objects re-sorted, preserving rule discovery and switching while constraining only the physical response space (panel B, **Figure 2**). Mean completion time was descriptively lower after redesign (963.1 ± 316.9 in the initial phase vs 731.4 ± 218.1 s in the refined phase; n=9 per phase; medians 877 vs 706 s). By comparing the results of the SUS and the Usability item of the VR-UXQ, it can be stated qualitatively that usability improved between the initial version and the revised version of ACTA-VR, rising from 61 out of 100 to approximately 80 out of 100 (converting the score of 4.00 out of 5 to a 100-point scale).

Ecological task interpretation

At the descriptive level, semantic category was identified first most often (9/19, 47.4%), followed by storage (6/19, 31.6%) and color (3/19, 15.8%); among the 17 completing all three criteria, color was most often identified last (9/17, 52.9%). Color was nonetheless eventually identified by 18/19 participants, and post-task checks confirmed largely correct knowledge of color, storage, and semantic category, indicating that the pattern reflects reduced spontaneous color salience rather than an inability to recognize it. Informal experimenter notes converged with this pattern, suggesting that the grocery unpacking narrative may have shifted attention toward semantically or functionally

meaningful properties. This is noteworthy because Task 2 was inspired by WCST-like set-shifting principles, in which color is typically an explicit perceptual dimension. Although exploratory, the finding suggests that ecological storytelling in VR may alter rule salience when conventional cognitive paradigms are translated into immersive contexts.

Lessons from the first evaluation: implications for VR cognitive assessment

Although ACTA-VR was born out of neuropsychological needs, the implementation and feasibility study highlighted tensions between task requirements, immersive interaction, and measurement, which became evident only with the target users. Therefore, we retrospectively translate these observations into preliminary considerations for cognitive assessment in VR (see **Table 1**).

Table 1. *Retrospective lessons from ACTA-VR deployment and implications for VR cognitive assessment.*

ACTA-VR observation	Assessment implication	Transferable consideration
Multiple valid strategies were possible, whereas free spatial clustering created additional demands at rule changes.	Fixed sequences may penalize valid strategies, while interaction freedom can introduce construct-irrelevant costs.	Preserve construct-relevant strategic freedom but constrain response mechanics where manipulation demands may overlap with the cognitive process being assessed.
Semantic and functional rules emerged more readily than color.	Ecological context may alter the spontaneous salience of otherwise valid task dimensions.	Pilot immersive scenarios for unintended salience effects and use explicit knowledge checks to distinguish contextual salience from lack of knowledge.
Familiar actions elicited expectations of real-world-consistent interaction. For example, in the grocery-sorting task, some participants appeared to privilege functionally meaningful criteria, such as storage location or food category, over color, consistent with how groceries may be organized in everyday life. Such mismatches interrupted the flow of the activity.	Interaction inconsistencies may redirect attention from the cognitive task and contaminate performance.	Prioritize stable, predictable affordances that are consistent with real-world interactions, and treat interaction breakdowns as a potential source of measurement interference.
Most participants were VR-naïve older adults, yet they engaged with complex immersive tasks; difficulties concerned specific interactions and questionnaire wording rather than VR use per se.	Limited prior VR experience should not be equated with an inability to perform immersive assessment.	Design for learnability and predictable interaction, and pilot both interfaces and evaluation instruments with older target users rather than assuming age-related technological limitations.

Discussion and conclusion

This paper presents the first deployment and evaluation of ACTA-VR, an IVR application for cognitive flexibility assessment developed within the VR-FlexAI project. The ACTA-VR deployment demonstrated that ecologically grounded IVR cognitive assessment entails more than merely reproducing familiar activities. In the initial version of Task 2, free spatial clustering introduced motor-spatial demands precisely at rule transitions, while the grocery unpacking narrative appeared to increase the salience of semantic and functional criteria relative to color. Thus, both interaction mechanics and contextual framing became part of the cognitive task itself. For VR-based assessment, this suggests preserving behavioral freedom when it is construct-relevant, while constraining interaction demands that may introduce unrelated variability. It also highlights that ecological realism is not cognitively neutral, as familiar contexts may alter the spontaneous salience of task dimensions.

The deployment also informs the design of VR cognitive assessment for older adults. Most participants were VR-naïve yet engaged with complex, goal-directed tasks. Each task was preceded by a simple familiarization phase based on the same interaction logic as the subsequent assessment, likely facilitating interaction learning before cognitive demands increased. The main difficulties, therefore, concerned specific mechanics rather than VR exposure itself. Short, task-consistent familiarization may consequently help separate interaction learning from the cognitive process being assessed. Overall, ACTA-VR illustrates how feasibility testing with intended users can guide the alignment of interaction design, ecological context, and cognitive assessment requirements.

References

Altmann, E.M., & Trafton, J.G. (2002). Memory for goals: an activation-based model. *Cognitive Science*, 26(1), 39-83. https://doi.org/10.1207/S15516709COG2601_2

Brooke, J. (1996). SUS: A 'Quick and Dirty' Usability Scale. In P.W. Jordan, B. Thomas, I.L. McClelland, & B. Weerdmeester (eds.), *Usability Evaluation in Industry*, pp. 189-194. London: Taylor & Francis.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64, 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3), 203-220. https://doi.org/10.1207/s15327108ijap0303_3

Libon, D. J., Swenson, R., Price, C. C., Lamar, M., Cosentino, S., & Bezdicek, O. (2024). *Digital assessment of cognition in neurodegenerative disease: a data driven approach leveraging artificial intelligence. Front Psychol. 2024; 15: 1415629*.

https://doi.org/10.3389/fpsyg.2024.1415629

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: a latent variable analysis. *Cognitive Psychology*, 41(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Monsell, S. (2003). Task switching. Trends in Cognitive Sciences, 7(3), 134-140. https://doi.org/10.1016/S1364-6613(03)00028-7

Parsons, T.D. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Frontiers in Human Neuroscience*, 9, 660. https://doi.org/10.3389/fnhum.2015.00660

Pupillo, C., Colombo, V., Prencipe, G., & Riva, G. (2026). VR-FlexAI: A Multimodal Approach for Assessing Cognitive Flexibility in Aging. *Cyberpsychology, Behavior, and Social Networking*, 29(8), 601-604. https://doi.org/10.1177/21522715261457493

Quaranta, D., L'Abbate, F., Pelosi, A., Arighi, A., Asoni, G., Bagattini, C., ... & Caffarra, P. (2025). Itel MMSE: a short phone screening test for cognitive decline. Italian Validation study by the SINdem Neuropsychology Working Group. *Neurological Sciences*, *46*(4), 1617-1627. https://doi.org/10.1007/s10072-024-07863-4

XR Compliance Matrix: Linking Application Contexts, Requirements and Verification Pathways

Christoph Runde, and Saahil Khanna

Virtual Dimension Center (VDC) w.V.

Corresponding author: christoph.runde@vdc-fellbach.de

Keywords: Extended Reality (XR); Compliance; Requirements Traceability; Verification; Standards; Regulatory Requirements

Introduction

Extended reality (XR) is moving from proof-of-concept and pilot environments into regular industrial and medical use. Regulatory sandboxes, where applicable, can support supervised experimentation, but they do not replace the requirements governing subsequent market entry or operational deployment [European Parliament and Council, 2024]. At operational deployment, all requirements applicable to the intended use must be identified and addressed: technical performance, ergonomics, occupational safety, data protection, product safety, and sector regulation; sandbox experience can accelerate but not replace this conformity assessment [European Parliament and Council, 2024]. For some companies, this transition is where orientation is lost: XR solutions that look convincing get deployed and are only later found inadmissible, causing rework or withdrawal, or investment is withheld because admissibility cannot be assessed in advance - a trust gap inhibiting adoption in both directions. The challenge is to identify applicable requirements, trace them to authoritative sources, and determine verification pathways early. This paper asks how context-specific XR requirements can be linked systematically to sources, measurable criteria, and verification methods. The XR Compliance Matrix (XRCM) addresses this by transferring established requirements-traceability and conformity-assessment principles into a cross-sector XR knowledge and navigation tool supporting exactly this sandbox-to-market transition; testing, evidence generation, and conformity assessment remain downstream.

State of the Art in Research and Technology

XR research provides requirements frameworks, KPI-based evaluation, technology-selection models, and requirements-driven validation, e.g., for medical VR-HMDs [Pedram et al., 2023], industrial AR [Quandt et al., 2018; Gerhard & Siewert, 2024], and context-weighted technology selection [Keiser et al., 2025]. In this reviewed literature, no approach was identified that links use-case-dependent requirement selection to normative traceability across sectors. Outside XR, this full chain is established practice: requirements traceability distinguishes a requirement's origin from its downstream verification [Gotel & Finkelstein, 1994]; regulatory requirements engineering operationalizes legal text into checkable requirements [Breaux et al., 2006]; and mature regulated domains formalize applicability, means of compliance, and evidence, e.g., EU MDR [European Parliament and Council, 2017] and aviation certification [FAA, 2017]. The XRCM contribution is

therefore not a new function but the cross-sector transfer of these principles into XR: Context - Authoritative Source - Requirement - Property/KPI - Verification Method.

Development and Structure of the XR Compliance Matrix

The XRCM was built iteratively from practical questions on technical suitability and regulatory admissibility, through three routes: (1) use-case-driven deduction; (2) analysis of normative and technical documents, including standards registers, for requirements, KPIs, and tests; and (3) reverse analysis of certified, qualified, or regulatorily cleared XR systems to trace intended use, approval scope, requirements, and evidence. Two design principles guided requirement inclusion: source-boundedness (priority for requirements traceable to an authoritative document - law, regulation, standard, technical rule, or official guideline - over freely formulated or literature-derived ones, keeping origin and binding status verifiable) and verifiability (preference for requirements with a measurable characteristic, KPI, or defined verification method: test, analysis, inspection, or demonstration). Methodologically, this combines structured multi-source document analysis with principles from requirements traceability and regulatory requirements engineering [Gotel & Finkelstein, 1994; Breaux et al., 2006].

The XRCM distinguishes industry-specific, application-specific, and general requirements, each broken down into an XR property (e.g., color fidelity, tracking, haptics) tied to a use case or vertical (e.g., healthcare, automotive, aerospace, construction, training). At the time of writing, it comprises 246 individually sourced entries across 55 aspects and 14 use-case/vertical categories: 191 general, 28 industry-specific, 27 application-specific. Its core data model comprises 13 fields connecting use context and requirement to characteristics, KPIs, testing, standards, and regulatory context (Table 1). Every entry also references concrete, named XR systems - from consumer HMDs to enterprise, medical, and defense systems - together with the basis on which they were evaluated, qualified, or cleared, spanning more than 200 XR products and XR platforms [VDC, 2025]: not only which requirements apply, but which available systems already demonstrate compliance, and on what basis.

***Table 1**. Abridged example of a complete XRCM data set [VDC, 2025].*

XRCM Field	Example: color fidelity
Requirements class	Application-specific requirements
Use cases and verticals	Design, Engineering & Planning Applications
Properties of XR	Color fidelity
Aspect	Color true
Requirements	Accurate, calibrated, perceptually consistent color reproduction
Characteristics	Color accuracy; luminance; chromaticity; consistency; stability; calibration quality
KPIs	xy chromaticity; luminance Y; ΔE; task/user metrics where applicable
Testing scope	Color rendering; white point/calibration; differentiation; perceived realism
Testing methodology	Spectrophotometer/colorimeter; reference comparison; calibration; user tests
Main standard	Colorimetry & optical display measurements
Applicable standards	CIE 1931, 1976; ISO 11664; IEC 62341-6-3, 63145-20/-22; IDMS; ISO 9241-305/-307
Regulatory context	No universal XR-specific acceptance threshold; application-specific criterion required
Example systems/devices	e.g., Apple Vision Pro; Varjo XR-3/-4; Meta Quest 3; Microsoft HoloLens 2

Evaluation through Two Example Use Cases

Use Case 1: Industrial Design in Mechanical Engineering

The first scenario is HMD-based product design, design review, and virtual sampling in white-goods mechanical engineering at a standard EU office workplace. XRCM screening identifies 21 potentially relevant requirement areas (18 direct, three conditional), including design fidelity, ergonomics, occupational safety, data protection, security, and product safety [VDC, 2025]. Occupational safety and cervical stress: ASR A6 requires low, balanced HMD weight and unhindered natural head posture and movement [Committee for Workplaces, 2024]; UL 8400 addresses biomechanical stress [UL Standards & Engagement, 2023]; the XRCM links these to mass, center of gravity, and biomechanical load. For color fidelity, the XRCM uses chromaticity, luminance, and ΔE: ISO/CIE 11664-6:2022 provides the CIEDE2000 measurement method for displays simulating reflective or transmissive objects [ISO/CIE, 2022], but no universally binding XR-specific acceptance threshold exists for virtual sampling - the XRCM makes such normative gaps explicit and locatable, turning an implicit unknown into an explicitly documented compliance gap.

Use Case 2: Intraoperative Surgical Navigation and Visualization

The second scenario is intraoperative AR visualization and navigation. The FDA classifies "*Orthopedic Augmented Reality*" as Class II (SBF) [FDA, 2025]; the cleared SpineAR SNAP configuration K243623 includes HoloLens 2 and Magic Leap 1 and defines positional and angular error criteria [FDA, 2024] - HoloLens 2 achieved mean errors of 0.76 mm and 1.73° against limits of <2.0 mm and <2°. This reverse-traces a cleared configuration to intended use, KPIs, and performance tests, not general medical certification of the HMD. For hygiene, EU MDR Annex I independently requires reduction of infection risks and, where necessary, safe cleaning, disinfection, and/or re-sterilization [European Parliament and Council, 2017]; the XRCM maps cleanability, disinfection resistance, material compatibility, and contamination risk to a separate verification pathway.

Table 2. *Four exemplary requirement-to-verification paths.*

Use case	Example Requirements	Source	KPI / criterion	Verification path	XRCM result
Industrial design	Cervical stress on Designer	ASR A6; UL 8400	Mass; center of gravity; biomechanical load	Weight/CoG measurement; biomechanical assessment	Test path fully defined
	Color fidelity	ISO/CIE 11664-6	xy; Y; ΔE00	Color measurement; reference comparison	Measurement path defined; threshold open
Intra-operative surgical navigation	Navigation / tracking	SBF; K243623	Position error [mm]; angular error [°]	Performance test	Regulatory reference path
	Hygiene	MDR Annex I	Cleanability; disinfectability; contamination risk	Validated reprocessing/testing	Separate evidence required

Discussion and Conclusion

Across both cases, the XRCM results in Table 2 fall into three recurring types: a directly executable test path (cervical stress, navigation accuracy), an operationalized but normatively open criterion (color fidelity), and a pointer to a separate evidence regime the XRCM surfaces but does not resolve (hygiene under MDR Annex I) - more useful than a single pass/fail verdict. More broadly, XR deployment involves a context-dependent bundle of requirements; the XRCM integrates their selection, operationalization, traceability, and available verification pathways, while individual standards or approvals cover only part of this space. The XRCM is not a certification or conformity-assessment tool: it prepares verification, while system testing, evidence generation, and formal conformity or approval decisions remain downstream. The demonstration does not establish

completeness or full validation; applicability depends on intended use, regulatory context, and source binding force. Future work will add explicit applicability and binding-status attributes following EU MDR and aviation-certification practice, introduce reusable compliance profiles analogous to Protection Profiles in Common Criteria [ISO/IEC, 2022], and extend coverage through a structured, formative expert plausibility check.

References

Breaux, T. D., Vail, M. W. & Antón, A. I. (2006). Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations. In: 14th IEEE International Requirements Engineering Conference, pp. 46–55. https://doi.org/10.1109/RE.2006.68.

Committee for Workplaces (Ausschuss für Arbeitsstätten, ASTA). (2024). ASR A6 – Bildschirmarbeit [Display Screen Work]. *Technical Rules for Workplaces. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Dortmund*.

European Parliament and Council. (2017). Regulation (EU) 2017/745 on medical devices (MDR). *Official Journal of the European Union L* 117, Annex I, 11.1–11.2.

European Parliament and Council. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Article 57 – AI regulatory sandboxes. *Official Journal of the European Union L*, 2024/1689.

Gerhard, D. & Siewert, J. L. (2024). Industrial Augmented Reality – A Framework for Defining Requirements. International *Journal of Advances in Production Research* 1(1), 11–23. https://doi.org/10.62743/uad.3419.

Gotel, O. C. Z. & Finkelstein, A. C. W. (1994). An Analysis of the Requirements Traceability Problem. In: *Proceedings of the IEEE International Conference on Requirements Engineering*, pp. 94–101. https://doi.org/10.1109/ICRE.1994.292398.

ISO/CIE. (2022). ISO/CIE 11664-6:2022 – Colorimetry - Part 6: CIEDE2000 colour-difference formula. *Geneva: International Organization for Standardization* / CIE.

ISO/IEC. (2022). ISO/IEC 15408-1:2022 – Information security, cybersecurity and privacy protection - Evaluation criteria for IT security - Part 1: Introduction and general model. Geneva: *International Organization for Standardization*.

Keiser, D., Niermann, D., Petzoldt, C. & Freitag, M. (2025). Challenges and Requirements for Augmented Reality in Industrial Assembly Systems. *Procedia CIRP* 136, 266–271. https://doi.org/10.1016/j.procir.2025.08.047.

Pedram, S., Kennedy, G. & Sanzone, S. (2023). Toward the validation of VR-HMDs for medical education: a systematic literature review. *Virtual Reality* 27, 2255–2280. https://doi.org/10.1007/s10055-023-00802-2.

Quandt, M., Knoke, B., Gorldt, C., Freitag, M. & Thoben, K.-D. (2018). General Requirements for Industrial Augmented Reality Applications. *Procedia CIRP* 72, 1130–1135. https://doi.org/10.1016/j.procir.2018.03.061.

UL Standards & Engagement. (2023). ANSI/CAN/UL 8400:2023 – Virtual Reality, Augmented Reality, and Mixed Reality Technology Equipment.

U.S. Federal Aviation Administration (FAA). (2017). The FAA and Industry Guide to Product Certification, 3rd Edition.

U.S. Food and Drug Administration (FDA). (2024). 510(k) Premarket Notification K243623 – SpineAR SNAP (SyncAR Spine). *Surgical Theater*, Inc.

U.S. Food and Drug Administration (FDA). (2025). Product Classification – Orthopedic Augmented Reality. 21 CFR 882.4560, *Product Code SBF*.

Virtual Dimension Center (VDC). (2025). Living Lab XR-Interakt. Freigabematrix XR und Metaverse. Technologie sicher einsetzen. Fellbach. https://doi.org/10.6084/m9.figshare.32200896.

Expressive Virtual Humans for Conversational Simulation in XR: Challenges and Solutions

Alberto Altozano, Julia Domingo-Ajenjo, Maria Eleonora Minissi, Mariano Alcañiz, and Javier Marin-Morales

Institute For Research On Human-Centered Technology

Universitat Politecnica de València, Valencia, Spain

Corresponding aurhor: aaltfer@upv.edu.es

Keywords: Virtual humans, conversational simulation, nonverbal behaviour, LLMs, XR

Introduction

Virtual Humans offer a way to rehearse social situations, from job interviews and customer encounters to conversations with simulated patients. Large language models (LLMs) have expanded the possibilities for simulating human behaviour (Park et al., 2023), while XR gives the conversational partner a visible presence. Earlier systems such as SimSensei already demonstrated how virtual interviewers could coordinate dialogue and nonverbal behaviour in sustained conversations (DeVault et al., 2014).

In conversational simulation, however, a useful response is not always a helpful assistant's response. A character may need to remain hesitant, become defensive, misunderstand a question or gradually feel reassured. Its words and actions should reflect a defined personality, personal history and situation. The aim is to simulate a particular person's reactions, including their emotional and behavioural changes.

Speech and lip synchronisation alone cannot express all these reactions. Gaze, facial expression and gesture can change how an utterance is understood. This is a longstanding research problem: BEAT generated context-sensitive nonverbal behaviour from text (Cassell et al., 2001), while SAIBA and the Behavior Markup Language (BML) separated communicative intent, behaviour planning and realisation (Kopp et al., 2006). These contributions emphasise the need to select, time and combine behaviours in context. A speaking avatar with a fixed animation loop leaves much of this expressive potential unused.

In this paper, we outline requirements and practical challenges for expressive virtual humans, followed by a modular approach to addressing them. We focus on connecting character profiles and LLM-directed nonverbal behaviour to a lightweight XR avatar. The discussion offers developers a practical basis for designing conversational simulations that bring together speech, expression and behaviour.

Requirements for an expressive virtual human

Figure 1 organises six capabilities from basic embodiment to expressive behaviour. Their relative importance depends on the scenario, but they share a purpose: conveying the character's response. The profile should guide both dialogue and nonverbal expression.

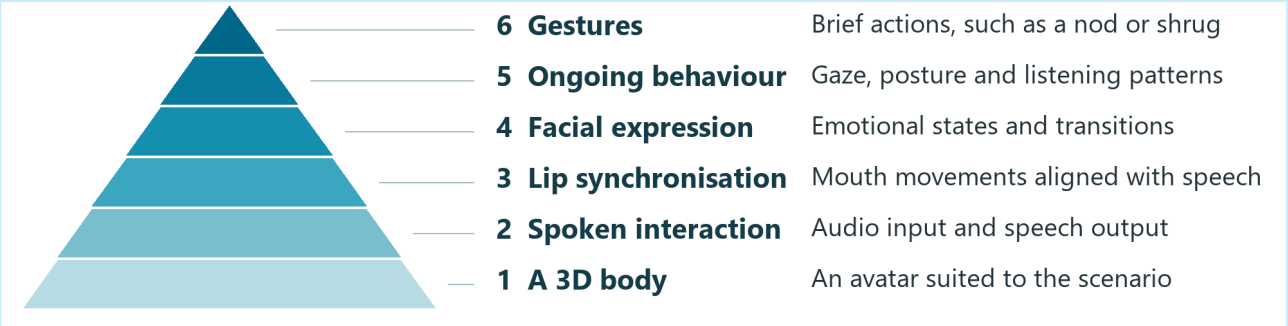


Figure 1. *Six capability layers for expressive virtual humans.*

Appearance and voice should be designed together: mismatched levels of facial and vocal realism can increase perceived eeriness (Mitchell et al., 2011). Lip synchronisation connects the voice to the body. Facial expressions convey emotion, behaviours shape gaze and posture, and gestures communicate a reaction. Their value comes from working together.

Consider an illustrative interview. The interviewer asks, “Why did you leave your last job?” The avatar looks away and replies, “I did not feel I was doing well there,” with a subdued expression and a brief shrug. Asked “Was there anything you enjoyed?”, it restores eye contact, softens its expression and answers, “Yes, helping new colleagues settle in,” with a small nod. Speech and nonverbal behaviour convey the changing reaction together.

Challenges in developing expressive virtual humans

Standardisation. A reply such as “play animation A” can trigger a gesture, but only if the prompt and application interpret that command identically. Adding or renaming behaviours means updating both. As the repertoire grows, maintenance increases.

Customisation. A new scenario requires more than a different name: the character needs a personality, a history and ways of reacting. Rewriting these instructions together with interaction rules for every character makes profiles harder to create and reuse, and ties character design to technical configuration.

Modularity. XR engines are suited to building interactive scenes, while speech and language processing often have better support in other environments. Keeping everything inside the XR application complicates integration; separating it requires a clear way to exchange user audio, generated speech and animation commands.

Platform coverage. An avatar that works on a desktop may fail in a browser or on a headset if its speech or lip-sync model needs an unsupported runtime. The challenge is to make the conversational system reusable across platforms without rebuilding it for each target.

Emotional consistency. Emotions should evolve realistically across interactions, responding to new stimuli while remaining consistent with the character’s defined personality and history.

Addressing the five challenges

Figure 2 shows the proposed solutions. The client performs the avatar's actions; the server connects its available behaviours, character instructions and speech processing.

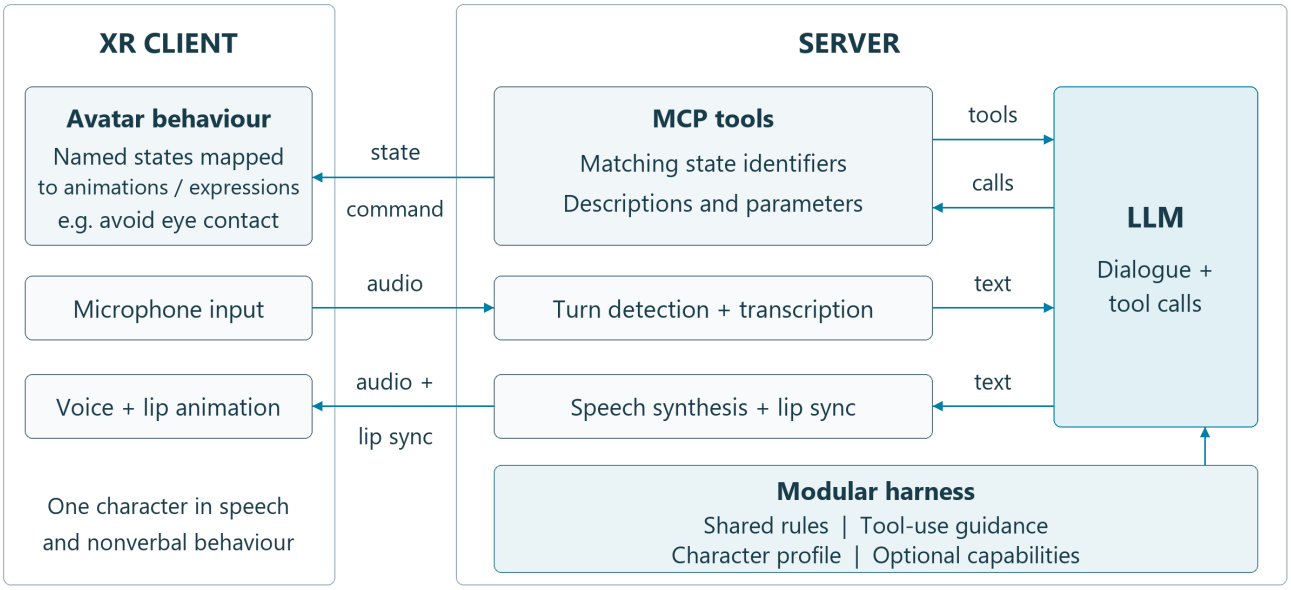


Figure 2. *Client–server architecture linking avatar behaviour, speech and character configuration.*

Standardisation through MCP tools. We propose publishing the available states through the Model Context Protocol (MCP), which describes tools and their parameters. Each published state maps to a matching client animation or expression. The server supplies the current definitions to the LLM and routes its tool calls to the client, avoiding a separately maintained command list. For example, selecting "avoid eye contact" activates the corresponding gaze behaviour. Definitions distinguish persistent behaviours, emotional states and brief gestures, so the model knows both what it can select and what each selection does.

Customisation through a modular harness. A reusable prompt template separates rules for staying in character and using MCP tools from a fillable character profile. The profile specifies personality, history and reactions, guiding both dialogue and nonverbal choices. Optional modules add instructions and tools for capabilities such as object interaction. This harness lets authors change the character or extend capabilities without rewriting instructions.

Modularity through a client–server split. The XR client acts as a virtual puppet, with animations, microphone input and audio playback. The server handles turn detection, transcription, the LLM, speech synthesis and lip-sync computation. Streaming connections, such as WebSockets, carry audio and events between them (Figure 2). The avatar and AI components can then be developed in environments suited to their respective tasks.

Platform coverage through a lightweight client. Moving inference to the server removes the need to run those models on each device. The client can be reused wherever rendering, animation, audio and networking are supported, reducing platform-specific dependencies.

Emotional state machine. User and avatar interactions provide stimuli whose emotional impact depends on personality. We propose using this impact to raise or lower state activation and drive transitions, with states representing, for example, Ekman's (1992) basic emotions.

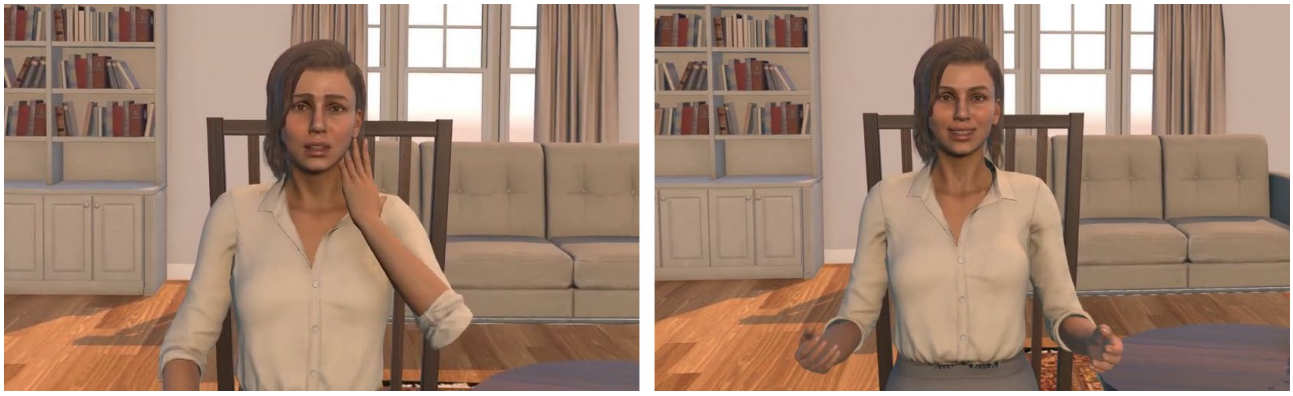


Figure 3. *The VR avatar displaying different expressions and gestures during conversation.*

Conclusions

For conversational simulation, a virtual human should express the reactions of the character it represents. Speech and lip synchronisation provide a foundation, while facial expression, ongoing behaviour and gesture allow the character to communicate more of its response. These capabilities should follow the same personality and narrative as the spoken dialogue.

The proposed approach connects these requirements to five development needs: standardised access to avatar actions, reusable character configuration, separation of AI processing from the XR client, broader platform coverage and emotional consistency. Combining MCP tools, a modular harness, a lightweight client and an emotional state machine offers a way to organise these responsibilities across conversational scenarios.

Funding

This work was supported by the VICTOR-XR project, funded by the European Union's Horizon Europe research and innovation programme (No. 101298864). It was also supported by the Spanish Ministry of Science, Innovation, and Universities through the VirtuVP project (No. PID2023-147284OB-I00), and by the Generalitat Valenciana under grant CIGE/2023/160.

References

Cassell, J., Vilhjálmsson, H.H. & Bickmore, T. (2001). BEAT: the Behavior Expression Animation Toolkit. *SIGGRAPH '01*, 477–486. https://doi.org/10.1145/383259.383315

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., ... & Morency, L. P. (2014, May). SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 1061-1068). https://www.ifaamas.org/Proceedings/aamas2014/aamas/p1061.pdf

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion* 6(3–4), 169–200.

Kopp, S., Krenn, B., Marsella, S., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R. & Vilhjálmsson, H. (2006). Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. *Intelligent Virtual Agents*, LNCS 4133, 205–217. https://doi.org/10.1007/11821830_17

Mitchell, W.J., Szerszen, K.A., Lu, A.S., Schermerhorn, P.W., Scheutz, M. & MacDorman, K.F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. i-Perception 2(1)

Park, J.S., O'Brien, J.C., Cai, C.J., Morris, M.R., Liang, P. & Bernstein, M.S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *UIST '23*, Article 2, 1–22. https://doi.org/10.1145/3586183.3606763

Cross-Platform Dual-Haptic Integration in XPBD with Real-Time Data Synchronization

Turhan Civelek[1], Kristoffer Waldow[1,2], and Arnulph Fuhrmann[1]

[1]*Institute of Media and Imaging Technology, TH Köln, Cologne, Germany*

[2]*Technical University of Munich (TUM), Human-Centered-Computing and Extended Reality Lab, TUM Klinikum, Orthopedics and Sports Orthopedics*

Corresponding author: turhan.civelek@th-koeln.de

Keywords: Haptic Interfaces, Extended Position-Based Dynamics, Extended Reality, Memory Communication

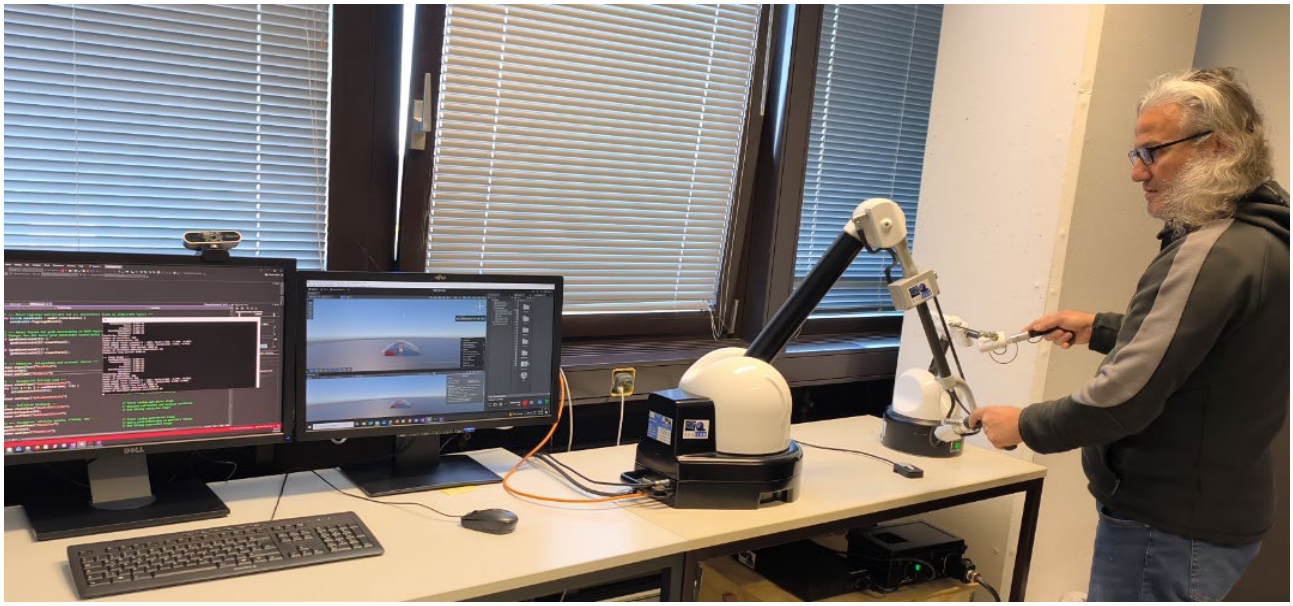


***Figure 49**. Experimental setup with two haptic devices.*

Background and System Overview

Virtual reality simulators are increasingly used for surgical training, where trainees have to acquire force-sensitive bimanual skills without putting patients at risk. Such training requires soft tissue that deforms plausibly and haptic devices that render the corresponding reaction forces in real time, ideally for both hands. Extended Position-Based Dynamics (XPBD) is widely used for real-time simulation of deformable objects because of its computational efficiency and stability (Macklin, Müller, & Chentanez, 2016). Existing work on haptic integration has focused on hardware abstraction, real-time force feedback, visual–haptic interaction, and modular system architectures, and single or dual haptic devices are usually integrated within a single development platform such as C++ or Unity (He, et al., 2026) (Zhang, et al., 2026). Synchronization across platform boundaries, multi-device interaction, and force stability remain open challenges.

Conducted within the VIRTOSHA project (Jansen, et al., 2026), this work presents a performance evaluation of a layered architecture that enables simultaneous and physically consistent interaction between two haptic devices and an XPBD-based deformable model. As illustrated in Fig. 2, architecture couples a C++ device-control and physics layer with Unity-based visualization through bidirectional shared-memory communication. The system uses Haption Desktop 6D and Virtuose 6D devices and was evaluated on a workstation with an Intel Core i7-6700, 16 GB RAM, and an NVIDIA RTX 2080 Ti GPU.

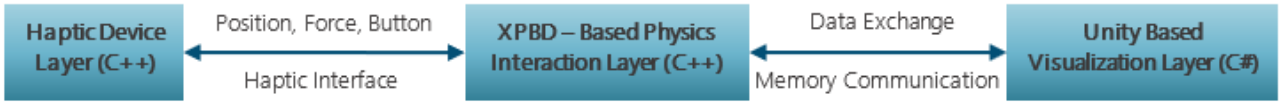


Figure .50. *System Architecture for Dual-Haptic Integration in an XPBD-Based Simulation Framework.*

The architecture is organized into three layers: Device control is handled through the Haption RaptorAPI, which exposes both the Desktop 6D and the Virtuose 6D behind a single interface, so that the two devices differing in workspace, kinematics and force range are addressed by the same control code. Each device is opened as an independent channel with its own connection, calibration and state machine, and is served by a dedicated thread at 1 kHz. Per cycle, the thread reads the current end-effector pose and button state from the RaptorAPI and writes back the force and torque computed by the simulation layer; The simulation layer represents each device by a proxy geometry that is coupled to the measured pose and takes part in collision handling against the deformable model. Reaction forces are derived from the constraint corrections of the XPBD solver at the contact positions before they are handed back to the corresponding device channel. The visualization layer is decoupled from both: Unity and the C++ process exchange data over a shared-memory segment with separate regions for mesh state and per-device channels, implemented as two unidirectional Windows memory-mapped buffers: a 1 MiB output buffer containing the three-dimensional particle positions followed by the force and position vectors of a both haptic channels, and a 56-byte input buffer containing seven floating-point interaction parameters per channel, which lets the renderer consume the most recent state at display rate without stalling the 1 kHz loops.

We therefore propose a system that defines the following main contributions: (i) integration of C++-based XPBD simulation, haptic control, and Unity visualization within a hybrid architecture; (ii) real-time synchronized data exchange; (iii) independent channel-based management of each haptic device; (iv) simultaneous, physically consistent dual-haptic interaction with a deformable model; and (v) a quantitative performance comparison of the XPBD system with and without haptic integration while bidirectional Unity–C++ communication over shared memory enables low-latency, deterministic visual–haptic synchronization.

Experimental Evaluation

Setup: We designed a small performance evaluation with the purpose of determining what the cross-platform coupling of two haptic devices costs at runtime and where in the pipeline that cost appears. Haptic rendering is the binding constraint here: the device threads must be served at 1 kHz, so every stage they depend on must complete within a 1 ms budget, while the visualization only must meet the display rate. We therefore compare the baseline simulation with the deployed dual-haptic system on the same model, workstation, and sampling procedure, so that the measured difference reflects the haptic layer together with the revisions required to sustain it.

Measures: The recorded metrics decompose the 1 ms budget along the layers of the architecture. SDtM is the cost of crossing the C++–Unity boundary and the only stage that scales with the number of devices, and HC the collision work added by the device proxies; these two carry the overhead of the integration. SCs is the dominant solver term and governs how far the model can be scaled, Tot verifies the per-frame budget, and PostS serves as a control that should remain unaffected by haptic integration. TF is the actual output of the system and indicates whether the transmitted magnitudes stay within the operating range of the devices. Times were recorded in

milliseconds and TF in newtons at one-second intervals; the data were analyzed using IBM SPSS Statistics.

Results

Figure 3 shows the deformation behavior of V1 (a−d) collision without haptic devices, (e−h) collision with haptic spheres, (i−l) grasping interaction; each row shows four consecutive time steps. Red arrows indicate the force vectors, which are transmitted to haptic devices.

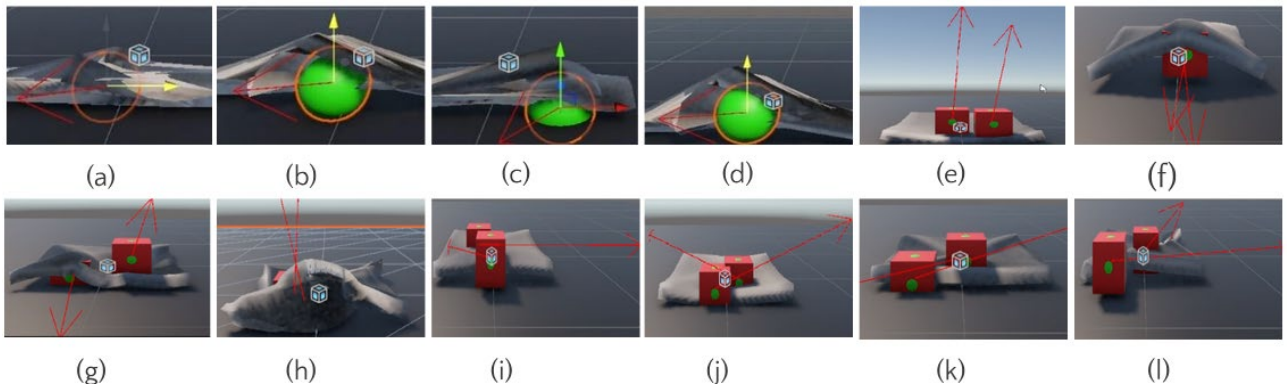


Figure 51*. Soft tissue deformation and reaction force directions.*

Table 7. *Descriptive statistics of XPBD performance metrics for V1 and V2 (N = 180 per configuration; times in ms, TF in N).*

Variable	Mean		Std. Dev.		Std. Error		Min.		Max.	
Version	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
SDtM [ms]	0.087	0.177	0.029	0.047	0.002	0.004	0.002	0.002	0.201	0.460
HC [ms]	0.009	0.014	0.005	0.007	0.000	0.001	0.004	0.006	0.027	0.051
SCs [ms]	1.350	0.704	0.232	0.127	0.017	0.009	1.005	0.518	2.543	1.134
PostS [ms]	0.003	0.003	0.002	0.003	0.000	0.000	0.002	0.002	0.023	0.400
Tot [ms]	1.372	0.733	0.235	0.134	0.017	0.010	1.024	0.536	2.567	1.202
TF [N]	2.840	1.660	2.059	1.693	0.153	0.126	0.000	0.000	9.255	7.880

Shapiro–Wilk tests rejected normality for all measures ($p < .001$), so the configurations were compared with the non-parametric Mann–Whitney U test, with effect sizes reported as $r = |z| / \sqrt{N}$, $N = 360$. As summarized in Tables 1 and 2, all parameters differ significantly ($p < .001$), the effects being large for SCs, Tot, SDtM and HC, and moderate for TF.

Table 8. *Mann–Whitney U test results and effect sizes for XPBD performance metrics.*

Variable	Mean Rank – V1/V2	U	z	P	R
SDtM [ms]	90.03	264.97	996	-15.403	< 0.001	0.812
HC [ms]	124.51	236.49	6122.5	-10.248	< 0.001	0.540
SCs [ms]	270.01	90.99	88,5	-16.378	< 0.001	0.863
Tot [ms]	269.65	91.35	153	-16.254	< 0.001	0.857
TF [N]	215.69	145.31	9865	-6.417	< 0.001	0.338

Discussion

SDtM roughly doubles, consistent with two additional position and force channels, and HC grows by a comparable factor, so that the integration stages together consume less than a fifth of the 1 ms haptic budget. The lower SCs and Tot in V2 follow from the development history described above: meeting the 1 kHz budget required restructuring the constraint-solving and data-handling paths, and these revisions more than compensate for the load of two device threads, leaving a larger share of the frame budget available for scaling the model. The smaller means and dispersion of TF likewise reflect operator-controlled contact, though the magnitudes remain within the continuous range of both devices, and PostS, as a control, is unchanged. Two limitations follow: the samples are consecutive observations of one run and therefore not independent, so the p-values should be read descriptively. Since V1 and V2 differ in both interaction source and codebase, the comparison characterizes the deployed system rather than the isolated cost of the devices

Conclusion and Future Work

This work integrated two haptic devices into an XPBD-based simulation through a hybrid C++–Unity architecture with deterministic shared-memory data flow. On a model of 900 particles and 3481 constraints, the additional cost remained confined to the communication and collision-handling stages, the total frame time stayed below the 1 ms budget of the 1 kHz device loop, and the force feedback was directionally consistent. Future work will scale the model to anatomically derived meshes, replace the fixed adhesion threshold with a tissue-specific rupture model, add a controlled ablation on an identical code base, and validate the setup with surgeons in the VIRTOSHA training scenario.

Acknowledgments. This work was partially funded by the state government of North Rhine-Westphalia and the European Union through the ERDF/JTF Program NRW 2021-2027 as part of the VIRTOSHA project.

References

He, L., Hu, M., Xiu, W., Wu, H., Zheng, S., Li, S., Dong, Q., & Hao, A. (2026). Energy-based haptic rendering for real-time surgical simulation. *Computers & Graphics*. https://doi.org/10.1016/j.cag.2025.104524

Jansen, A., Waldow, K., Pötter, S., Civelek, T., Steininger, M., Perret, J., Wellmann, M., Stein, S.-S., Lähner, D., Welle, K., Fuhrmann, A., & Krüger, B. (2026). VIRTOSHA - A VR Training Simulation for Osteosynthesis Procedures with Force Feedback and Tissue Simulation. *2026 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 921–930). Daegu, Korea: IEEE. https://doi.org/10.1109/VRW70859.2026.00171

Macklin, M., Müller, M., & Chentanez, N. (2016). XPBD: Position-based simulation of compliant constrained dynamics. *Proceedings of the 9th International Conference on Motion in Games* (pp. 49–54). Burlingame, California, USA: Association for Computing Machinery. https://doi.org/10.1145/2994258.2994272

Zhang, X., Li, S., Sun, W., Wang, Z., Zhang, H., & Jha, S. (2026). Effective simulation of nonlinear soft tissue deformation for interactive medical applications: An enhanced XPBD technique. *Biomedical Signal Processing and Control*, März. https://doi.org/10.1016/j.bspc.2025.108934

An open-source framework for prototyping virtual reality applications with force-feedback

Romane Couëdel, Thomas Howard, and Jerome Perret

Haption, France

Corresponding author: romane.couedel@haption.com

Keywords: Force-feedback, Physics Simulation, Godot

Abstract

Haptic feedback, and particularly kinesthetic feedback (also called force-feedback), is currently used in applications such as teleoperation, medical training, and virtual reality to enhance interaction, precision, and presence. However, developing force-feedback interactions requires the integration of a physics simulation, a graphical environment, and haptic devices, each having their own set of constraints and requirements. This work presents an open-source platform integrating the Godot game engine with the Bullet physics engine, providing a flexible environment for developing applications with high-performance force-feedback.

Introduction and Motivation

Haptic technologies enable users to receive physical feedback when interacting with a virtual environment. While vibrotactile feedback is natively supported by most game engines used for deploying virtual environments nowadays, force-feedback often requires complex software development, as well as very specific skills such as real-time programming, control theory and non-smooth physics simulation. The only development platforms supporting force-feedback out-of-the box today are either commercial and expensive (Thandapani et al 2027), or provide only forces (3D feedback) and no torques (6D feedback) (Conti et al., 2003); Home page of the CHAI3D library). This constraint can limit experimentation with new scenarios and make adapting them to different application requirements extremely complex.

In this context, Haption is developing a software platform to facilitate the development of applications using its force-feedback devices. The platform is based on open-source components, namely the Godot game engine (Repository of Godot Engine; Official website and download for Godot editor) and the Bullet physics simulation engine (Repository of Bullet Physics), without being strictly tied to these specific engines. Figure 1 shows a haptic device used with a demonstration built on this platform, in which the user can interact with a virtual object. This example illustrates the type of interactive haptic application that the platform is intended to support.

Beyond demonstrating the capabilities of force-feedback devices, the platform is designed as a modular and extensible framework for developing new haptic interaction scenarios. Its architecture cleanly separates the visual rendering, physics simulation, and haptic rendering processes, which operate concurrently at different update frequencies. Each component is abstracted behind a dedicated generic interface. This decouples components and allows core elements such as the

simulation or rendering engine or the force-feedback device, to be replaced without significantly affecting the rest of the system.



Figure 52. *User interacting with a demo built on the proposed force-feedback application architecture*

Architecture

The system relies on three concurrent loops, shown in Figure 2 for a simple example scene containing a manipulable cube and a floor: the visual rendering loop (Godot in our application), the physics loop (Bullet in our application), and the haptic rendering loop. Interfaces decouple these components, with each interface defining the functions and data required for communication between them.

- Visual rendering (60 Hz in our application): updates the scene representation by periodically querying the poses of physical objects through interfaces implemented by Bullet objects.
- Physics (240 Hz in our application): updates collisions, forces, gravity, and velocities. A frequency of 240 Hz was selected as a good compromise between simulation accuracy, overall performance, and CPU load. For manipulable objects, it sends their positions to the haptic interface and retrieves the forces applied by the user.
- Haptic rendering (1000 Hz in our application): this high frequency ensures stable and responsive interaction, in line with the recommendations of Adams and Hannaford (1999).

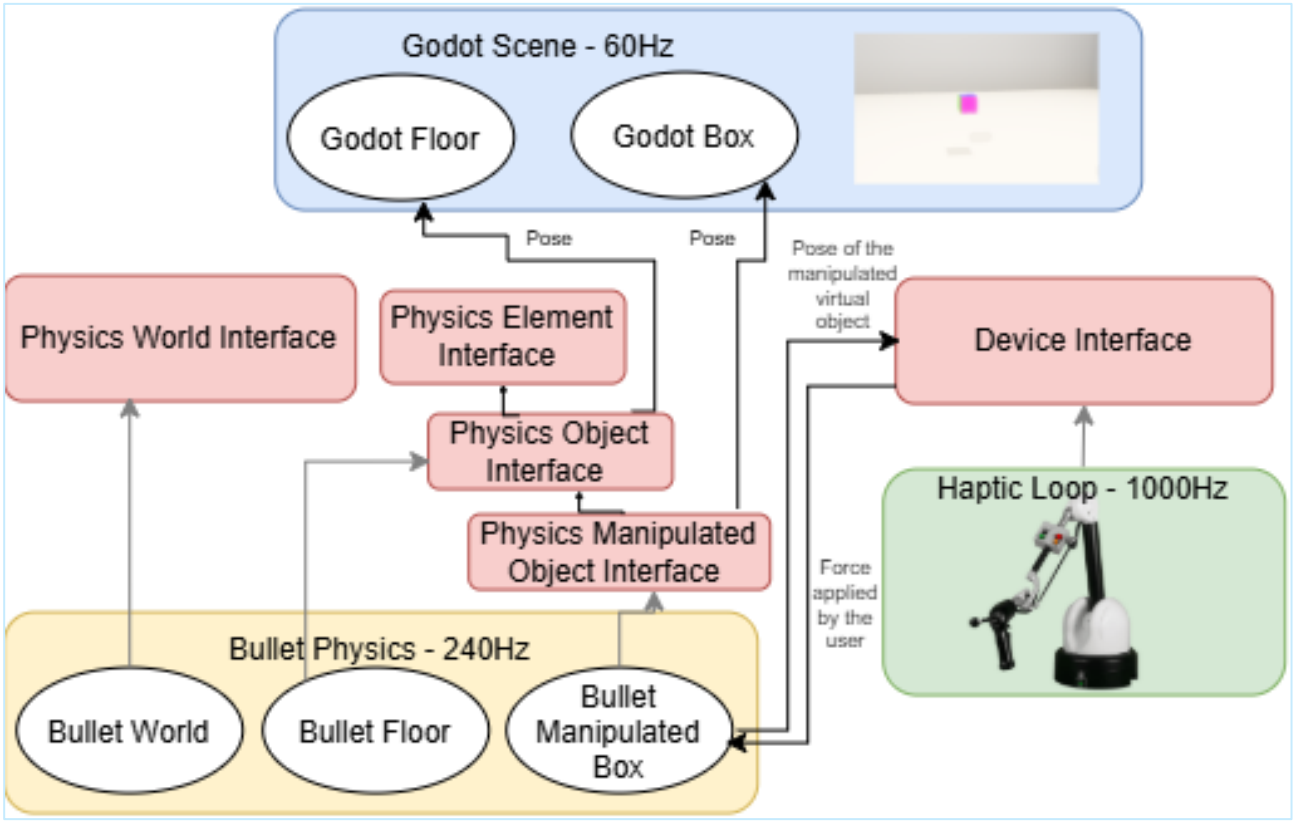


Figure 2. *Multi-rate architecture and data flow during runtime (simplified application example)*

Editor Integration

A GDExtension allows external code to be integrated directly into Godot in order to extend its functionality. The Godot editor automatically loads the elements defined in this GDExtension and exposes them as custom nodes. Figure 3 illustrates this integration in the editor: the “ManipulatedObject” node here a cube (BoxMesh), as well as the “PhysicsObject” node, representing the floor (PlaneMesh).

Any modification made in the editor, whether to the cube's size or the floor's height, is registered and passed on to Bullet when the physics world is initialized. Aesthetic elements, including lighting and the cube's color, remain configurable just as in a standard Godot project, independently of this physics synchronization.

Abstracting architectural complexity from the user’s perspective was a central consideration in the framework’s design. The scene is built like a standard Godot project, while some GDExtension nodes actually encapsulate the functionality of the three concurrent loops. This modularity also makes scene creation significantly faster, as components can be configured directly in the editor rather than rebuilt for each new scene.

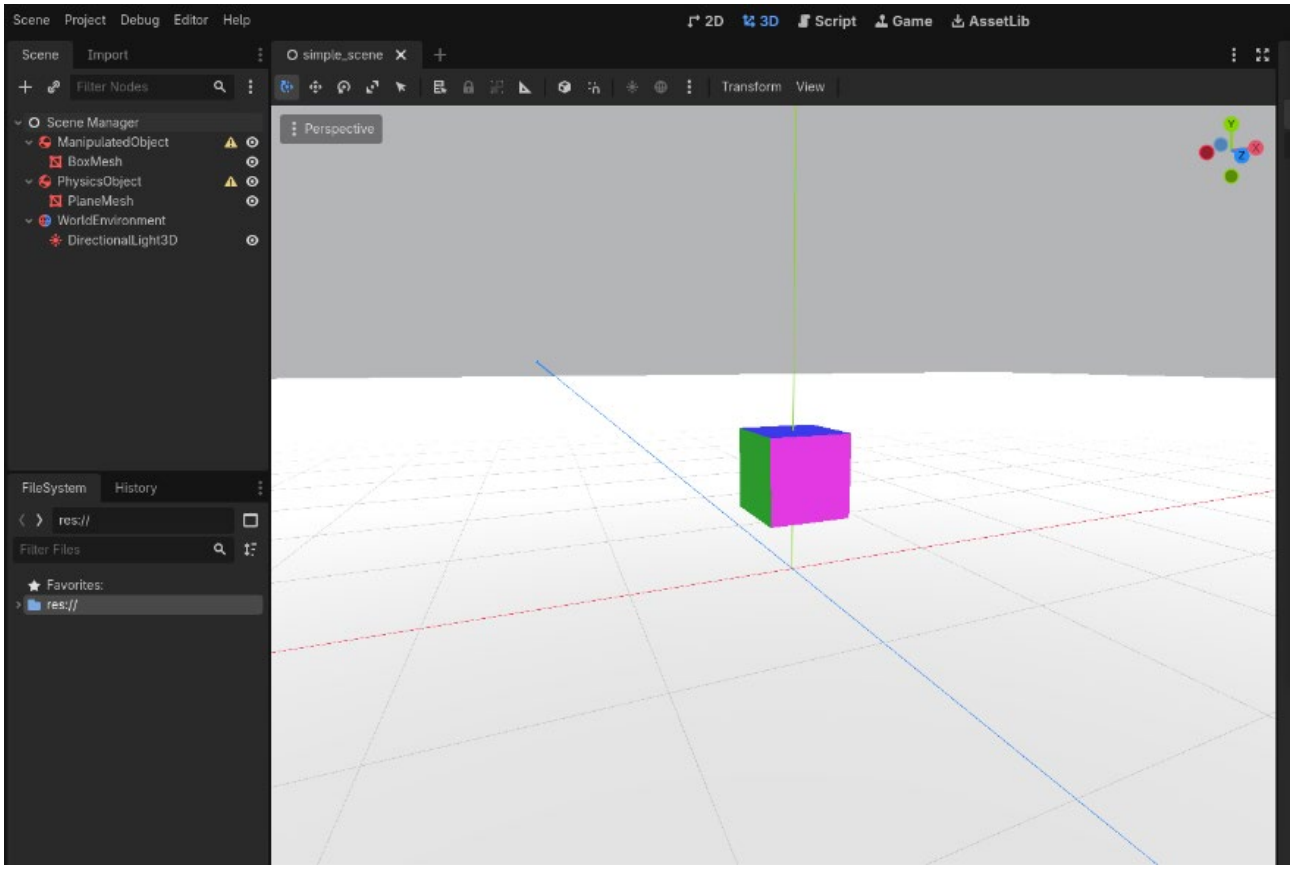


Figure 3. *Godot editor view of a simple scene with a manipulable cube and a floor*

Results

Based on the proposed architecture, two scenes for desktop haptic interaction were developed: one illustrating rigid-body interactions and the other illustrating interactions within viscous fluid. As shown in Figure 4, UI elements, visual assets, camera settings and lighting effects were easily integrated using the Godot editor, while remaining independent of the physical representation of objects.

In the first scene, several rigid objects can be manipulated and selected using a virtual hand modeled as a kinematic body. While the hand itself does not provide haptic feedback, the star, cube, sphere, and hammer generate force-feedback when interacting with the workbench. Each object is assigned a mass to simulate inertial effects during manipulation. At startup, the scene is automatically parsed to associate each visual object with a corresponding physical representation, defining the collision geometry and physical properties required for interaction. Simple objects, such as cubes and spheres, rely on native collision primitives, while more complex geometries are simplified using Volumetric Hierarchical Approximate Convex Decomposition (VHACD) at runtime, a method introduced by Mamou (et al) in 2016. This approach reduces the computational load while maintaining an adequate correspondence between visual geometry and physical interactions.

In the second scene, the spoon is the manipulated object, while the liquid inside the glass represents the viscous medium. The viscosity can be dynamically adjusted through UI controls, ranging from negligible resistance to high resistance that significantly limits the spoon's motion.

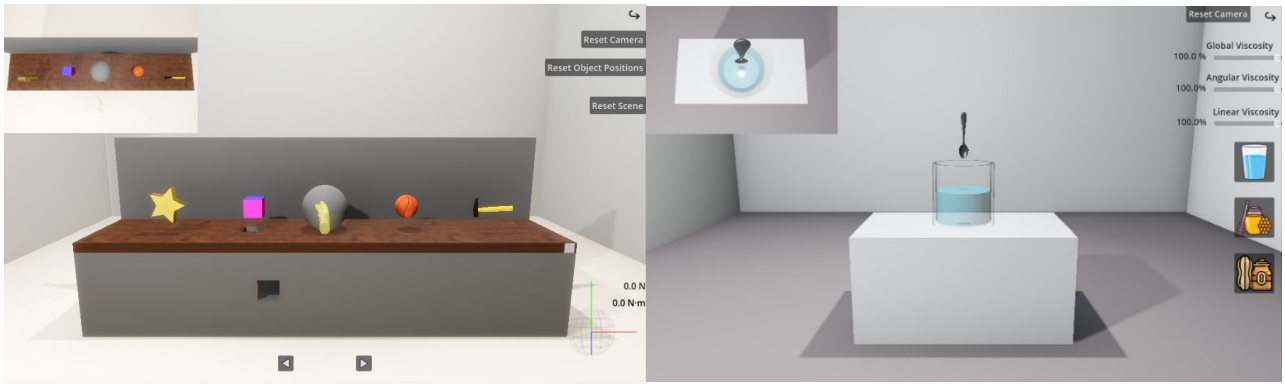


Figure 4. *(Left) Scene illustrating rigid-body contacts. (Right) Scene illustrating viscous fluid interaction.*

Conclusion and Future Work

The two scenarios presented in this work are fully operational and demonstrate the proposed architecture's ability to support different types of haptic interactions. Its modular design, based on generic interfaces, allows components to be extended or replaced without requiring major changes to the overall architecture.

This design also provides a foundation for adding new classes dedicated to other haptic functionalities and interaction models. Several scenarios are currently being investigated, including the teleoperation of a virtual robot, as well as the simulation of surface textures and deformable objects. These extensions will help evaluate the scalability of the architecture and its ability to support an increasingly diverse range of force-feedback applications.

Our development has been focused until now on a desktop experience. The framework can also be used with a HMD (thanks to Godot's built-in XR capabilities), but some development is still needed for a fully immersive experience. More specifically, we plan to add tools for calibrating the position of the force-feedback device in the scene, blending in a visual ghost of the force-feedback device, and managing frames of reference while the user moves in the virtual environment.

Together, these developments and the underlying application architecture will serve as one of the foundations for Haption's work in the Horizon Europe project Sensorama, aimed at developing the next generation of multisensory immersive XR.

References

Adams, R. J., & Hannaford, B. (1999) Stable haptic interaction with virtual environments. In *IEEE Transactions on Robotics and Automation*, vol. 15, no. 3, pp. 465-474.

Conti, F. Barbagli, F., Balaniuk, R., Halg, M., Lu, C., Morris D., Sentis, L., Vileshin, E., Warren, J., Khatib, O., & Salisbury, K. (2003) The CHAI Libraries. In *Computer Science Department, Stanford University, Stanford CA 94305, USA*

Home page of the CHAI3D library: https://www.chai3d.org/

Mamou, K., Lengyel, E., & Peters, A. (2016). Volumetric hierarchical approximate convex decomposition. *Game engine gems*, *3*, 141-158.

Official website and download for Godot editor: https://godotengine.org/download/3.x/windows/

Repository of Bullet Physics (3.27) : https://github.com/bulletphysics/bullet3

Repository of Godot Engine (4.5) : https://github.com/godotengine/godot-cpp

Thandapani, R. K. G. R., Capel, B., Lasnier, A., &. Chatzigiannakis, I. (2024). INTERACT: An authoring tool that facilitates the creation of human centric interaction with 3d objects in virtual reality. In: *MobileHCI '23 Companion: Proceedings of the 25th International Conference on Mobile Human-Computer Interaction, Article No.: 30, Pages 1 – 5,* https://doi.org/10.1145/3565066.3608250

MetaCity Liverpool: A Comprehensive, Multi-Modal Digital Twin Framework

Giuseppe Bellantuono[1], Simon Campion[1], Yang Zhang[1], Shaun Johnson[1], Wiktoria Ploch[1], Radoslaw Dorociak[1], Laurence Levers[1], Aye Chan Zay-Hta[1], Konstantin Vikhorev[1], Andrew Levers[1], and Weng Chan[1]

[1]*Virtual Engineering Centre (VEC), University of Liverpool, UK*

Corresponding author: G.Bellantuono@liverpool.ac.uk

Keywords: Digital twin, MetaCity framework, multi-modal deployment, geospatial data, real-time data, spatial AI, cloud streaming.

Introduction

Cities face increasing pressure from population growth, infrastructure demand, climate adaptation and economic change, making tools that help people understand complex urban systems increasingly valuable (Batty, 2018)(Ferré-Bigorra, Casals, & Gangolells, 2022). Most are delivered as desktop dashboards or web map clients; where immersive access exists, it is typically a separate, reduced build maintained in parallel with the primary model.

MetaCity Liverpool is a 210 km² digital twin of Liverpool, developed by the University of Liverpool's Virtual Engineering Centre (VEC) as an application of the wider MetaCity Framework, a platform for capturing, structuring and delivering city-scale digital twins. It combines high-fidelity geospatial data, real-time analytics, artificial intelligence and simulation in Unreal Engine 5, creating a photorealistic environment that is also a living data ecosystem.

City-Scale Capture and Platform Architecture

Liverpool was captured with a LEICA CityMapper 2 mounted on a light aircraft in just three days (Toschi, Remondino, Rothe, & Klimek, 2018), producing approximately 3.5 TB of processed data covering landmark buildings, transport infrastructure, streets and public spaces, and providing the spatial framework into which infrastructure, environmental and simulation data are integrated.

At the core of the framework is a real-time Unreal Engine 5 environment supported by a secure proprietary data backend and web-based frontend (Figure 1). The photogrammetry dataset is streamed using Cesium for Unreal, rendering city-scale geospatial content efficiently without sacrificing fidelity (Rantanen, Julin, Virtanen, Hyyppä, & Vaaja, 2023) (Figure 2). The environment is rendered and streamed from an NVIDIA GPU cluster, keeping the computational load off the client and delivering consistent performance regardless of access hardware (Shi & Hsu, 2015).

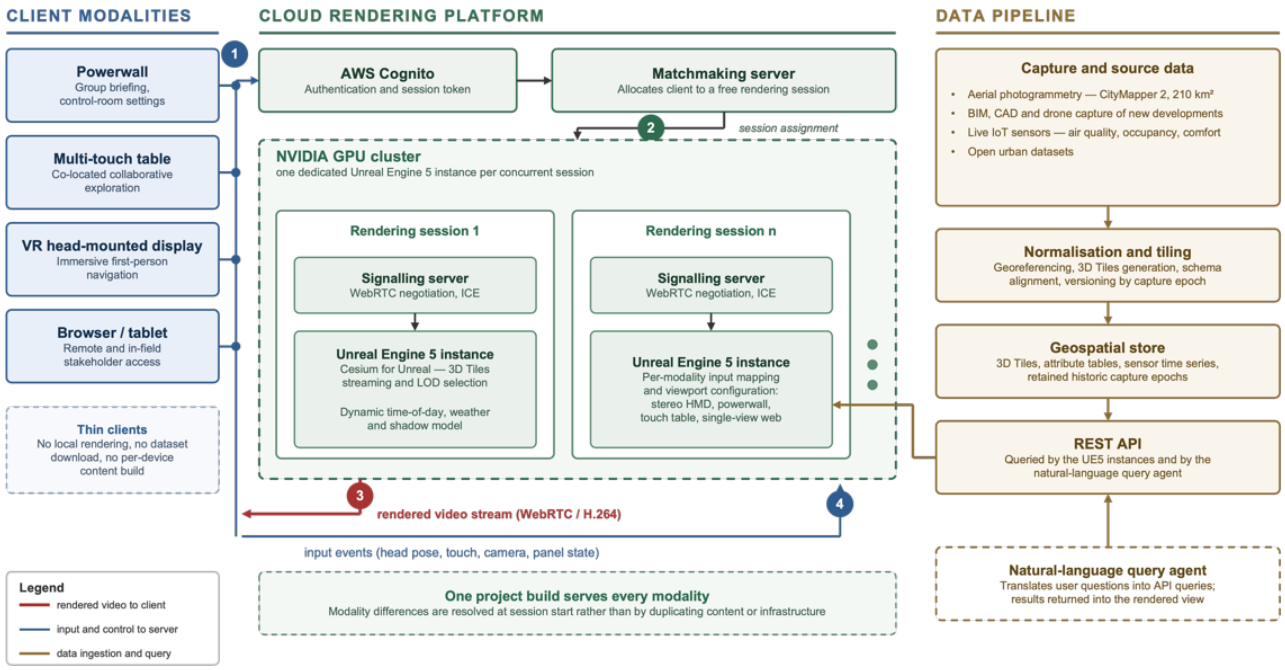


***Figure 53**. MetaCity Framework architecture. A client (1) authenticates via AWS Cognito and is allocated by the matchmaking server (2) to a session pairing a signalling server with a dedicated Unreal Engine 5 instance, which (3) streams video to the client and (4) receives its input events. A separate pipeline normalises aerial capture, BIM and live sensor data into the geospatial store, which serves 3D Tiles and REST API queries to every session.*



***Figure 2.** MetaCity Liverpool waterfront in Unreal Engine 5, with the Environment panel's live time-of-day, weather and fog controls.*

Multi-Modal Deployment

Because the environment is streamed rather than run locally, the same MetaCity instance can be deployed across multiple modalities without duplicating content or infrastructure: large-screen powerwalls for briefings and control rooms, multi-touch tables for collaborative exploration, VR headsets for immersive first-person navigation (Dembski, Wössner, Letzgus, Ruddat, & Yamu, 2020), and browser access from devices such as an iPad for workshops, public engagement or field use. The current cluster sustains up to 24 concurrent sessions.

Immersive Exploration and Dynamic Environments

Using game-engine technology as its visualisation layer, the framework offers a more immersive and intuitive experience than analysis-focused digital twins. Unreal Engine 5 provides dynamic weather, atmospheric effects and interactive day-night cycles, with realistic sun and shadow simulation for assessing the visual impact of buildings at different times of day and year (Schrotter & Hürzeler, 2020). Users can move seamlessly from a city-wide aerial view to ground-level exploration of streets and building entrances (Figures 3 and 4), supporting both strategic overview and human-scale assessment.

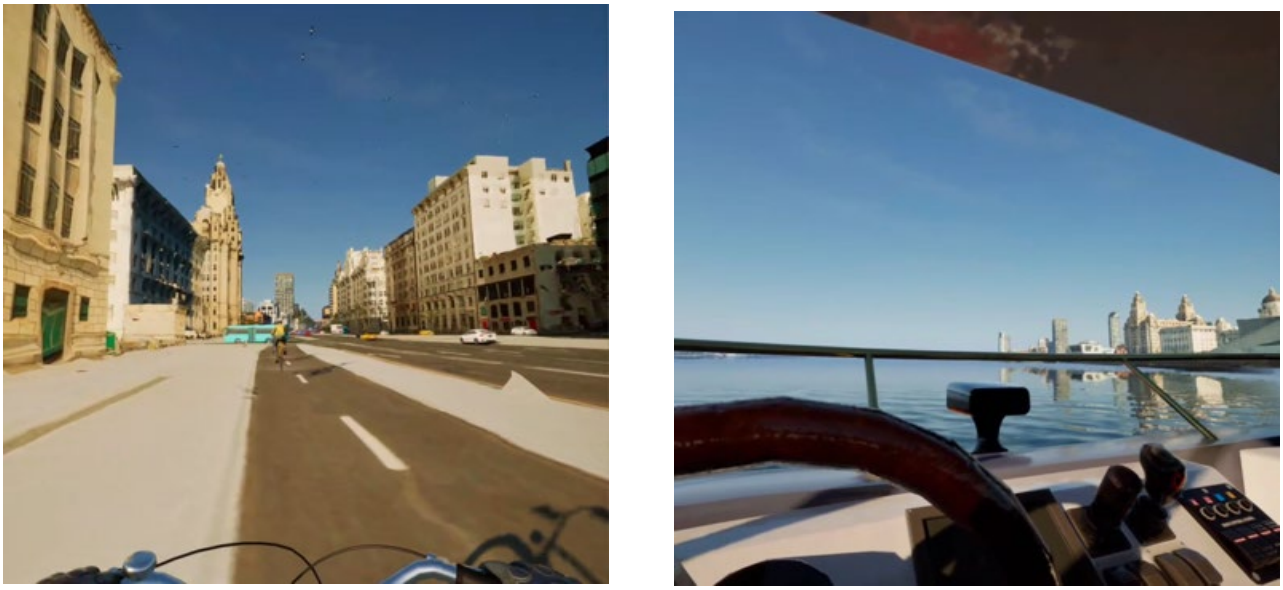


Figure 3. *First-person VR views: cycling past the Royal Liver Building (left) and viewing the skyline from a vessel on the Mersey (right).*



Figure 4. *Ground-level, third-person exploration at street scale.*

Real-Time Data and a Living Digital Twin

An AI-driven chatbot lets users query datasets in natural language, giving planners, researchers and stakeholders multi-dataset insights without specialist GIS expertise (Kanna & Kolbe, 2025). MetaCity Liverpool also ingests real-time sensor data: outdoor sensors stream live air quality readings, while indoor building sensors report occupancy, space utilisation and environmental

comfort, producing a continuously evolving representation for real-time monitoring and decision-making.

New developments can be integrated directly from CAD, BIM and drone-captured data so proposed changes are seen in their city context, and interior models enable seamless transitions from metropolitan scale to indoor navigation. Each new round of capture is added alongside earlier data rather than replacing it, building a historic record that allows past and present states of the city to be compared.

Integrated Datasets and Urban Simulation

The platform integrates neighbourhood population, crime statistics, infrastructure networks, air quality, building usage and large-scale vehicle and pedestrian simulations, explored through heat maps, Sankey diagrams, charts and dashboards in the same spatial context. Simulations informed by real-world mobility and footfall data let planners test the effects of new developments, road layouts, transport interventions and major events before implementation.

Applications and Future Direction

Beyond visualisation, MetaCity Liverpool explores how digital twins can act as long-term urban intelligence systems supporting master planning, infrastructure development, event preparedness, environmental monitoring and governance, with stakeholders exploring scenarios collaboratively in a shared immersive space. Spatial AI and analytics identify patterns and emerging challenges, while preserved historical data let users revisit past decisions in their original context alongside proposed developments.

Conclusion

By combining high-fidelity capture, cloud-streamed rendering, multi-modal deployment, AI, simulation and live data streams within a single framework, MetaCity Liverpool shows how city-scale digital twins can extend beyond visualisation into urban planning, digital infrastructure and spatial computing, and how technologies built for virtual worlds can help cities test future scenarios and engage stakeholders across devices and settings.

References

Batty, M. (2018). Digital twins. *Environment and Planning B: Urban Analytics and City Science, 45*(5), 817-820. https://doi.org/10.1177/239980831879641

Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., & Yamu, C. (2020). Urban digital twins for smart cities and citizens: the case study of Herrenberg, Germany. *Sustainability, 12*(6), 2307.

Ferré-Bigorra, J., Casals, M., & Gangolells, M. (2022). The adoption of urban digital twins. *Cities, 131*, 103905.

Kanna, K., & Kolbe, T. H. (2025). Advanced user interaction with urban digital twins using large language models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-G-2025*, 469-476. https://doi.org/10.5194/isprs-annals-X-G-2025-469-2025

Rantanen, T., Julin, A., Virtanen, J.-P., Hyyppä, H., & Vaaja, M. T. (2023). Open geospatial data integration in game engine for urban digital twin applications. *ISPRS International Journal of Geo-Information, 12*(8), 310. https://doi.org/10.3390/ijgi12080310

Schrotter, G., & Hürzeler, C. (2020). The digital twin of the city of Zurich for urban planning. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88(1)*, 99-112.

Shi, S., & Hsu, C.-H. (2015). A survey of interactive remote rendering systems. *ACM Computing Surveys, 47*(4), 1-29.

Toschi, I., Remondino, F., Rothe, R., & Klimek, K. (2018). Combining airborne oblique camera and LiDAR sensors: investigation and new perspectives. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1*, 437-444. https://doi.org/10.5194/isprs-archives-XLII-1-437-2018

Integration of Artificial Intelligence into Virtual Reality Training Systems for Public Speaking

Sarah Saufnay and Michaël Schyns

Liège – Management School of the University of Liège, LabXR, Belgium

Corresponding author: sarah.saufnay@uliege.be

Keywords: Public speaking, virtual reality, artificial intelligence, training

Introduction

Although speaking in public is common, many people struggle with it due to limited practice opportunities, which can cause significant anxiety (Kroczek & Mühlberger, 2023). Virtual Reality (VR) training helps address this by simulating social interactions, which are costly and difficult to reproduce in real life (Hirzle et al., 2023). Instead, once wearing a VR headset, users are immersed in a realistic presentation setting and face an audience of Intelligent Virtual Agents (IVAs), i.e., computer-controlled virtual humans that display social cues to mimic human presence (Kyrlitsias & Michael-Grigoriou, 2022). Moreover, Artificial Intelligence (AI) has expanded VR's interaction capabilities across various training fields (Hirzle et al., 2023). By analysing user behaviour, AI indeed enables the development of responsive Virtual Environments (VEs) and the generation of associated content. These advances have been especially relevant for Public Speaking Training (PST) environments, where AI can help assess speakers' performance, provide feedback, and customise training sessions (Chang et al., 2026).

Carefully designing the virtual audience's behaviour in response to the speaker's performance is essential. Both IVAs' verbal and nonverbal cues significantly influence audience perception (Glémarec et al., 2021), which, in turn, affects speakers' real-life confidence and performance (Kroczek & Mühlberger, 2023). In addition, incorporating AI into VR is essential to shift from static to dynamic simulations, but it presents challenges, including limited computational power, latency, data privacy, and potential dependence on external services.

In this context, a PST application in VR has been developed, allowing users to choose specific training configurations and interaction modalities based on their objectives. The system provides diverse presentation contexts for subsequent use in school programs or companies, and contributes at the technical level by integrating AI into standalone VR applications. Currently, this application is employed at the authors' University across multiple faculties to train students in public speaking and prepare them for future careers.

Platform Framework

The platform features a web-based companion interface alongside a standalone VR application. The web interface supports the preparation and management of training sessions, while the VR application provides the immersive training experience. Before entering VR, users can upload presentation materials such as slides and notes, which are then available in the VE. After each

training session, users can access performance feedback via the web interface to support reflection and skills improvement.

Training sessions can be customised by selecting a room, a role, and a scenario defining the audience's attitudes and training events. Once immersed, users can interact with their slides and virtual notes during practice, thereby reproducing realistic presentation settings. Sessions can also be recorded and replayed from a third-person perspective, either from the website or within the VE, allowing users to reflect more objectively on their performance (Zhou et al., 2021).

Training configurations

Several rooms, each corresponding to a public speaking context and associated with specific roles and scenarios, have been developed to cover a broad range of communication situations (Figure 1).

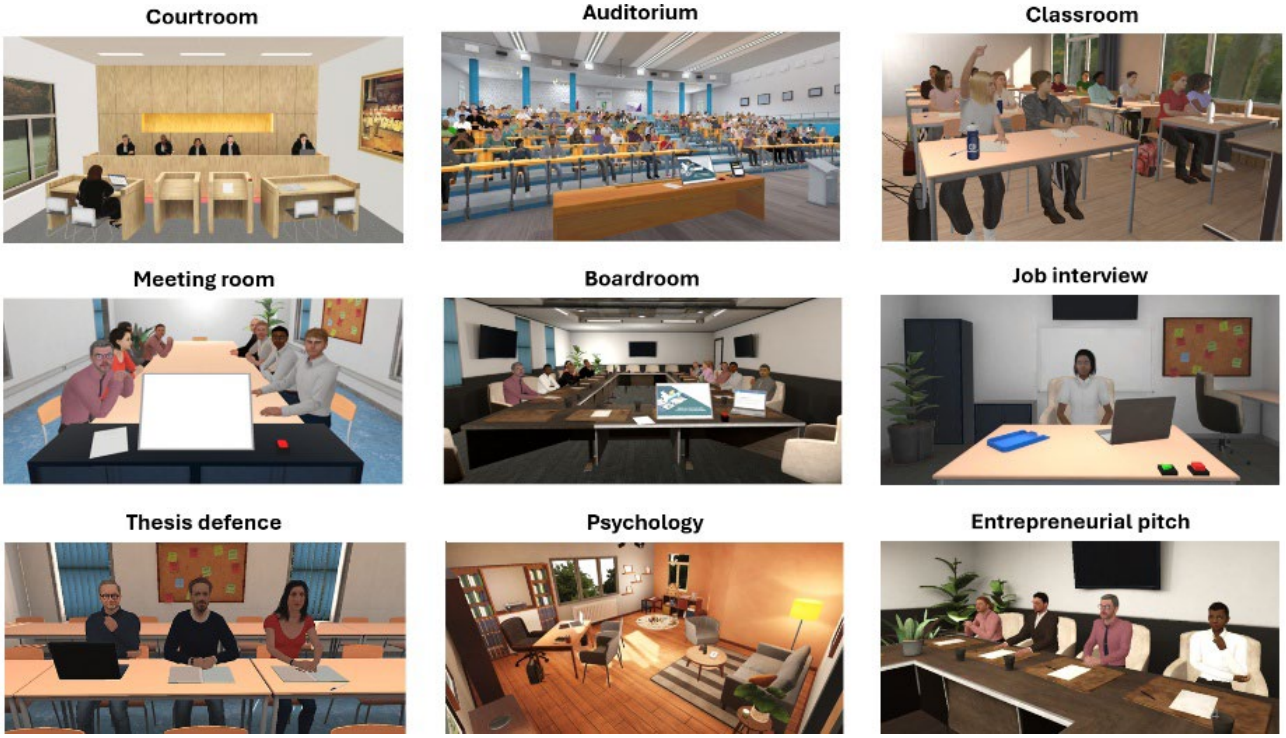


***Figure 54**. Overview of the available rooms*

The **Auditorium**, which replicates a university auditorium, represents a large-scale public speaking context, whereas the **Boardroom** and **Meeting Room** provide formal settings for business presentations. These environments feature configurable and validated virtual audiences, whose attitudes can range from attentive and supportive to disengaged or negative (Etienne et al., 2023; Saufnay & Schyns, 2025). More specialised environments target specific communication contexts. The **Courtroom** familiarises future lawyers with courtroom procedures and speaking constraints, while the **Thesis Defence** reproduces an academic jury using photorealistic avatars of actual professors, enabling the study of familiarity effects on stress, social presence, and performance. The **Pitch** environment simulates entrepreneurial presentations. In the latter two VEs, jury members can ask either prerecorded generic questions or specific ones, generated at runtime using the embedded AI system. Other environments can help prepare users for professional situations before internships. The **Classroom** focuses on teacher training and classroom management through pedagogical and disciplinary events occurring during immersion, such as students talking, using their phones, sleeping, or rocking their chairs. Speech recognition and intention detection enable the system to identify appropriate interventions from the users' speech and dynamically adapt the virtual students' behaviour. The **Psychology Internship** room exposes future psychologists to challenging

interactions, including patients questioning their credibility because of their age, remaining silent, or displaying strong emotional reactions. Finally, the **Office** supports job interview training from both the applicant and interviewer perspectives and can be used to investigate recruitment biases by using IVAs of different genders and ethnicities. Interactions may rely on predefined audio recordings or on conversational IVAs for more diversified discussions.

Intelligent and Adaptive Virtual Audiences

A key component of the platform is the relationship between IVAs' behaviours and users' perception of audience attitudes. These attitudes are modelled along valence and arousal, representing the positive or negative attitude towards the speaker and activation level of the audience response, respectively (Etienne et al., 2023). Attitudes are designed as combinations of nonverbal cues, including posture, gaze, head movements, and facial expressions. The perception of these multimodal behaviours was first studied and validated at the individual IVA level (Etienne et al., 2023) and subsequently investigated at the group level (Saufnay & Schyns, 2025) to determine how individual cues translate into an overall audience attitude. Indeed, group perception cannot be reduced to the perception of each member in isolation, as factors such as the proportion of agents displaying specific behaviours, their recurrence, or the IVAs' gender and position within the room may influence the perceived attitude.

Beyond scripted events, the application's long-term objective is to make audience reactions directly responsive to speakers' performance. Using openSMILE (Eyben et al., 2016), the system extracts acoustic features from the speaker's voice to estimate performance- and emotion-related dimensions, including valence, arousal, stress, and vocal dynamics. Speech alone provides valuable information for evaluating performance, yet it can be further enriched by tracking the user's nonverbal behaviour. For example, a monotonous delivery could reduce engagement, whereas a more dynamic presentation could elicit attentive or positive responses. This creates an interaction loop in which the speaker's behaviour is analysed and translated into changes in the VE.

Artificial Intelligence and Interactions

The audience may also ask questions during the presentation, requiring AI components for contextual interactions. A major technical objective achieved with this application is integrating AI locally on standalone VR headsets to preserve privacy, minimise response latency, and avoid dependence on external services (Wang et al., 2025). Lightweight models were therefore preferred.

Automatic Speech Recognition, such as Vosk, converts the user's speech into text. Depending on scenario complexity, the resulting text can be processed by either a dedicated intention-detection model or a Large Language Model (LLM). Intention detection is particularly useful when a restricted set of intents must be recognised. In the Classroom, for example, the system determines whether the user is asking students to stop a specific behaviour, triggering the relevant IVA response. More flexible scenarios, including job interviews or thesis defences, can instead rely on small LLMs, such as a 3B-parameter Llama model deployed through the UndreamAI framework, to generate context-specific questions based on the interaction and scenario. Generated responses can then be converted into speech through a Text-To-Speech component (e.g., Piper). Response latency remains an important constraint, as excessive delays reduce interaction naturalness. Whenever computational resources allow, components are executed locally on the VR headset.

Research Perspective

The platform is intended as both a training application and a research platform. Research objectives span the development pipeline, from socio-affective perception of multimodal IVAs' behaviour to automatic assessment of speaking performance and training effectiveness. The project also enables comparisons between predefined and AI-generated interactions. Prerecorded content offers greater experimental control, whereas generative AI may increase relevance and variability. Comparing these approaches can help determine the best design to achieve ecological validity and

a strong sense of presence and is considered future work. Overall, the platform aims to move from a static to an adaptive PST system. By combining configurable training environments, intelligent virtual audiences, performance metrics, recording and replay functionalities, the system is designed not only to reproduce public-speaking contexts but also to support skills transfer to real-world public-speaking situations.

References

Chang, E., Che, X., & Zhang, J. (2026). Designing and evaluating LLM-driven coaching agents for VR public speaking practice. Frontiers in Psychology, 17, 1886780. https://doi.org/10.3389/fpsyg.2026.1886780

Etienne, E., Leclercq, A.-L., Remacle, A., Dessart, L., & Schyns, M. (2023). Perception of avatars' nonverbal behaviors in virtual reality. Psychology & Marketing, 40(11), 2464–2481. https://doi.org/10.1002/mar.21871

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., & Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing, 7(2), 190–202. https://doi.org/10.1109/TAFFC.2015.2457417

Glémarec, Y., Lugrin, J.-L., Bosser, A.-G., Collins Jackson, A., Buche, C., & Latoschik, M. E. (2021). Indifferent or enthusiastic? Virtual audiences animation and perception in virtual reality. Frontiers in Virtual Reality, 2. https://doi.org/10.3389/frvir.2021.666232

Hirzle, T., Müller, F., Draxler, F., Schmitz, M., Knierim, P., & Hornbæk, K. (2023). When XR and AI meet: A scoping review on extended reality and artificial intelligence. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery. https://doi.org/10.1145/3544548.3581072

Kroczek, L. O. H., & Mühlberger, A. (2023). Public speaking training in front of a supportive audience in virtual reality improves performance in real-life. Scientific Reports, 13(1), 13968. https://doi.org/10.1038/s41598-023-41155-9

Kyrlitsias, C., & Michael-Grigoriou, D. (2022). Social interaction with agents and avatars in immersive virtual environments: A survey. Frontiers in Virtual Reality, 2. https://doi.org/10.3389/frvir.2021.786665

Saufnay, S., & Schyns, M. (2025). Perception of virtual audiences: Influence of gender and nonverbal behavior. In Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents (IVA '25) (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3717511.3747085

Wang, Y., Guizani, M., & Hossain, M. S. (2025). Artificial intelligence for virtual reality: State of the art, challenges, and future perspectives. ACM Transactions on Multimedia Computing, Communications, and Applications, 21(11). https://doi.org/10.1145/3769090

Zhou, H., Fujimoto, Y., Kanbara, M., & Kato, H. (2021). Virtual reality as a reflection technique for public speaking training. Applied Sciences, 11(9), 3988. https://doi.org/10.3390/app11093988

XREL: a Hybrid eXtended REality Learning experience

Erica Rocchi, Anna Ferrarotti, Matteo Pestelli, Manuela Viscontini, Marcos Valdes, Matias Guerra, Sara Baldoni, Federica Battisti, and Marco Carli

[1]*Dept. of Industrial, Electronic and Mechanical Engineering, Roma Tre University, Rome, Italy*

[2]*National Inter-University Consortium of Telecommunications (CNIT), Parma, Italy*

[3] *Virtual Immersions in Science, Pisa, Italy*

[4] *Dept. of Information Engineering, University of Padova, Padova, Italy*

Corresponding author: erica.rocchi@uniroma3.it

Keywords: eXtended Reality; immersive learning; multisensory media; 360° video; haptics; olfactory display; Quality of Experience

Introduction

Multi-sensory immersive technologies are increasingly integrated into educational settings. However, the extent to which each component of the experience (i.e., immersion, and non-visual sensory modalities), contributes to user satisfaction and learning remains insufficiently characterized. While haptic interaction has been repeatedly associated with improved conceptual learning outcomes in embodied XR environments (Gu, et al., 2025), non-visual modalities such as olfactory and thermal stimulation remain comparatively understudied (Świdrak, et al., 2024). Furthermore, existing evidence largely stems from studies isolating a single modality in brief, highly controlled tasks, leaving unaddressed what a complete, multi-scene lesson achieves when evaluated against standard institutional delivery formats. To bridge this gap, XREL (eXtended REality Learning), the educational pilot of the Horizon Europe project Hybrid eXtended reAliTy (HEAT), transforms a conventional lecture into a hybrid multisensory experience (Murroni, Muntean, Montagud, & Popescu, 2024), in which learners explore otherwise inaccessible virtual spaces and photorealistic reconstructions of real environments, interact with a holoported teacher, and receive haptic, olfactory, and thermal stimuli synchronized with the narrative. The system serves two purposes: demonstrating a complete hybrid XR learning pipeline and assessing, through a controlled three-group comparison, how multisensory immersion affects Quality of Experience (QoE), presence, engagement, cognitive load, usability, and knowledge retention with respect to two control conditions constituted by a traditional lecture and a 360° video presenting the same content.

The XREL experience

XREL takes learners from a classroom to three destinations in the Solar System: the Sun, Mercury and Mars. The hybrid XR module and the 360° video are both delivered through an HMD and last 30 and 7 minutes respectively, while the frontal lesson uses no HMD and lasts approximately 30 minutes. The 360° video and the frontal lesson consist of the three planetary segments alone; the hybrid XR module adds a familiarization phase and a classroom scene, and is therefore organized into five parts:

1. **Familiarization:** A short onboarding scene allows participants to adjust to the HMD and to the WeArt TouchDIVER Pro gloves[3], which has been selected to render the haptic feelings. The familiarization with the glove consists in grasping and releasing a simple object until the correspondence between real and virtual hand feels natural. The purpose of this phase is to avoid any influence of the novelty of the hardware on the following tests. The instructor, physically located at a remote site, and the student are captured volumetrically and holoported into the scene, so that they share a single "virtual" space and address each other in real time.

2. **The classroom:** The lesson starts in a real teaching room, captured and reconstructed as a navigable 3D scene rendered with 3D Gaussian Splatting (Kerbl, Kopanas, Leimkuehler, & Drettakis, 2023), with 6DoF. Beginning in a realistic learning setting is deliberate: it gives learners a known spatial reference before being immersed in the learning scenes.

3. **The Sun:** The first learning segment consists in providing information about the Sun. Learners navigate in the outer space towards the Sun, and, as they approach the star, an infrared lamp, driven by a Shelly Gen3 dimmer[4], raises radiant heat synchronized with the visual content, so that distance becomes something perceived through physical sensation rather than solely through a visual representation.

4. **Mercury**: The second learning segment focuses on the peculiarities of the planet Mercury and is divided into four phases (Pestelli, Rocchi, & Carli, 2026). The planet's day–night thermal dichotomy, roughly 430 °C against −180 °C, is first made tangible through paired hot and cold zones rendered by the gloves' thermal actuators. Learners then take the planet in hand and explore its surface texture, locate Caloris Planitia on a rotating model, and finally run their fingers over a large-scale reconstruction of the same basin, where ridges and crater cavities are rendered through force feedback.

5. **Mars**: Two 360° captures place learners on the Martian surface while the INOv3 diffuser[5] releases sulphur and iron-oxide cues. Olfactory stimuli allow the direct perception of the chemistry of the planet's regolith, information that standard lectures convey as an abstract line in a table of composition.

Olfactory and thermal stimulation are triggered by the instructor over a WebRTC control channel, which also governs the order of the segments and when each begins. Every

[3] WeArt. (2026). TouchDIVER Pro gloves. Retrieved from WeArt: weart.it/touchdiver-pro-haptic-gloves/
[4] Shelly. (2026). Shelly Dimmer Gen3. Retrieved from Shelly: shelly.com/products/shelly-dimmer-gen3?_pos=2&_sid=ff0fa6f11&_ss=r
[5] Inovasens. (2026). Inovasens. Retrieved from Inovasens: inovasens.fr/

stimulus is thus delivered at a known point of the narrative and can be attributed to a specific segment during the analysis.

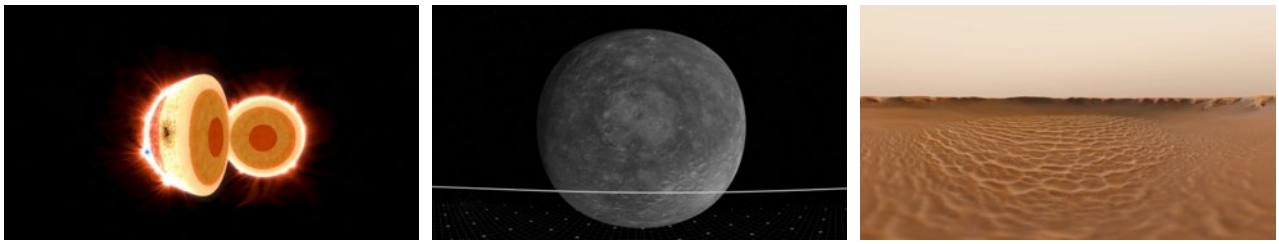


***Figure 55**. Stills from the 360° video. i) The Sun; ii) Mercury and iii) Mars.*

Study design. XREL is evaluated through a between-subjects study with 21 participants per group, hereafter Group A for the hybrid XR experience, Group B for the 360° video and Group C for the frontal lecture. Volunteers were recruited from the general public and were assigned to the three conditions according to availability. The resulting groups are comparable in age (24.3 ± 2.2, 25.1 ± 3.3 and 23.3 ± 3.4 years for Groups A, B and C) and in prior familiarity with VR (2.67 ± 1.02, 2.90 ± 1.09 and 2.24 ± 1.30 on a five-point scale), with no significant differences (Kruskal-Wallis, $p = 0.073$ and $p = 0.137$). The 360° and frontal versions cover the same three planetary segments of the hybrid XR, with similar narration and identical learning objectives. Group B receives the lesson as a stereoscopic 360° video played back on the same HMD, reusing the visual content of the full experience (Group A) but with three degrees of freedom and no holoportation, interaction or multisensory stimulation; it therefore isolates immersion-as-viewpoint from the multisensory and social layers, while representing a realistic low-cost deployment option for schools and museums. Group C receives the same content from a teacher in a physical classroom, supported by two tangible props: a 3D model of Mercury and a manually operated scent diffuser. Table 1 summarizes the three conditions. All groups follow the same protocol: a pre-experience questionnaire collecting demographics, prior knowledge of the topic and familiarity with XR; stimulus-specific questionnaires administered during the lesson; an immediate post-experience session measuring QoE (ITU-T, 2022), multisensory experience (Covaci, et al., 2019), presence (Orduna, Pérez, Gutiérrez, & García, 2022), cognitive load (Hart & Staveland, 1988), usability (Brooke, 1996) and cybersickness (Pérez, Oyaga, Ruiz, & Villegas, 2018) through standardized questionnaires; and a knowledge-retention test repeated one month later. The nominal durations differ, 30 minutes for Groups A and C against 7 for Group B, because the hybrid XR and classroom conditions include interaction, questions and exploration time that a linear 360° video cannot accommodate. The expository content, and the material covered by the knowledge test, are identical across conditions. Data collection has been completed, and the analysis is in progress.

***Table 9**. Comparison of the three modalities.*

Dimension	Group A - Hybrid XR	Group B - 360° video	Group C - Traditional
Visual scene	3DGS reconstruction + Virtual scenes, 6DoF	Stereoscopic 360° capture, 3DoF	Real classroom
Instructor	Holoported volumetric presence	Pre-recorded voice-over	Real presence
Interaction	Object manipulation, dialogue with the instructor	Passive viewing, gaze-based navigation only	Passive lesson with two physical sensory props
Tactile	WeArt TouchDIVER Pro gloves	None	3D model of Mercury
Olfaction	INOv3 scent diffuser	None	Manual scent diffuser
Thermal cue	Infrared lamp (Shelly Gen3 dimmer)	None	None

Conclusions

XREL brings photorealistic environment capture, holoportation and synchronized haptic, olfactory and thermal stimulation into a single educational experience, deployed in three delivery modes that an educational institution could plausibly adopt. The study described here is designed to establish whether the cost and complexity of a full multisensory setup are repaid in engagement and in knowledge retention. Future work will report the results of the three-group comparison, including the one-month retention data.

Acknowledgements

This work has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101135637 (HEAT – Hybrid Extended reAliTy).

References

Brooke, J. (1996). SUS: A quick and dirty usability scale.

Covaci, A., Trestian, R., Saleme, E. B., Comsa, I.-S. a., Santos, C. A., & Ghinea, G. (2019). 360° Mulsemedia: A Way to Improve Subjective QoE in 360° Videos. *Proceedings of the 27th ACM International Conference on Multimedia*.

Gu, P., Li, Y., Ji, H., Zhang, Y., Zhang, Y., & Luo, Y. (2025). The Role of Haptic Interaction in Embodied Extended-Reality Learning: A Three-Level Meta-Analysis. *Educational Psychology Review*. https://doi.org/10.1145/3343031.3350954

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in psychology*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

ITU-T. (2022). ITU-T P.1320: Quality of experience assessment of extended reality meetings. *ITU-T*.

Kerbl, B., Kopanas, G., Leimkuehler, T., & Drettakis, G. (2023). 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*.

Murroni, M., Muntean, G., Montagud, M., & Popescu, V. (2024). Toward Interactive and Distributed Hybrid Extended Reality Experiences. *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences Workshops (IMXw '24)*. Association for Computing Machinery. https://doi.org/10.1145/3672406.3672424

Orduna, M., Pérez, P., Gutiérrez, J., & García, N. (2022). Methodology to assess quality, presence, empathy, attitude, and attention in 360-degree videos for immersive communications. I*EEE Transactions on Affective Computing*, 2375--2388.

Pérez, P., Oyaga, N., Ruiz, J. J., & Villegas, A. (2018). Towards Systematic Analysis of Cybersickness in High Motion Omnidirectional Video. *10th International Conference on Quality of Multimedia Experience (QoMEX)*, (S. pp. 1-3). Cagliari, Italy. https://doi.org/ 10.1109/QoMEX.2018.8463377

Pestelli, M., Rocchi, E., & Carli, M. (2026). The Impact of Haptics on User Presence in Virtual Reality. *Proceedings of the 2026 ACM International Conference on Interactive Media Experiences (IMX '26)*. Athlone, Ireland: Association for Computing Machinery. https://doi.org/10.1145/3788851.3812846

Świdrak, J., Georgiou, O., Banakou, D., Matamala-Gomez, M., Pochwatko, G., & Seinfeld, S. (2024). Beyond audiovisual: novel multisensory stimulation techniques and their applications. *Frontiers in Virtual Reality*, 1491854. *https://doi.org/10.3389/frvir.2024.1491854*

Interactive VR360 for Scientific Heritage: A Gaze-Based Approach to Demonstrating Historical Instruments

Zoltán Kárpáti[1], Katalin Németh[1], George Gkanias[2], Christos Lougiakis[2], Maria Boile[2], and Maria Roussou[2]

[1]*Leopoly*

[2]*ATHENA Research and Innovation Center, Greece*

Corresponding author: *vivi@di.uoa.gr*

Keywords: Scientific Heritage, Cultural Heritage, 360° Video, Immersive Virtual Reality, Immersive Learning

Abstract

Historical scientific instruments are challenging to interpret because understanding them may require communicating not only their appearance but also how they were assembled, operated, and used in scientific practice. This application paper presents VR360, which combines immersive 360° video with gaze-based interaction to support the presentation of such procedural and contextual information. We demonstrate the approach through a nineteenth-century carbon arc projector reconstructed within a historical microbiology laboratory. Users observe its assembly and operation while accessing contextual information, alternative viewpoints, and knowledge questions through gaze-based interaction, with optional 3D object inspection. We discuss the potential of interactive VR360 as a lightweight approach for presenting operational aspects of scientific heritage and supporting the reuse of existing immersive heritage content.

Introduction

Historical scientific instruments are not only material artefacts but traces of scientific and educational practice. Their significance may depend on understanding how they were assembled, adjusted, operated, and used to produce or communicate scientific knowledge (Lourenço and Gessner, 2014; Ludwig and Weber, 2013). This presents a challenge for heritage interpretation: instruments preserved in museums may be fragile, incomplete, obsolete, or unsuitable for operation, while their appearance alone may reveal little about how they functioned.

Immersive Virtual Reality (VR) can address this challenge by reconstructing historical instruments and enabling users to experience aspects of their operation (Bekele et al, 2019; Johnson-Glenberg, 2018). However, highly interactive VR applications commonly rely on complex locomotion and interaction techniques that users must learn (Bekele et al, 2018; Makransky and Petersen, 2021). Interactive 360° video experiences provide an alternative immersive format, combining panoramic video with contextual information, navigation, and interactive hotspots (Argyriou et al, 2020; Kabassi, 2019; Osman et al, 2009, Škola et al, 2020). Gaze-based interaction can further reduce interaction requirements.

This paper presents VR360, an interactive approach for communicating the operation and historical context of scientific instruments. The approach combines immersive demonstration with gaze-based information access, alternative viewpoints, 3D object exploration, and knowledge questions. It is demonstrated through an experience centred on a nineteenth-century carbon arc projector reconstructed within an early twentieth-century microbiology laboratory. The application additionally illustrates how digital assets and educational content originally developed for an interactive VR experience can be repurposed for an alternative immersive format.

The Interactive VR360 Approach

The VR360 Viewer presents immersive 360° video through a VR head-mounted display, enriched with spatially embedded interactive elements. The core narrative and informational interactions are gaze-based and do not require handheld controllers; direct manipulation of embedded 3D objects currently requires controllers. At selected points, video playback pauses and users can access four principal types of interaction:

1. **Information panels**, providing technical and historical context;
2. **Viewpoint hotspots**, allowing users to observe a process from alternative pre-recorded perspectives;
3. **3D object hotspots**, providing explorable models of relevant components; and
4. **Question panels**, introducing short knowledge questions.

The experience combines the temporal characteristics of video with self-paced exploration. This is particularly relevant to scientific instruments, where users may need both to observe how an operation unfolds over time and to examine the individual components involved.

Authoring the Experience

To support the creation of interactive VR360 experiences, the framework is being extended within the <Anonymous> project with a web-based no-code authoring environment, intended to reduce the technical expertise required to produce immersive content. The editor allows creators to upload or select 360° video and visually navigate the panoramic scene while positioning interactive elements directly within it. A timeline interface controls when these elements appear and disappear and how video playback responds to interaction.

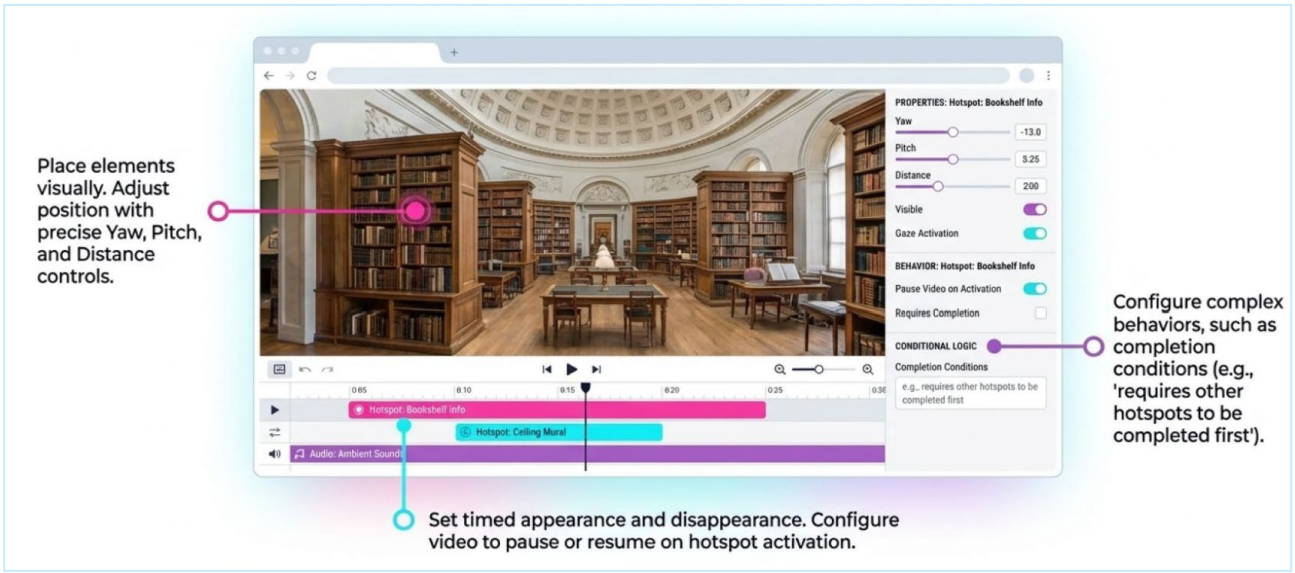


Figure 1. *The VR360 Creator Canvas*

Authors can add information hotspots containing text, images, or 3D content; dialogue hotspots presenting questions and selectable answers; and movement hotspots connecting different 360° scenes. Behavioural rules can additionally define conditions and branching, for example requiring completion of an interaction before progression or directing different answers to different content. An integrated localization editor supports multilingual content.

The complete experience can be previewed in the browser before publication, including its gaze-based interaction and narrative flow. This separation between content authoring and the final immersive presentation is particularly relevant to cultural heritage, where curators, educators, and experience designers may need to construct and adapt interactive narratives without developing a dedicated VR application for each experience.

Application Case: The Carbon Arc Projector

The application centres on a carbon arc projector manufactured by P.H. Pellin in Paris during the late nineteenth century and currently exhibited at the History Museum of the National and Kapodistrian University of Athens. The instrument is presented within a virtual reconstruction of a microbiology laboratory in the early twentieth century. The interpretive objective is to communicate the projector not simply as a museum object, but as an operational instrument situated within historical scientific practice.

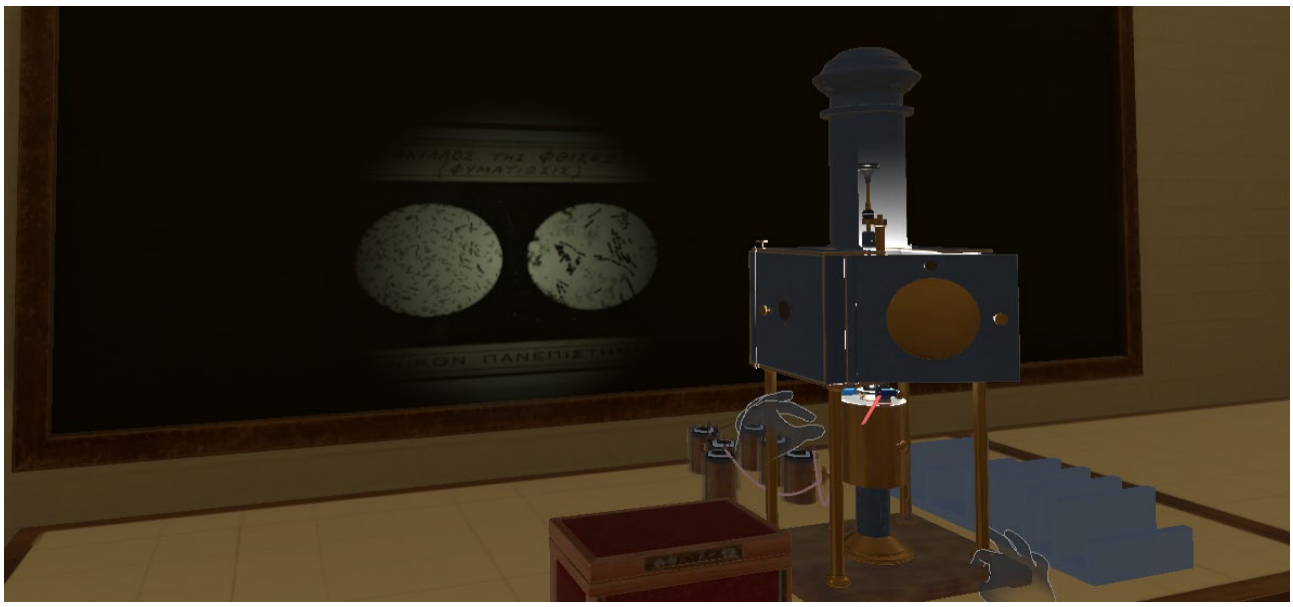


Figure 2. *The assembled projector is used to view a slide relevant to Malaria.*

The digital environment and projector were originally developed for an interactive VR application in which users assembled the instrument by positioning its components, including the internal mechanism, carbon rods, mirror, and lens. Once assembled, the projector was used to display historical educational material.

For the VR360 application, the existing virtual laboratory, projector model, animations, historical material, and educational narrative were reused. 360° footage was captured from several virtual camera positions in Unity, while selected instrument components were exported as separate 3D objects for interactive inspection. In the resulting experience, users observe the projector being assembled through short immersive video sequences. At meaningful points, playback pauses and gaze-activated hotspots allow users to inspect the current component in 3D, access an explanation, select a more informative viewpoint, or answer a short question. Interaction therefore follows a recurring sequence:

Observation → Reflection → Interaction → Continuation.

After assembly, users observe the projector in operation within a historical lecture on malaria inspired by early twentieth-century public-health education in Greece. Projected slides connect the technical operation of the instrument with its historical purpose as a medium for communicating emerging scientific knowledge.

The application presents the instrument at three interconnected levels: as a material object, through its components; as an operational system, through the assembly and projection process; and as part of scientific practice, through its historical educational context.

Discussion and Conclusions

The VR360 application illustrates the potential of interactive 360 video for communicating scientific heritage. Immersive video makes the operation of an instrument visible within its reconstructed spatial context, while interactive pauses, alternative viewpoints, contextual information, and 3D inspection provide opportunities to explore the demonstrated process beyond passive viewing.

Interactive VR360 occupies a different position from fully interactive VR rather than replacing it. Direct manipulation offers forms of embodied engagement that 360° video cannot reproduce. Conversely, gaze-based interaction removes controllers and reduces the interaction conventions that visitors must learn. This may make the format particularly relevant to short museum experiences and audiences with limited XR experience.

The case demonstrates opportunities for content reuse. Existing 3D models, animations, historical material, and educational narratives can provide source material for alternative immersive experiences rather than remaining tied to a single VR application. Future work will examine visitor engagement and learning potential for procedural understanding.

References

Argyriou, L., Economou, D. & Bouki, V. (2020). Design methodology for 360° immersive video applications: The case study of a cultural heritage virtual tour. *Personal and Ubiquitous Computing* 24(6), 843–859.

Bekele, M.K., Champion, E., McMeekin, D.A., Rahaman, H. & Pierdicca, R. (2019). A comparison of immersive realities and interaction methods: Cultural learning in virtual heritage. *Frontiers in Robotics and AI* 6, 91. https://doi.org/10.3389/frobt.2019.00091

Bekele, M.K., Pierdicca, R., Frontoni, E., Malinverni, E.S. & Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. *ACM Journal on Computing and Cultural Heritage* 11(2), 1–36.

Johnson-Glenberg, M.C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI* 5, 81. https://doi.org/10.3389/frobt.2018.00081

Kabassi, K., Amelio, A., Komianos, V. & Oikonomou, K. (2019). Evaluating museum virtual tours: The case study of Italy. *Information* 10(11), 351. https://doi.org/10.3390/info10110351

Lourenço, M.C. & Gessner, S. (2014). Documenting collections: Cornerstones for more history of science in museums. *Science & Education* 23(4), 727–745.

Ludwig, D. & Weber, C. (2013). A rediscovery of scientific collections as material heritage? The case of university collections in Germany. Studies in History and Philosophy of Science Part A 44(4), 652–659.

Makransky, G. & Petersen, G.B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review* 33(3), 937–958.

Osman, A., Wahab, N.A. & Ismail, M.H. (2009). Development and evaluation of an interactive 360° virtual tour for tourist destinations. *Journal of Information Technology Impact* 9(3), 173–182.

Škola, F., Rizvić, S., Cozza, M., Barbieri, L., Bruno, F., Skarlatos, D. & Liarokapis, F. (2020). Virtual reality with 360-video storytelling in cultural heritage: Study of presence, engagement, and immersion. *Sensors* 20(20), 5851. https://doi.org/10.3390/s20205851

Inmersidra: Bridging Virtual Gestures and Physical Action in XR for Asturian Cider Pouring

Irene Bouzón, and Roberto Abad

Technological Center for Information and Communication (CTIC), W3C Spain Office host, Ada Byron 39, 33203 Gijón, Spain

Corresponding author: irene.bouzon@fundacionctic.org (I. Bouzón)

Keywords: Virtual reality, 3D user interfaces, embodied interaction, cultural heritage, hybrid XR, serious game, Asturian cider.

Introduction

Extended Reality (XR) is increasingly being used not only for visualization, but also for designing embodied forms of interaction that connect digital environments with physical actions and tangible outcomes. In this context, cultural heritage offers a particularly relevant application domain, as immersive and game-based experiences can foster participation, accessibility, and engagement with traditions that may otherwise remain difficult to communicate to broader audiences. This is especially relevant for living and intangible heritage, where gestures, rituals, and social practices are central to the experience itself.

Inmersidra is presented as a demonstrator and application case: a hybrid XR experience that recreates the traditional Asturian cider-pouring gesture through a synchronized interaction between a virtual environment and a physical device. Asturian cider culture was inscribed on UNESCO's Representative List of the Intangible Cultural Heritage of Humanity in 2024, underlining its social and cultural significance beyond the regional context.



Figure 56. *Asturian cider-pouring traditional gesture. Photographs by Miki López, reproduced with permission.*

Asturian cider pouring, or "*escanciado*", is the traditional serving technique in which cider is poured from above the head into a tilted glass held below the waist, so that the stream hits the rim and helps awaken the cider's natural effervescence. Beyond its functional purpose, the gesture is highly embodied, requiring posture, coordination, timing and precision, and it is traditionally transmitted through imitation and oral practice. These characteristics make it an ideal case for exploring how XR can translate a culturally meaningful real-world action into an immersive 3D interaction metaphor. Inmersidra explores how XR can be used to reinterpret such a tradition as a participatory, embodied and publicly shareable experience.

System Overview

Inmersidra combines a VR headset, an interactive 3D environment, and a connected physical cider-pouring device. In the experience, the user enters a simulated cider house and performs the pouring gesture through a step-by-step tutorial, which is in turn translated to a synchronized physical action: the nearby device pours real cider into a real glass. The virtual component is implemented as a PC/VR application in Unity using OpenXR for compatibility with a wide range of headsets. The cider stream is rendered as a particle system with collision detection, enabling the system to estimate pouring height and the impact point on the virtual glass. These parameters are used to compute an accuracy score reflecting how closely the user reproduces the traditional "*escanciado*" gesture.

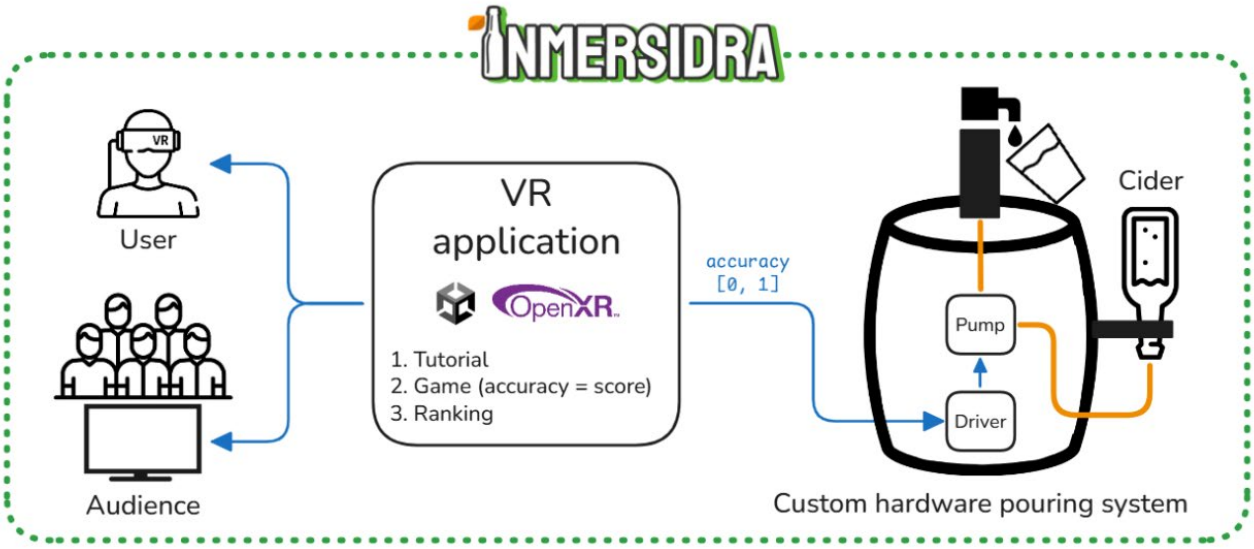


Figure 57. *Overview of the Inmersidra system architecture.*

To connect this virtual interaction with a real-world outcome, the Unity application communicates with a custom-built physical pouring device integrated into a traditional wooden cider barrel. The barrel houses the electronics, tubing, power supply, pump system, high-current driver, and sealed components required to automatically deliver cider from a freshly opened bottle to the serving glass. Physical pouring is directly governed by VR performance over a serial connection signal: cider is actuated only when virtual gesture precision exceeds minimum thresholds, while pump flow is modulated in real time based on the accuracy score. Accurate VR execution triggers a continuous physical pour into the real glass, whereas poor gesture accuracy reduces or halts liquid flow, establishing an immediate physical feedback loop.

From an interaction design perspective, Inmersidra can be understood as a 3D user interface based on embodied metaphor. The user does not interact through abstract menus or detached commands, but through the enaction of a culturally meaningful movement that already exists in the real world. The experience therefore relies on the user's body as the primary interface and on gesture as the central interaction metaphor. The core design challenge is not only technical synchronization, but also experiential coherence: the user must perceive that the action performed in VR and the

result observed in the real world belong to the same continuous interaction. Inmersidra addresses this by aligning posture, timing, visual cues, and physical output so that the gesture feels meaningful and causally legible.



Figure 58. *Left: Inventor Martin Cooper experiencing Inmersidra. Right: External display of the VR session, showing the user's avatar, the cider-pouring interaction, and the gamified interface (time and score feedback).*

Conceived for public deployment in fairs and showcases, the system features a guided tutorial, a gamified ranking phase, and a secondary display for surrounding audiences. This structure combines individual immersion with collective spectatorship, leveraging serious gaming and edutainment principles to lower entry barriers and encourage active participation.

Cultural heritage dissemination often faces the challenge of making local or intangible practices understandable and engaging to people who are unfamiliar with their context. XR offers a way to address this challenge by transforming heritage into spatial, interactive and experiential media. Inmersidra contributes to this direction by focusing not on the representation of an object, but on the transmission of a gesture-based social ritual. This is particularly important in the case of Asturian cider culture, where meaning emerges not only from the beverage itself, but from the associated spaces, gestures, social codes and performative practices. By reinterpreting cider pouring as a hybrid XR interaction, Inmersidra makes this ritual more accessible to wider audiences while preserving its performative dimension, resulting in an experiential mediation of the heritage.

Discussion and Concluding Remarks

Inmersidra illustrates how hybrid XR systems can be leveraged to connect embodied interaction with tangible real-world effects in a cultural heritage context. Rather than focusing on visual reconstruction alone, the project foregrounds the design of an interaction loop in which a culturally meaningful gesture drives both virtual feedback and physical actuation. This makes the experience relevant to core topics such as 3D user interfaces, embodied interaction metaphors, novel input devices, tangible interfaces, cultural heritage dissemination, and serious gaming. The system also highlights how cultural heritage and playful interaction can be productively combined, supporting both individual immersion and shared spectatorship. The serious game elements do not replace the heritage dimension, but support accessibility, repeatability, and audience participation. As such, Inmersidra can be positioned at the intersection of hybrid XR interaction design, cultural heritage dissemination, and experiential edutainment, offering a demo case for how XR can bridge virtual and physical action. To date, the application has been publicly deployed across multiple itinerant events and a permanent demonstrator, recording over 560 completed session score submissions from diverse user demographics, proving system robustness in real-world public settings.

This paper introduced Inmersidra as a hybrid XR experience in which a user's gesture in virtual reality triggers a synchronized physical cider-pouring action. The project illustrates how XR can

connect virtual and material interaction through embodied metaphors while serving as a vehicle for intangible cultural heritage dissemination. In doing so, Inmersidra provides a concrete, audience-tested example of how XR can support cultural engagement, playful participation, and hybrid human-computer interaction in public settings.

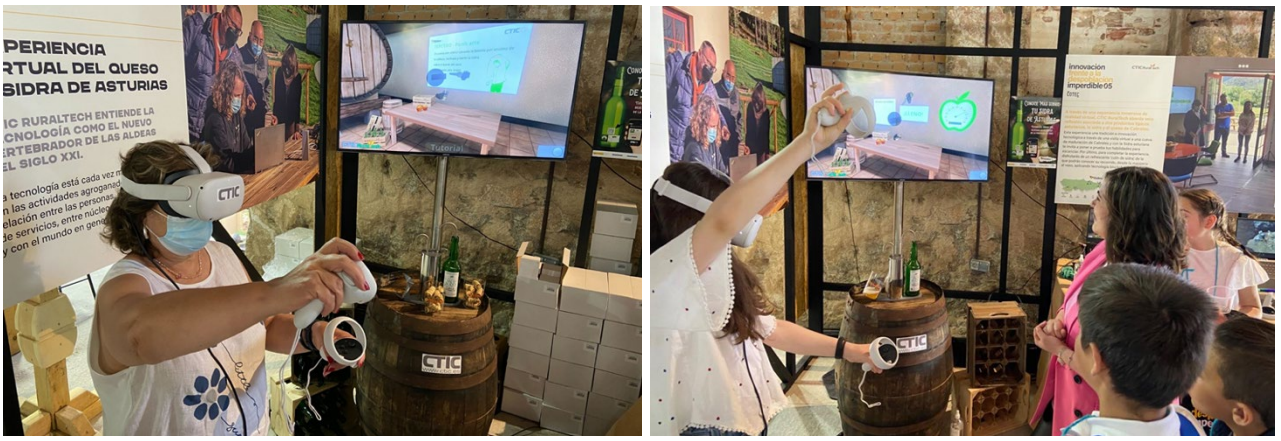


Figure 59. *Inmersidra booth at events, engaging both older adults (left) and children (right), highlighting its inclusive design for diverse audiences.*

Acknowledgements

The authors acknowledge TACTIX (CER-20251027), a CERVERA Excellence Network funded by the Spanish Ministry of Science, Innovation and Universities through CDTI, for supporting dissemination activities related to this work. AI was used for English-language translation and editing; the authors reviewed the final text and remain fully responsible for its content.

References

Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S., & Gain, J. (2018). A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *Journal on Computing and Cultural Heritage* (JOCCH), 11(2), 7:1-7:36. https://doi.org/10.1145/3145534

Hornecker, E., & Buur, J. (2006). Getting a grip on tangible interaction: A framework on physical space and social interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, 437–446. https://doi.org/10.1145/1124772.1124838

UNESCO. (2024). Asturian cider culture. Representative List of the Intangible Cultural Heritage of Humanity. Available online: https://ich.unesco.org/en/RL/asturian-cider-culture-01959 (accessed 3 August 2026).

Building the Ecosystem: A Multi-Level Approach to Scaling XR in Higher Education

Geraldine McDermott

Building the Ecosystem: A Multi-Level Approach to Scaling XR in Higher Education

Corresponding author: geraldine.mcdermott@tus.ie

Keywords: Extended Reality (XR), higher education, capacity-building

Introduction

Extended Reality (XR) adoption in higher education has frequently been characterised by isolated experimentation: a single enthusiastic lecturer, a one-off pilot module, and limited institutional follow-through once initial funding or novelty fades. Research on barriers to the adoption of immersive learning consistently identifies access, affordability, inadequate staff training, interoperability, and a lack of reusable content as recurring obstacles to moving beyond the pilot stage (see Albishiri & Blackmore, 2025; Radianti et al., 2020; Upadhyay et al., 2024). Addressing these barriers arguably requires more than isolated projects; it requires a deliberate strategy operating simultaneously across multiple levels of the higher education system. This paper presents a case study of a five-strand, multi-level approach to building sustainable capacity for XR in teaching and learning, developed and led by the author over a four-year period. The initiatives span national, cross-institutional, institutional, and discipline-specific levels: (i) a national one-day symposium on XR in education and training; (ii) a cross-institutional Open Course on "Teaching with XR", developed collaboratively across five Higher Education Institutions (HEIs) in Ireland and awarding a digital badge to successful participants; (iii) an institutional XR Community of Practice (CoP) bringing together educators, researchers, and XR professionals; (iv) a larger-scale, funded project embedding immersive learning within Nursing education at the author's university; and (v) a portfolio of funded learning enhancement projects led by discipline-based lecturers across a range of disciplines.

This paper sets out that model and shares transferable lessons for institutions and practitioners seeking to scale XR adoption.

A Multi-Level Model for Scaling XR Adoption in Higher Education

Each of the five initiatives included above targets a distinct stage in the XR adoption pathway within the higher education sector. However, while individual initiatives have merit, it is the combination of these activities, positioned strategically along a growth continuum, that is proposed as a model for XR adoption in institutions or sectors seeking to move beyond isolated XR projects. The model below illustrates the five levels of capacity building. Each level is presented from the author's perspective and considers the challenges and opportunities against the broader backdrop of challenges for the immersive sector, as outlined by the XR4Europe report (2025).

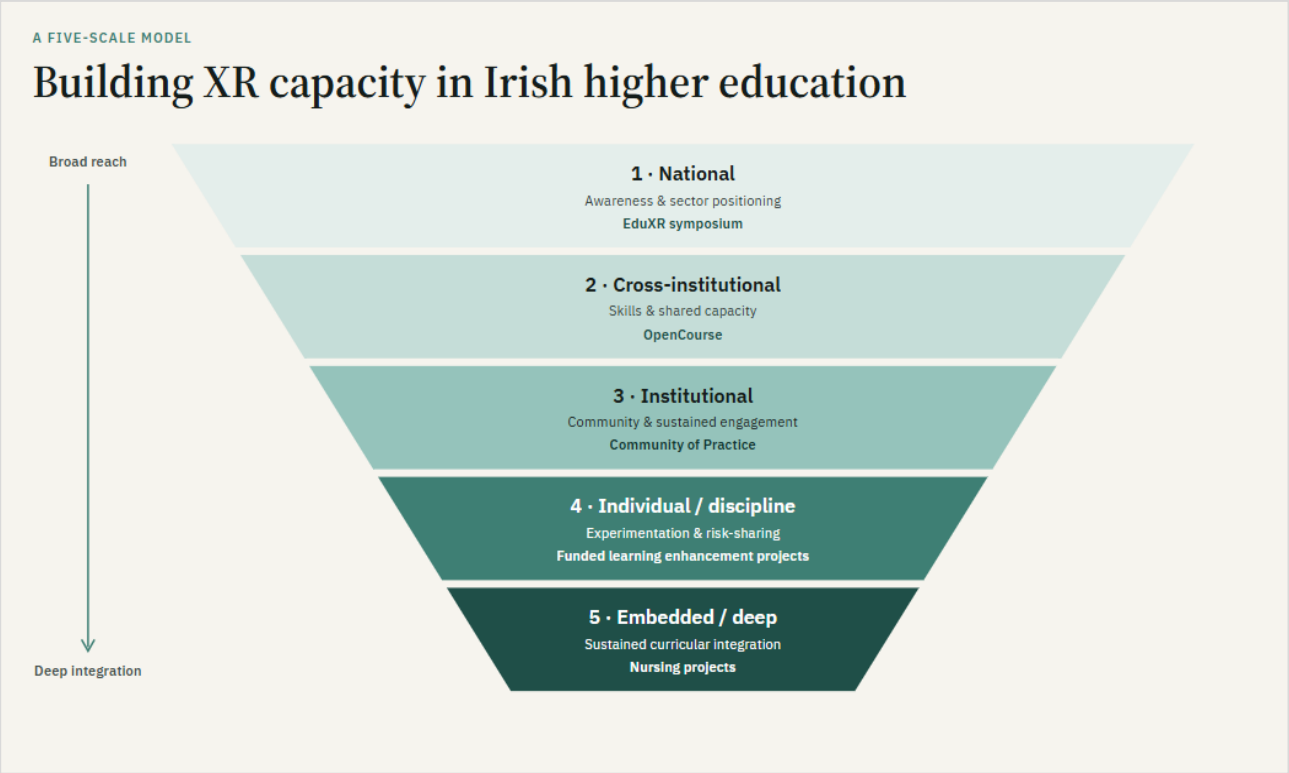


Figure 60. *A five-scale model for building XR capacity in Irish higher education. (Note: Graphic created using Claude.ai)*

Initiatives and Outcomes

National Symposium (EduXR)

The national symposium on EduXR is the result of a collaboration between XR practitioners and professionals from five further and higher education institutions in Ireland and positions itself as an annual event to share experiences, strengthen networks, and forge new connections within the growing XR community in Ireland. Viewed through the lens of value co-creation, the symposium is considered a “co-creative ecosystem” (Ting et al, 2026, p.1), where its value “lies in the dialogue, relationships, and shared meaning-making that unfold across and beyond it” (ibid). Understanding the importance of the social, functional and emotional values of a networking event such as the EduXR symposium (Zhang et al, 2024), the organising committee agreed on guiding principles from the outset, notably collegiality, sustainability, affordability and a commitment to diversity, in its selection of keynote speakers, and academic and research track presentations.

Service-Dominant Logic (Vargo & Lusch, (2016)) offers a complementary lens, moving towards an understanding that value is co-created for attendees and organisers beyond the symposium programme. Rather than the passive consumption of resources or presentations, the shared meaning-making, engagement and interaction amongst attendees at the symposium create value, which is “dynamic and co-constructed” (Ting et al., 2026, p. 4).

Collaboration has emerged as a central theme for the national EduXR symposium, and understanding the role of educators, researchers, industry partners, and learners in contributing to value co-creation is a key objective for the organisers. Making space for dialogue across different industry and education sectors to question, exchange knowledge and build relationships is at the core of the symposium design (Verbeke, 2015).

Cross-Institutional Open Course: "Teaching with XR: Designing and Delivering an Immersive Learning Experience"

This course is the result of a successful funding application to the Higher Education Authority (HEA) in Ireland. XR experts and practitioners from five higher education institutions are responsible for developing a 25-hour, facilitator-led course on the use of XR in an educational context, hosted on a national Open Course platform. The rationale for developing this professional development opportunity is to move beyond the workshops and non-formal learning to a more sustainable, structured space within a community of XR practitioners. Participants are awarded a digital badge upon completion of the course assessment activities, providing recognition of their learning through a popular digital badge architecture/ecosystem. Digital badges as a recognition mechanism are increasingly used in higher education to credential non-formal professional learning at a more granular level (Stefaniak & Carey, 2019). The process of cross-institutional development revealed a need to align pedagogical approaches and combine expertise and resources across partner institutions. The addition of a facilitator badge (learners are allowed to facilitate the course at their own institution once the badge is awarded) supports the course's sustainability, as facilitators can customise it to suit their local context.

An Institutional XR Community of Practice

Communities of Practice (CoPs) are widely used in higher education to sustain staff engagement with new practices or around a particular common interest. However, the CoP's sustainability is also flagged as a recurring challenge (Baker & Beames, 2016). The institutional CoP comprises educators, researchers, and professional staff, and its mission is "to advance the transformative potential of extended reality in higher education by sharing innovative teaching and learning practices, pioneering interdisciplinary research, and contributing to industry-focused solutions" (Technological University of the Shannon, 2025). To address issues of longevity, the CoP is grounded in the values of the university, i.e. sustainability, inclusivity, and collaboration. It also aims to foster equitable access to immersive technologies, support lifelong learning, and cultivate partnerships that contribute to resilient and forward-looking societies. The cadence and focus of the activities vary, and the challenges continue to include sustained engagement and "aliveness" (Wenger et al., 2002, p.50). However, taking cognisance of Wenger et al's (2002) principles for cultivating communities of practice has contributed to its sustainability, while a commitment to a collective review of its goals creates value for its members.

Embedded Immersive Learning in Nursing

An EU-funded initiative (National Technological University Transformation for Recovery and Resilience (N-TUTORR)) provided an opportunity within the author's institution to "deliver a best-practice suite of initiatives and opportunities for the learner, for staff development and supported by necessary enabling technologies" (NTUTORR, 2024). Aligning with the digital transformation pillar, the author proposed adopting immersive learning to align with the HEA's healthcare capacity-building process (HEA, 2022). General and Mental Health Nursing programmes are offered by the author's HEI, and it was considered that immersive learning could provide additional opportunities for situated learning. Engaging with the programme coordinator and subject matter expert was crucial to identifying the pedagogical objectives and clinical scenarios that would help meet these. The initiative was successful among second- and third-year BSc in General Nursing students (*N=130*), who used a clinical simulation platform and voted overwhelmingly to increase the use of virtual reality in their programme (Dempsey & McDermott, 2025). The sustained curriculum integration depended on the investment and commitment of the Nursing and Healthcare department. 2026 marks the third year that virtual reality has been used in nursing education at the university, with the expansion to mental health nursing organised this year. This project demonstrates how XR can move from experimentation to embedded practice.

Funded Learning Enhancement Projects

Small, funded learning enhancement projects lower the barrier to entry for discipline-based lecturers to trial XR in their own teaching context. Within an Irish context, the HEA's Strategic Alignment of Teaching and Learning Enhancement (SATLE) fund promotes " innovation,

collaboration, and the enhancement of teaching and learning practices across Irish higher education institutions (HEI), with an emphasis on sustainable development, digital transformation and academic integrity" (HEA, n.d.) This fund is managed by individual HEIs and, in the author's context, is distributed across small-scale learning enhancement projects and communities of practice. To date, domains represented include Social Sciences, Life Sciences, Business, Design, Apprenticeship Education, Student Wellbeing and the university library. Collaboration among colleagues across a distributed campus is embedded in the application criteria, and the author provides pedagogical support. The technical infrastructure within the university is advancing to support these projects, and, as with many small-scale projects, the challenge for the university is embedding the immersive learning experience beyond the pilot's lifetime (Moolman, 2025).

Conclusion and Future Work

This paper has presented a multi-level model for scaling XR adoption in higher education, grounded in five initiatives spanning discipline-embedded, institutional, cross-institutional and national levels. Each initiative provides an opportunity for collaborative interactions, the cornerstone for successful XR adoption across education. The approach offers a transferable framework for institutions and practitioners, who are seeking to move XR adoption beyond isolated pilots toward sustainable, embedded practice. Mapping existing initiatives to the model can provide XR practitioners with a roadmap for the future, as it is argued that the combination of discipline, institutional, cross-institutional and national initiatives are needed for effective XR adoption.

References

Albishri, B., & Blackmore, K. (2025). Duality in barriers and enablers of augmented reality adoption in education: a systematic review of reviews. *Interactive Technology and Smart Education*, 22(2),167–191.

Baker, A., & Beames, S. (2016). Good CoP: What Makes a Community of Practice Successful? *Journal of Learning Design*, 9(1), 72-79.

Barngrover, M. (2025). XR funding in H2020 and HEU: Public report. XR4Europe. https://xr4europe.eu/wp-content/uploads/H2020-HEU-Virtual-Worlds-Report-Public-Version.pdf [Accessed 13 September 2026].

Dempsey, L. & McDermott, G. (2025). Implementing and Embedding Virtual Reality in an Undergraduate BSc in General Nursing Programme. In McDermott, G. (Ed). *Shaping the Future of Immersive Technology in Education and Training*. Pressbook. https://tus.pressbooks.pub/eduxr25/chapter/laura-dempsey-and-geraldine-mcdermott-implementing-and-embedding-virtual-reality-in-an-undergraduate-bsc-in-general-nursing-programme/

HEA. n.d. Strategic Alignment of Teaching and Learning Enhancement Funding in Higher Education. https://www.teachingandlearning.ie/funding/ [Accessed 14 September 2026].

HEA. 2022. HEA Capacity Building 2022: New and Expanded Courses. https://hea.ie/policy/he-reform/hea-eoi-2022-new-and-additional-courses/ [Accessed 13 September].

HEA. 2024. NTUTORR Transforming Learning. https://www.transforminglearning.ie/ [Accessed 12 September].

Moolman, J. (2025). VISIEN Virtual and Immersive Solutions for Integrated Educational Needs. VISIEN-Integrating-Immersive-Technologies-into-Education-A-REEdI-Framework.pdf [Accessed 13 September].

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Stefaniak, J., & Carey, K. (2019). Instilling purpose and value in the implementation of digital badges in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), 44.

Ting, H., Kumarusamy, R., Vuong Le Hong, L. Y., Villamin, J. M., Soonsan, N., & Tan, K. L. (2026). To present or to be present? Rethinking the value of the academic conference. *Asian Journal of Business Research Volume*, 16(2).

Technological University of the Shannon (2025). TUS XR Special Interest Group Mission, Values and Strategic Pillars. [Unpublished report].

Upadhyay, B., Chalil Madathil, K., Hegde, S., Anderson, D., Wooldridge, E., Presley, D., Perez, L., & Reid, B. (2024). Barriers toward the implementation of extended reality (XR) technologies to support education and training in workforce development programs. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 68 (1), 265-269. Sage CA: Los Angeles. https://doi.org/10.1177/10711813241275080

Vargo, S.L., & Lusch, R.F. (2016). Institutions and axioms: an extension and update of service-dominant logic. *Journal of the Academy of Marketing Science*, 44(1), 5-23.

Verbeke, J. (2015). Designing academic conferences as a learning environment: how to stimulate active learning at academic conferences?, *Constructivist Foundations*, 11(1), 98-105.

Wenger, E., McDermott, R., & Snyder, W. (2002). Communities of Practice: A Guide to Managing Knowledge. *Harvard Business School Press*, Cambridge, USA.

Zhang, X., Singh, S., Li, J., & Shao, X. (2024). Exploring the effects of value co-creation strategies in event services on attendees' citizenship behaviors: the roles of customer empowerment and psychological ownership. *Journal of retailing and consumer services*, 76, 103619. https://doi.org/10.1016/j.jretconser.2023.103619

Instructor-led Mobile Augmented Reality for Synchronous Lessons

David Kunz[1], Pramod Biradar[1], Marvin Schobert[1], Jörg Franke[1], Jens Fürst[1], Robert Köhler[2], Sven Haschke[3], Christoph Runde[3]

[1]*Institute for Factory Automation and Production Systems, FAU Erlangen-Nürnberg*

[2]*Robert Bosch GmbH*

[3]*Virtual Dimension Center w.V.*

*Corresponding author: david.kunz@faps.fau.de

Keywords: Education; Instructor-led; Mobile Augmented Reality; Synchronous; In-Classroom

Introduction

In synchronous education, where an instructor guides learners in real time, mobile augmented reality (MAR) allows virtual 3D objects to be integrated into both physical and virtual classrooms. To adapt the flow of the lesson in accordance with their teaching and presentation style, instructors require tools for moderation beyond what most current applications offer. This publication takes a look at the background and requirements and showcases a sample implementation for technology training from an industrial use case.

Background and Use Case

MAR uses smartphone or tablet screens to visualize interactive 3D objects and animations or virtually add to real objects. This enables views that are difficult to achieve in real life, such as making the invisible visible, size variations, and showing interactions through animations. Furthermore, a virtual visualization permits showing objects at any time and place even if the real object is unavailable or difficult to attain. When used in education it has been shown to improve students‘ spatial understanding of complex 3D geometries and interactions (Papakostas et al., 2021). MAR is especially well suited for education due to the widespread availability of suitable devices, users' familiarity with the controls, and instructor's ability to observe their classroom and thus manage and orchestrate it. But current applications mostly focus on application scenarios for self-exploration by the learners in individual or group tasks.

This leaves a gap for when educators want to embed MAR visualizations in instructor-led, synchronous sessions, e.g. for discussing complex products such as DC-DC converters during in-house training. Unlike asynchronous use, where scaffolding must be pre-authored into the application, synchronous sessions let instructors add scaffolding live. This enables them to adapt to the pace, knowledge and background of their audience but requires orchestration tools that let instructors steer what learners see and do.

In our use case which is part of the XR-Upskill project, trainers use internal products such as a DC-DC converter across several training formats: onboarding for new recruits, refresher courses for

existing staff, reskilling for employees moving into new focus areas, and in-house product training. Depending on the target group the expected learning outcomes, framing and therefore depth at which the component is discussed vary. At the same time, new versions of the DC-DC converter are released regularly and producing a new physical cutaway model is expensive and ties the training to a single location.

Virtual models are well-suited for this task as they are not limited to one place and relatively quick and inexpensive to update to new models. At the same time, the instructor-led, synchronous training setting benefits from the guidance and experience of the expert instructors and makes adaptation to the individual target audience possible. In our AR classroom setting (Kunz et al. 2026), instructors aim to convey the task, operating principle, core components, energy flow and heat management of the object in a locked mode where only they can change views, but also give the learner the option to freely explore the object afterwards without changing applications. To simplify use in the given context, the application should work within the browser and enable a seamless transition between both modes within the same webpage. Shared local anchors are not required for the selected use case.

The following showcases the design and implementation of an AR application for synchronous learning that follows an asymmetric control model, enabling the instructor to guide the learners' attention to the AR object by switching between an instructor mode and an exploration mode.

System Design and Implementation

The AR application is split into two modes: a locked instructor mode and an exploration mode. In the former, the instructor selects which step is visible and if additional information such as component name, info or images are shown in a pop-up. In the exploration mode, the learners are able to freely change the steps and select components to obtain additional information.

The application is built on a client-server architecture in which the AR client runs entirely within the user's web browser and communicates with a cloud-based backend service for real-time session state synchronization. This architectural separation keeps the client-side requirements limited to standard web browser capabilities, while the backend primarily handles state synchronization rather than computationally intensive processing. Consequently, the backend can serve an arbitrary number of connected clients without requiring dedicated infrastructure management. The overall data flow follows a unidirectional update model: the trainer client writes session state to the Firebase Firestore backend, and all connected student clients receive updates through real-time listeners without polling or manual refreshing.

The architecture consists of four principal layers, see Figure 1. The authoring layer comprises the Unity Editor and the Needle Engine export pipeline, which together produce the deployable web application assets. The presentation layer consists of the browser-based AR client, implemented in TypeScript using Needle Engine's component system and rendered through the WebXR Device API. The synchronisation layer is implemented using Firebase Firestore, which maintains the shared session state document and distributes real-time updates to all connected clients. The hosting layer uses the Vercel deployment platform, which serves static web application assets to client devices over Hypertext Transfer Protocol Secure.

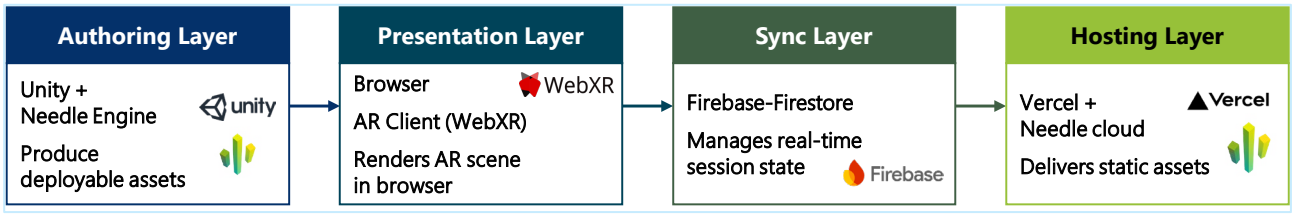


Figure 61. *System architecture overview.*

The AR client is the primary user-facing component of the platform, responsible for rendering the three-dimensional DC-DC converter model, managing assembly step progression, handling user interaction, and communicating session state changes to and from the Firebase backend. The client is implemented as a TypeScript-based web application produced by the Needle Engine export pipeline from the Unity scene. The data flow of the platform follows a publish-subscribe pattern mediated by Firebase Firestore, in which the trainer client acts as the sole publisher of session state and all student clients act as subscribers. This asymmetric data flow model ensures that session state has a single authoritative source, eliminating the possibility of conflicting concurrent writes from multiple clients during a structured training session. Figure 2 shows the instructor view of the application on a smartphone. The main difference in the learner mode is that the mode selection is hidden.

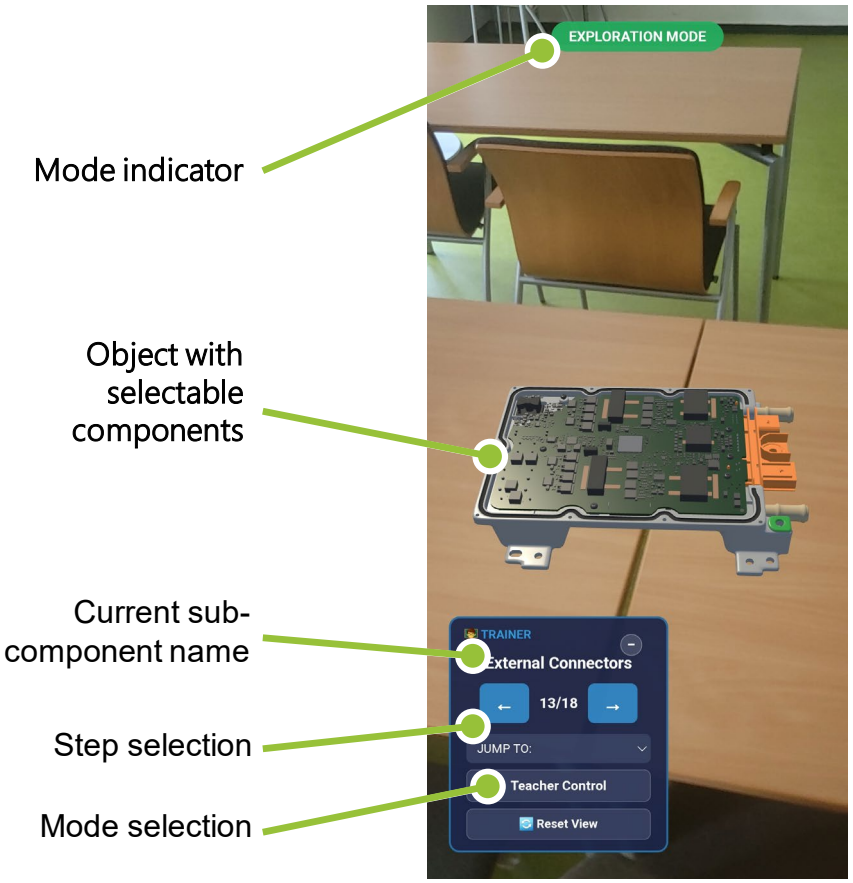


Figure 62. *Instructor view on a smartphone with highlighted UI elements.*

The model was prepared based on a CAD model which was remodelled as .glb in Blender. The 18 subdivisions of the model are based on the assembly of the converter and selected in close consultation with the instructors.

Preliminary Evaluation and Results

The application was continuously evaluated by instructors who checked the clarity of the interface and controls as well as correctness of the model and of the information displayed in the pop-ups. A first evaluation with a small group of learners from the main target audience also yielded positive feedback. Instructors rated the asymmetric control model and ability to switch between the two states as well-suited to dynamic synchronous training formats. The final version tested successfully on multiple Apple and Android devices, with a PC browser version as fallback.

One recurring criticism concerned the level of geometric detail: the model for MAR is necessarily less detailed than the CAD data that instructors and learners with an industry background are accustomed to. This criticism shows that users' expectations of fidelity, shaped by their professional experience, strongly influence how they perceive the quality of educational extended reality. Adapting the application to other models is straightforward, with the main effort being model preparation. Licensing costs are low and depend on the institution type and number of developer seats.

Summary and Outlook

The presented application illustrates how MAR can be integrated into instructor-led, synchronous learning formats in which 3D models are central to the learning objectives. It combines instructor control over the pacing of the session and over the scaffolding of the content with the learners' freedom to explore the models on their own.

In doing so, this work addresses the lack of AR experiences designed for instructor-led settings rather than for individual, self-paced use. Next steps include an in-class evaluation with the target groups described above, functional extensions such as letting instructors highlight components or direct attention to specific regions of the model, and the application of the approach to other products, for example inverters.

This work was conducted within the XR-Upskill research project, funded through the InnoVET Plus program of the German Federal Ministry of Education, Family Affairs, Senior Citizens, Women and Youth (BMBFSFJ) and implemented by the Federal Institute for Vocational Education and Training (BIBB).

References

Kunz, D., Schobert, M., Hofmann, B., Rehman, H., & Fürst, J. (2026). Immersive Lehrkonzepte für Lean Management: Augmented Reality in synchronen Lernumgebungen. Zeitschrift für wirtschaftlichen Fabrikbetrieb, 121(6), 438-442. https://doi.org/10.1515/zwf-2026-1084

Papakostas, C., Troussas, C., Krouska, A., & Sgouropoulou, C. (2021). Exploration of Augmented Reality in Spatial Abilities Training: A Systematic Literature Review for the Last Decade. Informatics in Education, 20(1), 107-130. https://doi.org/10.15388/infedu.2021.06

Dog Assisted Virtual Environment (DAVE)

Camran Qadeer[1], Marie-Claire Pagano[1], James Oxley[2], Christy Georgiou[1], Craig Rothwell[1], Giuseppe Maria Bellantuono[1], Matthew Butcher[1], Georg Myers[1], Andy Levers[1], Simon Campion, and Carri Westgarth[1]

[1]Virtual Engineering Centre (VEC), University of Liverpool, UK and the Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, in collaboration with Dogs Trust

[2]IVC EVIDENSIA

Corresponding author: C.Qadeer@liverpool.ac.uk

Keywords: Virtual reality, animal behaviour education, human-dog interaction, perspective taking, bite prevention, immersive learning.

Abstract

DAVE (Dog Assisted Virtual Environment) is a virtual reality (VR) application, developed by the University of Liverpool and Dogs Trust, that serves as both a research tool for studying how people respond to canine warning signals and as an educational resource for bite prevention. Rebuilt as a single standalone Meta Quest 3 application, it features a behaviourally validated Labrador co-designed with animal behaviour specialists and a four-mode learning sequence. We summarise evaluations of earlier versions and the current build, including a randomised controlled study in which an informative Tutorial mode increased confidence in recognising aggression, and discuss limitations.

Introduction

DAVE serves a dual purpose: a research tool for studying how people respond to signs of aggression in dogs, and an educational resource that improves understanding of canine body language and early warning signs of aggression to help prevent bites. Educational interventions can improve children's knowledge and behaviour around dogs, but the evidence base is weak (Shen et al., 2017). The application has been developed as individual modules over several years, with three papers published on its inception and use (Oxley et al., 2022a, 2022b, 2024); here we present the complete application, recently optimised for consumer virtual reality (VR) headset deployment with all modules available within a single build. Previous iterations were tethered to a laptop and HTC Vive, which were cumbersome to transport and set up; it has since been rebuilt for the standalone Meta Quest 3 (MQ3).

The primary technical challenge was creating the Labrador. Running natively on a VR headset demanded a careful balance between visual realism and performance: the dog had to look and behave convincingly while still running smoothly enough to avoid disrupting the experience. To achieve this, the polygon budget for the dog mesh was reduced, fur was simplified and lighting was baked rather than real-time. Modelling and animation were carried out in Autodesk Maya, while the VR experience itself was developed in Unity and deployed on a MQ3. The move required a Unity

engine upgrade and a native, on-device performance profile, and all animations were rebuilt to a much higher visual standard.

Users can move freely towards the dog, with realism sufficient to feel co-present with it. Controller vibration simulates the sensation of a bite when triggered, while built-in headphones provide realistic auditory cues such as panting, whines and growls.

The contribution is twofold: a physically and behaviourally validated VR dog model, co-designed with animal behaviour specialists rather than adapted from a generic asset, and a four-mode structure that takes the user through a deliberate learning sequence - from building empathy through the dog's perspective, to recognising escalating warning signs, to consolidating that knowledge explicitly, and finally, practising calm, positive interaction.

Method

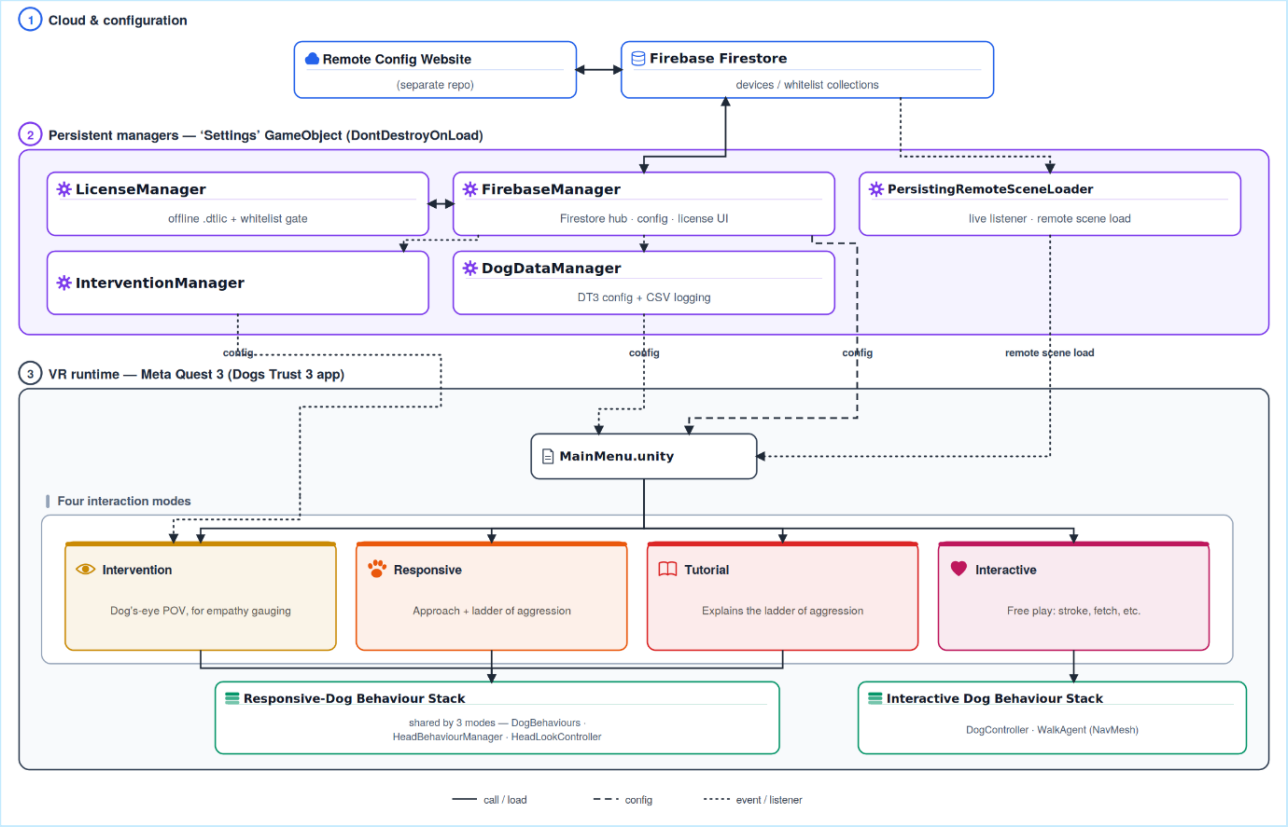


Figure 1. *System architecture, structured in three layers governing configuration, persistence and delivery. Administrators can update device access, licensing and scene setup via a website that writes to a Firestore database, without redeploying to headsets. Persistent manager singletons handle configuration, licence validation and data logging.*

Once configured, the application presents four interaction modes: Intervention, Responsive, Tutorial and Interactive (Figure 2). The first three form the learning sequence described above; the fourth offers a positive interaction with a dog showing no signs of aggression. The first three share a 'Responsive Dog' Behaviour Stack governing head and body behaviour; Interactive mode uses a separate stack with a 'free play' walking agent.

In Intervention mode, perspective taking allows the user to experience the environment from the dog's point of view, including simulated canine vision and hearing, with the aim of building empathy (Herrera et al., 2018). In Responsive mode, set in a park or home, the dog progresses through the ladder of aggression as the user approaches, displaying lip licking and turning away to growling and snapping (Shepherd, 2002), and ignoring these warning signs results in a virtual bite.

Tutorial mode explains the dog's actions and motivations across the same ladder of aggression through audiovisual content. In Interactive mode, the user may play fetch, pet, and prompt it to roll over for a belly rub, intended as a safe space for positive interaction, including for people who may have been bitten.



Figure 2. *The four DAVE interaction modes: Intervention (dog's-eye perspective), Responsive (ladder of aggression), Tutorial (explanatory content) and Interactive (free play).*

Evaluation

As a research tool, an earlier version was evaluated in a pilot study in which 16 university students approached an aggressive or non-reactive dog; participants stayed significantly further from the aggressive dog (although three were ‘bitten’) and perceived it as realistic (Oxley et al., 2022a). DAVE was also used to measure the effect of training: after teaching on canine behaviour, 40 veterinary students stopped sooner when approaching the aggressive dog in video form (Oxley et al., 2024). Most recently, a randomised controlled study with 47 adults recruited from council offices, a delivery office and the university used the current version's Intervention and Tutorial modes as interventions and Responsive mode as the outcome task (Pagano et al., 2026). The tutorial increased confidence in recognising aggression and avoiding bites, with weak evidence that participants stopped further from the dog than controls (around 48 cm on average); perspective-taking showed little to no effect. Participants with higher trait imagination (fantasy) scores approached the dog more closely, an effect the tutorial appeared to reduce in exploratory analysis. A scoping review also found that the behavioural fidelity of virtual dogs in prior studies was rarely justified against canine ethology (Oxley et al., 2022b), a gap DAVE’s co-design process addresses. Ethical approval is reported in the original studies.

Discussion

In the latest study, the tutorial mainly taught warning signs and safe behaviour, whereas perspective-taking conveyed how the situation felt to the dog (Pagano et al., 2026), supporting the use of the modes in sequence; adding a voice-over that explains the dog’s inner state may also strengthen the Intervention mode.

Limitations remain. Only the most recent study used the current build, so earlier findings may not transfer directly. Samples have been small and adult, and most participants in the latest study were current or former dog owners, so effectiveness for children remains untested, and it is unknown whether behaviour towards a virtual dog reflects behaviour around real dogs or is retained over time. In Intervention mode the user controls only the dog's head, and embodiment was moderate (Pagano et al., 2026). Finally, the performance trade-offs and the use of a single breed may limit realism and generalisation to other dogs, although a Labrador was chosen deliberately as a common breed that is not generally perceived as aggressive.

Choosing the right technology also matters, as added immersion does not guarantee better learning (Makransky et al., 2019); in the right context, however, immersive technology can offer an experience closer to real life than any other medium, and DAVE is a strong example of this - a multisensory learning experience achievable only through VR.

Acknowledgements

This research was funded by Dogs Trust Canine Welfare Grants.

References

Shen, J., Rouse, J., Godbole, M., Wells, H.L., Boppana, S., & Schwebel, D.C. (2017). Systematic review: interventions to educate children about dog safety and prevent pediatric dog-bite injuries: a meta-analytic review. *Journal of Pediatric Psychology*, 42(7), pp.779-791. https://doi.org/10.1093/jpepsy/jsv164

Oxley, J.A., Meyer, G., Cant, I., Bellantuono, G.M., Butcher, M., Levers, A., & Westgarth, C. (2022a). A pilot study investigating human behaviour towards DAVE (Dog Assisted Virtual Environment) and interpretation of non-reactive and aggressive behaviours during a virtual reality exploration task. *PLoS ONE*, 17(9), p.e0274329. https://doi.org/10.1371/journal.pone.0274329

Oxley, J.A., Santa, K., Meyer, G., & Westgarth, C. (2022b). A systematic scoping review of human-dog interactions in virtual and augmented reality: the use of virtual dog models and immersive equipment. *Frontiers in Virtual Reality*, 3, p.782023. https://doi.org/10.3389/frvir.2022.782023

Oxley, J.A., Meyer, G., Butcher, M., Bellantuono, G., Levers, A., & Westgarth, C. (2024). Veterinary students' proximity to and interpretation of a simulated “aggressive” dog before and after training. *Scientific Reports*, 14(1), p.3209.

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: a large-scale comparison of traditional and virtual reality perspective-taking. *PLoS ONE*, 13(10), p.e0204494. https://doi.org/10.1371/journal.pone.0204494

Shepherd, K. (2002). Development of behaviour, social behaviour and communication in dogs. In: *BSAVA Manual of Canine and Feline Behavioural Medicine*. Gloucester: BSAVA, pp.8-20.

Pagano, M.-C., Qadeer, C., Georgiou, C., Rothwell, C., Campion, S., & Westgarth, C. (2026). Virtual reality bites: how intervention type and learner imagination influences dog safety intervention efficacy. [Manuscript under review].

Makransky, G., Terkildsen, T.S., & Mayer, R.E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction, 60, pp.225-236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Posters and Demos

GhostLights: Investigating Fog-of-War Alternatives in Asymmetric Collaborative Mixed Reality

Charles Bailly, Barbara Schiavi and Julien Castet

Immersion, Bordeaux, France

Corresponding author: charles.bailly@immersion.fr

Keywords: Mixed Reality, Asymmetric collaboration, Information visualization

Introduction

Mixed Reality (MR) offers great capabilities for collaboration as it allows groups of users to share virtual and/or mixed content to work together. Nonetheless, collaborating in MR is not necessarily symmetric. Users can be in different locations (Sykownik e al., 2020) or have hardware with heterogeneous capabilities and immersion levels (Bréhault et al., 2020; Drey et al., 2022) (for instance, a tablet for Augmented Reality versus a Virtual Reality Head-Mounted Display). Such cases can also lead to asymmetry of environments (Gilbert et al., 2025).

This work focuses on another aspect: the asymmetry of Information (Harris et al., 2016) occurring when MR users do not have the same access rights to virtual content. This choice was motivated by an industrial scenario around factory production lines represented in MR, as illustrated in Figure 1. According to Ouverson et al. (2021), it corresponds to an asymmetry of Participation (different user roles and levels of control) and an asymmetry of Fidelity (different level of detail about production lines). In this scenario, confidentiality is key as users manipulate sensitive industrial data. Real industrial Digital Twins (DTs) raise significant cybersecurity concerns as continuous synchronization expands the attack surface and could create severe tampering and system disruptions (Allison et al., 2023; Coppolino et al., 2023). Therefore, different user roles and accreditation levels exist. Users must collaborate and take decisions despite not visualizing nor interacting with the same elements.

To investigate this kind of asymmetric collaboration scenario, we took inspiration from the fog of war mechanic (Tryhorn et al., 2023). In strategy video games, unexplored or unseen areas of the game map can be partially or totally obscured from the player's view (Hagelback & Johansson, 2008). Our goal is to investigate the impact of different fog of war approaches for MR content on subjective feeling of factory confidentiality and 2) on mental workload for users. The contributions of this work-in-progress are twofold. First, we propose the design and implementation of GhostLights, a set of fog of war approaches for information sharing in asymmetric MR scenarios based on a World-in-Miniature representation (Danyluk et al., 2021). We then discuss the lessons learned from an initial experimental evaluation of GhostLights.

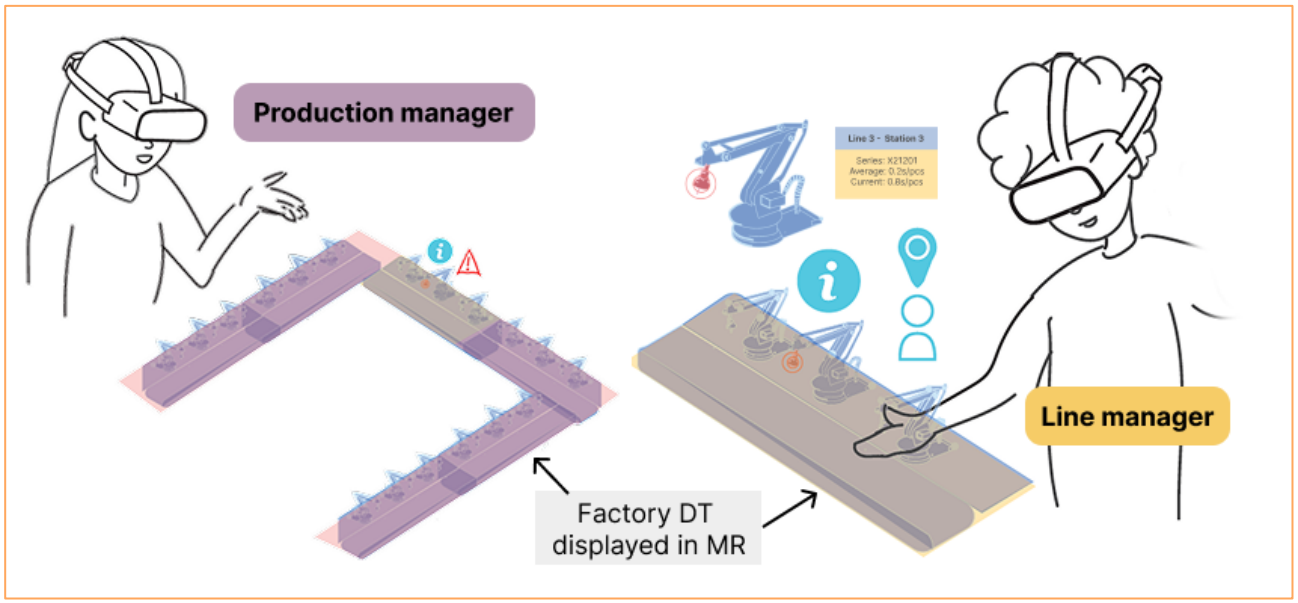


Figure 63. *Overview of the scenario motivating this work: asymmetric MR collaboration around a virtual factory representation. Production managers have elevated rights while Line managers only have limited view of production lines.*

Design of GhostLight techniques

To align with the main concept of fog-of-war, we opted for a common, easy to understand metaphor: using torchlights to reveal hidden objects. By using XR controller as torchlights, the Production manager can point at virtual objects to reveal and share them with other users. The Production manager can also attach filters to each XR controller to affect only the selected category of objects with the torchlight cone. We designed 4 types of visualizations:

StrictOnOff. is the technique prioritizing the most factory confidentiality. Target objects are only shared with others as long as 1) they are currently within the torchlight cone and 2) they match with the corresponding filtering batteries. Otherwise, they are fully hidden for other users, as if they did not exist.

PassiveGhost is based on partially revealed scene objects displayed as transparent "ghost copies". Objects currently affected by a torchlight are fully revealed. However, when outside of the torchlight cone, ghost copies remain visible for all users. Only the outline (location and approximate shape) of targeted objects is shared. Ghost copies are not synchronized with their target object until they are affected again by the torchlight.

ActiveGhost has many similarities with *PassiveGhost*. Nonetheless, the Line Manager can still interact with ghost copies. Any change performed by the Line Manager appears as a suggestion that can be accepted or rejected by the Production Manager. The Line Manager is notified when their suggestions are accepted or rejected.

Finally, a Baseline technique without torchlight was implemented. The Production Manager only had to select each factory object one by one to reveal or hide it.

GhostLights techniques were implemented in Unity (version 2022.3.18) and Netcode for GameObjects (version 1.8). Netcode was used to synchronize the 3D factory scene and its objects, using Remote Procedure Calls (RPCs) to alter the state of factory objects depending on torch interactions. Ghost copies looked like blurred approximations of their original objects by using a custom shader producing a colored deformation wave effect.

Initial experimental study

24 participants (10 female, 14 male) from 16 to 52 years old (*sd = 8.2*) were recruited for the study from nearby companies. Participants passed the experiment as randomized pairs: one playing the role of a Production manager, the other one acting as Line manager. The task was directly inspired by our target scenario. Participants had to identify and correct three failures on a factory DT, represented as a World-in-Miniature factory room in AR. To do so, they had to locate failure locations, buy the corresponding repair supplies and apply them. The Production manager user had to find the balance between efficiently collaborating and protecting the factory confidentiality. On the contrary, some repairing actions could only be performed by the Line manager.

We used a within-subject study design, with 4 randomized blocks (one per technique, Latin-square ordering). Each block consisted of a training phase of 2 tasks, followed by 6 recorded tasks. Overall, each group of participants corrected a total of 96 factory failures (4 blocks of 2 training tasks + 6 recorded tasks each, 3 failures per task). For each task, we recorded the total completion time, number of fixed factory failures, torchlight usage and filters. We also used Raw-TLX questionnaires and conducted semi-structured interviews to gather additional qualitative feedback from participants.

Lessons learned

GhostLights techniques need better mechanisms to avoid unintentional reveals. Overall, the torchlight metaphor used by GhostLights techniques was welcomed by participants. It was judged easy to understand, fluid and enjoyable. The major drawback of GhostLights seems to be their risk of unintentional reveals of other objects. Most participants preferred sharing objects one by one to avoid accidental reveals, even if it was perceived as slower or more cumbersome.

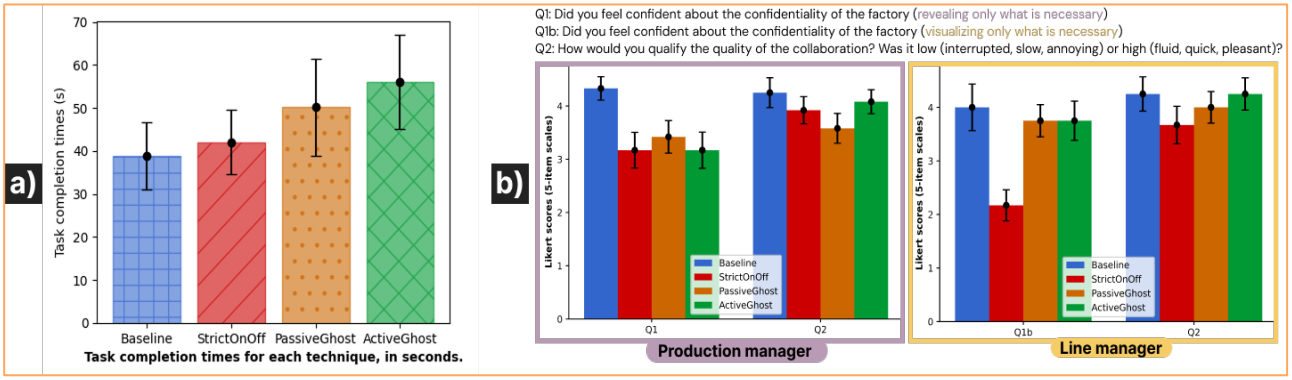


Figure 64: a) Average task completion times per technique. b) Subjective feeling of factory confidentiality (Q1 and Q1b) and perceived quality of the collaboration (Q2) with the different techniques. Reported using 5-item Likert scales. In both figures, error bars are 95% CIs.

Ghost copies offer an interesting compromise. Techniques with ghost copies offered new strategies and collaboration possibilities for participants. Besides, these techniques lead to a better feeling of factory confidentiality than *StrictOnOff.* We hypothesize that Line managers perceived any additional revelation with *StrictOnOff* as unintentional and critical. Meanwhile, they may have been a bit more relaxed about *PassiveGhost* and *ActiveGhost* since it included ghost copies anyway. Additional revelations could thus be intentional in this case. Moreover, the additional validation mechanism from *ActiveGhost* may also help Line managers feel that Production managers are in control and can limit errors by refusing suggestions. This mechanism may nonetheless be a double-edged sword as it was strongly disapproved by a few participants. We thus recommend designers to carefully consider their targeted users and context before integrating such an approach.

As future work, our goal is to further explore the potential of ghost copies while allowing a finer control over the 3D selection of highlighted objects. Our second envisioned study will also target a more realistic and complex factory environment to get closer to ecological conditions.

Acknowledgment

SAFE-6G project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101139031.

References

Allison, D., Smith, P., & Mclaughlin, K. (2023). Digital Twin-Enhanced Incident Response for Cyber-Physical Systems. In *Proceedings of the 18th International Conference on Availability, Reliability and Security*, in ARES '23. New York, NY, USA: Association for Computing Machinery, Aug. 2023, pp. 1–10.

Bréhault, V., Dubois, E., Prouzeau, A., & Serrano, M. 2025). A Systematic Literature Review to Characterize Asymmetric Interaction in Collaborative Systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, in CHI '25. New York, NY, USA: Association for Computing Machinery, Apr. 2025, pp. 1–19. https://doi.org/10.1145/3706598.3713129.

Coppolino, L., Nardone, R., Petruolo, A., Romano, L., & Souvent, A. (2023). Exploiting Digital Twin technology for Cybersecurity Monitoring in Smart Grids. In *Proceedings of the 18th International Conference on Availability, Reliability and Security*, in ARES '23. New York, NY, USA: Association for Computing Machinery, Aug. 2023, pp. 1–10. https://doi.org/10.1145/3600160.3605043.

Danyluk, K., Ens, B., Jenny, B.,. & Willett, W. (2021). A Design Space Exploration of Worlds in Miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, in CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, pp. 1–15. https://doi.org/10.1145/3411764.3445098

Drey, T., Albus, P., Der Kinderen, S., Milo, M., Segschneider, T., Chanzab, L., ... & Rukzio, E. (2022, April). Towards collaborative learning in virtual reality: A comparison of co-located symmetric and asymmetric pair-learning. In *Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1-19). https://doi.org/10.1145/3491102.3517641.

Gilbert, D., Bose, A., Kuhlen, T. W., & Weissker, T. (2025). Pascal-a collaboration technique between non-collocated avatars in large collaborative virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, *31*(5), 3525-3535. https://doi.org/10.1109/TVCG.2025.3549175.

Hagelback. J.., & Johansson, S. J. (2008). Dealing with fog of war in a Real Time Strategy game environment. In *2008 IEEE Symposium On Computational Intelligence and Games*, Dec. 2008, pp. 55–62. https://doi.org/10.1109/CIG.2008.5035621.

Harris, J., Hancock, M., & Scott, S. D . (2016). Leveraging Asymmetries in Multiplayer Games: Investigating Design Elements of Interdependent Play. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, Austin Texas USA: ACM, Oct. 2016, pp. 350–361. https://doi.org/10.1145/2967934.2968113.

Ouverson K. M., & Gilbert, S. B. (2021). A Composite Framework of Co-located Asymmetric Virtual Reality', *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, pp. 1–20, Apr. 2021, https://doi.org/10.1145/3449079.

Sykownik, P., Emmerich, K., & Masuch, M. (2020) Like in the Good Old Times, but Virtual - A Case for Simulating Co-located Multiplayer Games in VR. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*, Virtual Event Canada: ACM, Nov. 2020, pp. 379–383. https://doi.org/10.1145/3383668.3419885.

Tryhorn, D., Dill, R., Hodson, D. D., Grimaila, M.R., & Myers, C. W. (2023). Modeling fog of war effects in AFSIM. *J. Def. Model. Simul.*, vol. 20, no. 2, pp. 131–146, Apr. 2023, https://doi.org/10.1177/15485129211041963.

.

Navigating XR Standardization: A Living, Multidimensional Standards Register for XR and Virtual Worlds

Christoph Runde[1,2], Arjun Rai Gupta[1,2], Saahil Khanna[1]

[1]Virtual Dimension Center (VDC) w.V. | [2]StandICT

Corresponding author: christoph.runde@vdc-fellbach.de

Keywords: Extended Reality (XR); Virtual Worlds; Standardization; Standards Register; Standards Classification

Problem Context and Motivation

Extended Reality (XR) and Virtual Worlds are developing into cross-sector digital infrastructures for design, manufacturing, training, healthcare, public services and collaboration. Their global deployment depends on heterogeneous devices, software, networks, data models, identities and digital assets working together. This creates recurring challenges around interoperability, portability, data exchange, security, privacy, accessibility, safety and long-term compatibility across products, platforms and organizational boundaries.

Standardization is essential for addressing these challenges. Standards define common interfaces, data formats, terminology, reference architectures, measurement procedures and quality requirements. They enable independently developed components to interoperate, reduce integration uncertainty, support interchangeability and market access, and protect investment decisions. For industry and public-sector users, open and internationally coordinated standards can also reduce vendor dependency and contribute to digital sovereignty.

XR standardization is particularly difficult to navigate because XR is a cross-cutting technology rather than a single technical domain. Relevant work is distributed across many Standards Development Organizations (SDOs), consortia and working groups and often originates in adjacent fields such as networking, media coding, geospatial information, cybersecurity, digital identity, artificial intelligence and human-computer interaction. The resulting question is how this heterogeneous and continuously changing ecosystem can be made searchable and navigable for different stakeholder needs.

State of the Art, Research Gap and Methodology

Numerous standards landscapes and roadmaps already address XR, the Metaverse and Virtual Worlds. They organize the field by technology, SDO, system architecture, application or use case. The ETSI Virtual World Standards Landscape Report, for example, analyzed 912 standards and 354 technical reports across eight Virtual Worlds domains [ETSI, 2025a]. These studies provide valuable snapshots, but their document-based form is not designed for continuous updating, interactive filtering or simultaneous navigation across several classification perspectives.

The research gap is therefore not the absence of landscape studies, but the lack of a persistent infrastructure combining systematic discovery, editorial curation, multidimensional classification and continuous maintenance. A single hierarchy is insufficient because the same standard can be relevant to several technologies, applications, industries and methodological or governance topics at once.

Within this work, standards discovery follows the iterative procedure established in the ETSI landscape report. New standards, technical reports, working groups and SDOs are identified through SDO websites, existing landscapes, industry or consortium registers, and normative or informative references in already identified XR standards documents. Related publications and organizational activities are then followed, while metadata are reviewed, completed and normalized before feeding into the register [ETSI, 2025a].

For classification, eight existing classification approaches were analyzed for their structure, strengths and limitations: the VDC taxonomy, ViWISSO, EUOS, NATO XR profiles, CyberLAEND, the EU Rolling Plan for ICT Standardization, the Metaverse Standards Forum, and the XRSI model. The resulting schema combines top-down synthesis with bottom-up validation against the actual register and allows non-exclusive relevance scoring across complementary dimensions. The classification methodology is documented in an internal TWG working report [StandICT.eu TWG Virtual Worlds, 2026a]. Human editorial review remains part of every registration and correction cycle.

Implementation and Results

The XR Standards Register is implemented as a continuously maintained online knowledge base rather than as a static catalogue. The August 2026 snapshot contains more than 2,500 combined records covering standards and related documents, SDOs and Standards Working Groups (SWGs). Besides formal standards, the register includes specifications, technical reports, recommendations, guidelines, white papers and other resources with material relevance to XR and Virtual Worlds [VDC, 2026].

Each record follows an editorial lifecycle with the states: new, in preparation, published, re-check and reject. Editors screen submissions for scope and duplication, complete and verify metadata before publication, and return published entries to re-check when a document, status or classification changes. Rejected records are retained internally for traceability. This separates public information from preliminary or uncertain records while supporting continuous maintenance.

The classification layer (Fig. 1) separates five principal dimensions - Technology, Application Fields, Industries/Verticals, Society/Market/Policy and Methodologies - with additional dimension profiles possible in parallel (D6: NATO profiles; D7-D8: StandICT EUOS profiles).

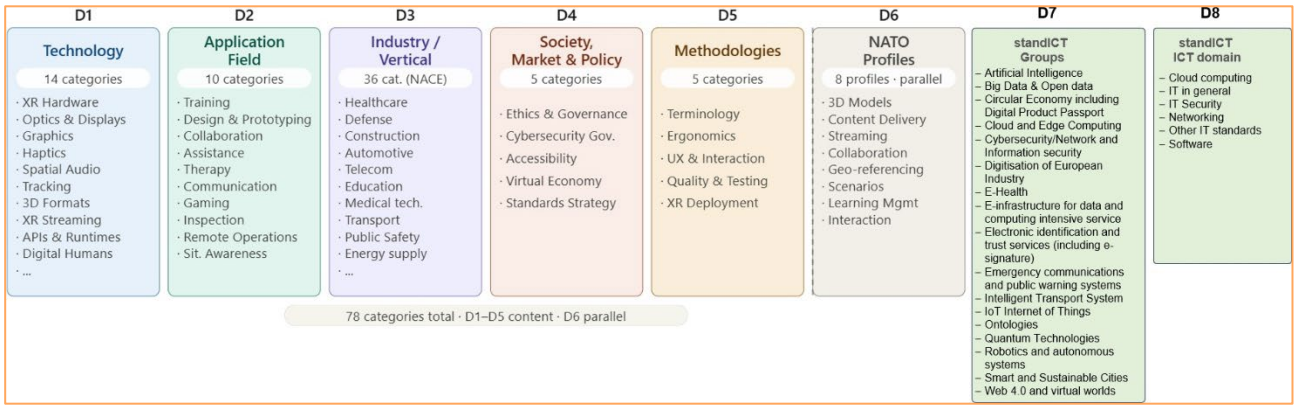


***Figure 1**. The Standards Register's Overall Classification System*

The implementation uses 78 category scores between 0.0 and 1.0. An AI-assisted pipeline evaluates metadata and descriptive information and returns relevance scores plus reasoning, allowing multi-label classification to scale while retaining expert oversight.

For the August 2026 analytical snapshot, 1,919 source records were classified; a separate non-redundant catalogue layer contains 1,858 unique entries. Search and filter functions combine metadata such as type, status and issuing organization with multidimensional classification tags. The same knowledge base also serves as the empirical basis for an ongoing Virtual Worlds Standards Landscape analysis within the StandICT.eu Technical Working Group Virtual Worlds, currently documented as an internal draft [StandICT.eu TWG Virtual Worlds, 2026b].

Evaluation, Impact and Outlook

The register is evaluated as a living knowledge infrastructure (Fig. 2). Its scope, metadata, relevance assessments and classifications are continuously reviewed with experts from the StandICT.eu Technical Working Group Virtual Worlds. Missing standards, questionable inclusions, outdated metadata and misclassifications are fed back into the re-check workflow, complemented by technical consistency checks in the automated classification pipeline.

This recurring review provides continuous quality control, but it does not imply permanent completeness or a formally measured semantic classification accuracy. The ecosystem changes continuously and source metadata vary in quality. A quantitative assessment of classification accuracy and a user-centered usability study therefore remain useful future validation steps.

The impact extends beyond individual searches. Developers and industrial users can identify relevant standards; researchers can analyze standardization structures; and SDOs, policymakers and coordination bodies can identify activity clusters, overlaps and potential gaps. This directly addresses the ETSI recommendation to improve the accessibility and transparency of existing standards and projects through better registries [ETSI, 2025b].

Operational maintenance and analytical use reinforce each other. Findings from landscape analyses and expert reviews feed back into the register, improving coverage, metadata and classification over time. Reusing the same evidence base avoids rebuilding isolated datasets for each study and supports subsequent gap analyses, coordination activities and standardisation roadmaps.

Current development at VDC extends the register towards an AI-supported XR standardization knowledge system. A retrieval-augmented generation (RAG) environment is being built to enable natural-language queries over available standards with answers grounded in the underlying source documents. In parallel, an agent-based monitoring system is being developed to identify new standards and detect changes in publication status, editions, prices and organizational metadata.

The architecture remains human-in-the-loop: agents support discovery and monitoring, the RAG system supports access to document content, and experts retain editorial responsibility.

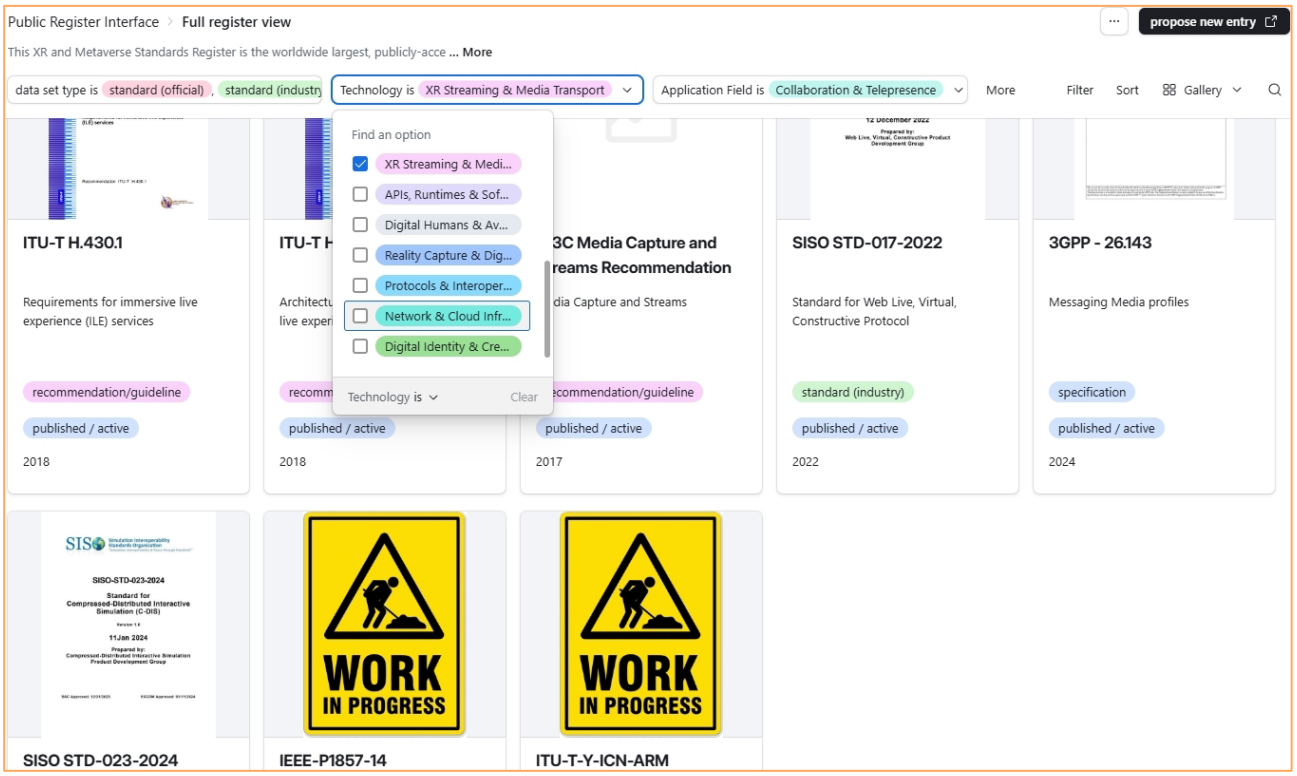


Figure 2. *User Interface of the Current XR and Virtual Worlds Standards Register*

References

ETSI. (2025a). Augmented Reality Framework (ARF); Virtual World Standards Landscape Report. *ETSI GR ARF 010 V1.1.1. Sophia Antipolis: ETSI*. https://www.etsi.org/deliver/etsi_gr/ARF/001_099/010/02.01.01_60/gr_ARF010v020101p.pdf

ETSI. (2025b). Augmented Reality Framework (ARF); Executive Summary of the Virtual World Standards Landscape Report and the Virtual World Standards Recommendations Report. *ETSI GR ARF 012 V1.1.1. Sophia Antipolis: ETSI.* https://www.etsi.org/deliver/etsi_gr/ARF/001_099/012/02.01.01_60/gr_ARF012v020101p.pdf

StandICT.eu TWG Virtual Worlds. (2026a). Development of a Multi-Dimensional Classification System for XR, Metaverse, Virtual Worlds and Immersive Technology Standards. *Internal working report*, v2.0, May 2026 (unpublished internal).

StandICT.eu TWG Virtual Worlds. (2026b.) Virtual Worlds Standards Landscape Report. *Internal draft, v28*, August 2026 (unpublished; work in progress).

VDC. (2026). XR Standards Register. *Online knowledge base*, accessed Aug. 2026. https://www.vdc-fellbach.de/infrastruktur-assets/xr-metaverse-normenregister

From Concept to Deployment – A Physical AR Workstation for Washing Machine Disassembly, Inspection and Component Recovery

Mario Lorenz[1], Sebastian Knopp[1] and Arwa Own[1]

Chemnitz University of Technology, Professorship for Production Systems and Processes, Reichenhainer Straße 70, 09126 Chemnitz, Germany

Corresponding author: mario.lorenz@mb.tu-chemnitz.de

Keywords: Augmented reality; circular economy; disassembly; component recovery; ergonomics; inspection

Motivation

The transition from linear manufacturing toward circular production increases the importance of recovering reusable components from returned products. Disassembly is particularly challenging in this context because, unlike assembly, returned products can differ substantially in condition even when they belong to the same product model. Augmented Reality (AR) can support operators by providing task-related information directly during manual processes, and AR-based instructions have been extensively investigated for industrial assembly, training, and maintenance applications (Wang et al., 2022). Applications for systematic disassembly and component recovery, however, remain less established.

Within the Horizon Europe projects DiCiM (https://www.dicimproject.eu/) and CREDIT (https://www.creditproject.eu/), an AR-AI-(Artificial Intelligence)-workstation was therefore conceived to support the recovery of components from returned washing machines and inspect used components. An earlier publication presented the initial concept of this station, including the planned mechanical lift-and-tilt mechanism, monitor-based AR guidance, and the overall concept for AI-supported validation (Lorenz et al., 2025). At that stage, the physical structure was still under development and completion of the station, integration of the required hardware, and subsequent user evaluation were outstanding.

The present work reports the next development stage: the physical workstation has been completed, integrated, deployed at the industrial use-case partner Gorenje, and is currently undergoing practical testing (Figure 1). The focus of this work-in-progress contribution is therefore the physical design of the station, the handling of the washing machine throughout disassembly, and the resulting operator workflow.

Workstation Design for Industrial Disassembly

The workstation was designed around the physical requirements of washing-machine disassembly and the need to access components from above, the sides, rear, front, and underside without repeatedly repositioning the appliance manually. The frame measures approximately 1.70 m in width, 2.30 m in height, and 1.50 m in depth. Two electric linear actuators lift the complete washing machine by up to 1 m and allow it to be tilted by up to 45°. This creates defined working positions for

different disassembly phases and reduces prolonged work close to floor level or in awkward postures. An integrated lifting aid supports the final removal and transfer of the washing unit. The station incorporates reinforced anti-tip structures, mechanical movement limits, emergency-stop functions, position sensors, appliance restraints, and shielding of moving parts. These measures maintain stable and controlled handling while the machine is raised or tilted. To extend the workstation from component recovery toward reuse-oriented processing, two inspection modules are positioned adjacent to the disassembly area. A height-adjustable table with controlled illumination and an Intel RealSense camera supports inspection of recovered printed circuit boards, while a dedicated washing-unit fixture enables tightness and bearing-condition checks using pressure and vibration sensing. The modules are arranged so that recovered components can be transferred from disassembly to inspection without leaving the workstation context. The monitor-based AR system links these physical stations at workflow level by providing the corresponding disassembly and inspection steps and allowing results to be recorded. Detailed inspection methods and algorithms are outside the scope of this contribution; here, the inspection extensions are considered primarily as part of the integrated physical workstation and material-handling concept within the industrial workflow.

Disassembly and Handling Procedure

The resulting procedure combines the disassembly sequence with deliberate changes in the physical position of the washing machine. The washing machine is first placed in the workstation in its normal upright position. The operator initially removes the upper cover and components that can be conveniently accessed from above and from the front or rear. The lift function can then be used to adapt the working height to the task and operator. For components located underneath the machine, the complete washing machine is raised and subsequently tilted to approximately 45°. This provides access to the lower region while the appliance remains mechanically supported by the workstation. The operator can therefore work from a standing position rather than manually tipping the appliance or performing prolonged work close to floor level. After completion of the lower disassembly steps, the washing machine is returned to the upright position. Remaining front-accessible components can then be recovered. During the final stages, the integrated lifting aid supports removal of the washing unit from the appliance and movement of the recovered assembly away from the central working area. The physical configuration therefore does more than accommodate the AR technology: machine positioning becomes an integral part of the digitally supported disassembly procedure. The workstation provides defined physical states, upright, raised, and raised-and-tilted, that correspond to different groups of disassembly operations.

AR Support and Current Deployment

The physical station is complemented by a monitor-based AR guidance system and inspection system. The software components for stepwise disassembly guidance, inspection and task support have been implemented. Multiple camera views can be incorporated into the workstation, while a touchscreen monitor provides the operator with the instructions required for the current disassembly step. (Lorenz et al., 2025; Own et al., 2025) The detailed software architecture is outside the scope of the present contribution. Instead, the relevant aspect from a workstation perspective is that the digital guidance and the physical manipulation of the appliance form one combined process: instructions can be associated not only with component-removal activities and component inspections but also with the required physical configuration of the washing machine and workstation. Following development and integration, the station has now been deployed at Gorenje and entered the practical testing phase. This transition from laboratory development to an industrial environment is particularly important for evaluating whether the dimensions, accessibility, movement sequences, and physical interaction concept remain practical under actual disassembly conditions.

Preliminary User Feedback and Work in Progress

At the current stage, two participants (N = 2) have interacted with the deployed station and provided initial qualitative feedback. The preliminary feedback regarding the overall user experience has been positive, particularly concerning the general handling concept and the support offered by the workstation during the disassembly procedure. These observations must be interpreted strictly as early qualitative feedback rather than as a formal usability study. The current sample is too small to derive quantitative or generalisable conclusions regarding usability, ergonomics, workload, or process efficiency. The ongoing evaluation therefore focuses on the complete workstation workflow under practical conditions. Of particular interest are the physical accessibility of components in the different machine positions, the usefulness of the lifting and tilting functions, the handling of recovered components, interaction between physical workstation operations and digital instructions, and potential modifications required after longer-term use.

Conclusion

The deployed AR workstation integrates lifting, 45° tilting, safe multi-directional access, and component handling with monitor-based AR guidance for washing machine disassembly. Initial qualitative feedback from two users is positive. Ongoing testing will evaluate usability, ergonomics, workflow integration, and practical suitability for circular-economy component recovery in industrial operation at Gorenje facilities.

Acknowledgement

The authors acknowledge financial support from the European Union's Horizon Europe research and innovation program under the grant agreement No. 101091536 (DiCiM project) and No.101138182 (CREDIT project). The authors used ChatGPT (OpenAI) for language editing and manuscript preparation support. All content was reviewed and approved by the authors, who take full responsibility for it.



Figure 65. *AR workstation for washing-machine component recovery an inspection.*

References

Lorenz, M., Okoniewski, M., Own, A., & Knopp, S. (2025). Industrial Augmented Reality – Concept for an AR-AI-Station for Component Recovery from Washing Machines for Circular Economy. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 1384–1385). IEEE. https://doi.org/10.1109/VRW66409.2025.00338

Own, A., Knopp, S., & Lorenz, M. (2025). Industrial Augmented Reality – Feasibility of AI-based Validation of AR Instructed Component Recovery for Washing Machines in Circular Economy. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 1382–1383). IEEE. https://doi.org/10.1109/VRW66409.2025.00337

Wang, Z., Bai, X., Zhang, S., Billinghurst, M., He, W., Wang, P., Lan, W., Min, H., & Chen, Y. (2022). A comprehensive review of augmented reality-based instruction in manual assembly, training and repair. *Robotics and Computer-Integrated Manufacturing*, *78*, 102407. https://doi.org/10.1016/j.rcim.2022.102407

Appearance Control via Environmental Light Modulation: Performance Limits with a Transmissive LCD

Toshiyuki Amano[1], and Ayu Shigeta[2]

[1]*Graduate School of Systems Engineering, Wakayama University, Japan*
[2]*Analytics Platform Division, Panasonic Digital Co., Ltd, Japan*

Corresponding author: amano@wakayama-u.ac.jp

Keywords: Spatial Augmented Reality, Environmental Light Modulation, Transmissive LCD

Introduction

Spatial Augmented Reality (SAR), widely known through projection mapping, enables the direct manipulation and enhancement of real-world appearances. Its applications include virtual restoration of cultural heritage, visual assistance, and design support, etc. However, most SAR techniques rely on additive light projection, which increases the radiance of the target surface. As a result, it remains difficult to create a natural perceptual experience in which the object itself appears to have changed.

To address this limitation, previous studies have reconstructed environmental illumination using multiple projectors (Takeuchi, 2024) or generated pseudo-environmental illumination using mirror reflections (Yasui, 2024). Other approaches use environmental light itself as a light source for projection or display in bright environments (Yamamoto, 2018; Koizumi, 2013). However, these methods either require specialized configurations or leave uncontrolled environmental light incident on the target, limiting the visibility of appearance manipulation.

Our previous work proposed a more direct approach that modulates incident environmental light using a transmissive liquid crystal display (LCD), inspired by stained glass, where transmitted light is controlled through selective attenuation (Shigeta, 2025). While that work demonstrated the feasibility of the framework under several illumination conditions, this paper focuses on its performance limits. We quantitatively analyze how illuminance ratio, spatial frequency, and propagation distance affect the visibility of appearance manipulation.

Principle and Performance Measures of Environmental Light Modulation

The proposed method controls the appearance of real objects by modulating environmental light with a transmissive LCD panel, as shown in Fig. 1. Each LCD pixel controls the amount of transmitted light, and the modulated environmental illumination reaches the target surface. Let $\mathbf{P} \in [0,1]^3$ denote the RGB pixel value displayed on the LCD, and let $\tau_{\max}$ and $\tau_{\min}$ denote the maximum and minimum transmittance of the LCD. If the incident illumination is $\boldsymbol{I}_a$, the modulated illumination is expressed as $\boldsymbol{I}_{pL}(\boldsymbol{P}, \boldsymbol{I}_a) = (\tau_{\max} - \tau_{\min})\boldsymbol{P} + T_{\min}\boldsymbol{I}_a$. The illuminance ratio is defined as $T = \tau_{\max} / \tau_{\min}$, where $\tau_{\max}$ and $\tau_{\min}$ are measured using a white reference board. A larger T indicates that LCD can produce stronger contrast between bright and dark transmitted states. The observed appearance is

determined by the modulated illumination and the surface reflectance. Let $K = diag(k_R, k_G, k_B)$ denote the surface reflectance and $\boldsymbol{I}_e$ denote indirect environmental illumination reaching the target. The observed intensity is modeled as $\boldsymbol{I}_c = K(\boldsymbol{I}_{pL} + \boldsymbol{I}_e)$.

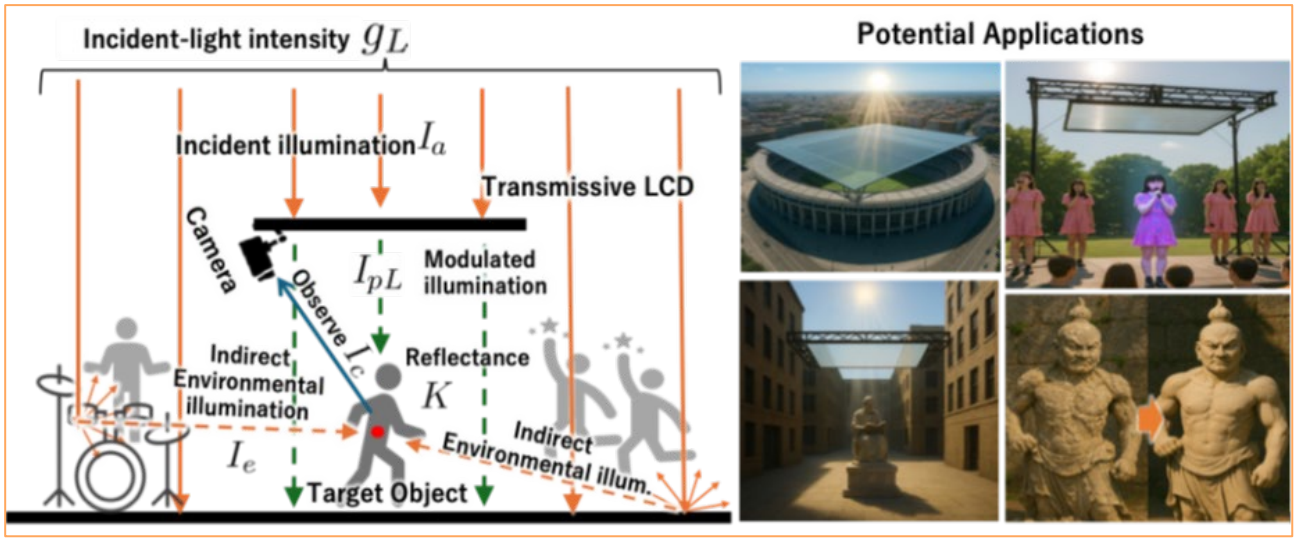


Figure 66. *Concept of Environmental light modulation and its potential applications.*

Using a pre-calibrated correspondence between the camera image and the LCD, the system estimates the reflectance $\hat{K} = \boldsymbol{C}./(M\boldsymbol{P} + \boldsymbol{C}_0)$ from the captured image and the known LCD pattern where $\mathbf{C}$ is the captured RGB value, M is the color-mixing matrix between the LCD and the camera, $\boldsymbol{C}_0$ is the captured image in the minimum-transmission state, and ./ denotes element-wise division. In the figure, g_L denotes a global change in incident-light intensity. Because this gain scales both the captured image and the estimated appearance, the feedback process can compensate for moderate brightness changes. The estimated appearance based on $\hat{K}$ is processed to generate a target image, and the LCD pattern is iteratively updated so that the observed appearance approaches the target. This feedback process follows the appearance-control framework [Amano and Kato 2010], replacing additive projector illumination with environmental light modulation.

To evaluate the spatial performance, we measure the spatial frequency response (SFR) by displaying stripe patterns with spatial frequency f and measuring the contrast ratio R on the target surface. The measured SFR is fitted by $R(f,T) = (R_{\max}(T) - 1)\exp(-af^k) + 1$. This empirical model describes the monotonic decay of contrast with increasing spatial frequency. The $R_{\max}(T)$ represents the maximum achievable contrast under illuminance ratio T, and a and k are effective fitting parameters.

Experimental Results

We first evaluated the proposed method under outdoor sunlight to examine how indifect environmental illumination affects appearance manipulation. In the preliminary experiment shown in Fig. 2(a), the transmissive LCD was placed 40 cm from the target object, and direct sunlight on the ground marked by (a) was approximately 90,000 lx. The transmitted illuminance under the LCD marked by b was 9,000 lx for the maximum-transmission state and 8,300 lx for the minimum-transmission state, resulting in an illuminance ratio of $T = 1.08$. Under this condition, we observed almost no visible appearance change due to the strong indirect environmental illumination. Quantitative evaluation of the hue shift manipulations were only $2.2°$ and $0.5°$ for target shifts of $-28.7°$ and $+28.7°$ in median values, respectively.

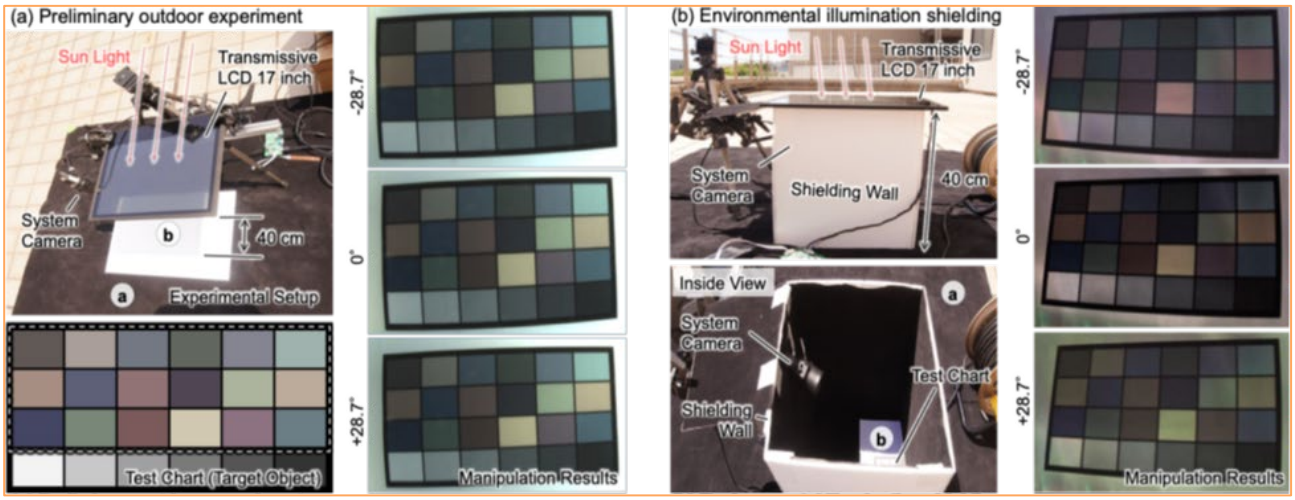


Figure 2. *Outdoor experimental setup and hue-manipulation results with and without environmental-light.*

We then introduced shielding walls to suppress indirect environmental illumination around the target, as shown in Fig. 2(b). In this controlled outdoor condition, the transmitted illuminance marked by b became 3,000 lx for the maximum-transmission state and 1,500 lx for the minimum-transmission state, increasing the illuminance ratio to $T = 2.0$. As a result, the measured median hue shifts were $-33.0°$ and $27.0°$, which were close to the target shifts of $-28.7°$ and $+28.7°$. These results indicate that appearance manipulation becomes feasible when indirect environmental illumination is sufficiently suppressed.

Quantitative Evaluation of Operational Performance

To quantitatively evaluate the operational performance of the system, we measured the spatial frequency response (SFR) in an indoor controlled environmental illumination setup. Since the proposed method displays an attenuation pattern on the LCD, the spatial frequency corresponds to the frequency of the projected transmittance pattern on the target surface. Stripe patterns with spatial frequencies from 0.018 to 1.754 cycles/cm were displayed on the transmissive LCD, and the resulting contrast ratio was measured on a diffuse white target placed 60 cm from the LCD. The illuminance ratio T was varied from 1.09 to 22.

As shown in Fig. 3, higher illuminance ratios preserved higher contrast, while contrast decreased with increasing spatial frequency. Using $R = 3$ (shown with the red dashed line), based on ISO 9241-303 visual performance guidelines, as a practical threshold, the maximum usable spatial frequency under the best condition of $T = 22$ was approximately f_{max}=0.35 cycles/cm. This indicates that environmental-light modulation is most suitable for low-frequency appearance changes.

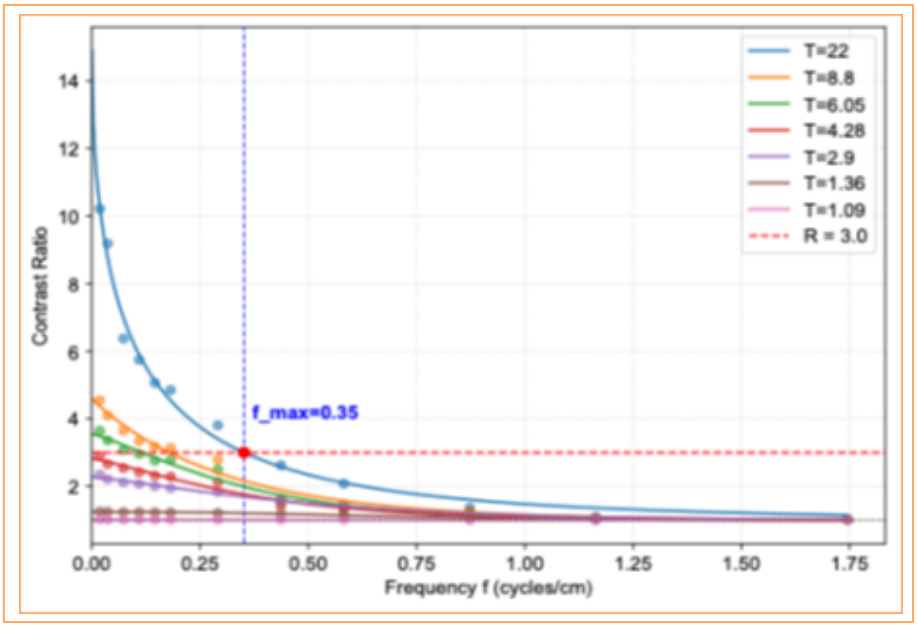


Figure 3. *Spatial frequency response under varying indirect environmental illumination levels.*

Conclusion

This paper evaluated an appearance-control method based on environmental light modulation using a transmissive LCD. Outdoor experiments showed that hue manipulation was ineffective under strong indirect environmental light, but became feasible with shielding for blocking it. SFR measurements showed that the higher illuminance ratio preserved contrast, while high-frequency patterns were attenuated, suggesting that the method is promising for low-frequency appearance changes such as color and materiality alternation under strong sunlight.

Acknowledgement

This research was supported by JSPS KAKENHI Grant Numbers JP24K02979, JP25H01886.

References

Takeuchi, M., Kusuyama, H., Iwai, D. & Sato, K. (2024). Projection mapping under environmental lighting by replacing room lights with heterogeneous projectors, *IEEE Transactions on Visualization and Computer Graphics*, 30(5), 2151-2161. https://doi.org/10.1109/TVCG.2024.3372031

Yasui, M., Iwataki, R., Ishikawa, M. & Watanabe, Y. (2024). Projection mapping with a brightly lit surrounding using a mixed light field approach, IEEE Transactions on Visualization and Computer Graphics, early access. https://doi.org/10.1109/TVCG.2024.3372132

Yamamoto, K., Omomo, K., Takazawa, K. & Ochiai, Y. (2018), Solar Projector, ACM SIGGRAPH 2018 Posters, Article no. 49.

Koizumi, N. (2013). Sunny Day Display: Mid-air image formed by solar light, ACM SIGGRAPH Asia 2013 Emerging Technologies.

Shigeta, A. & Amano, T. (2025). Investigation of Environmental Light Blocking Appearance Manipulation. *Proceedings of the International Display Workshops*, pp. 1091-1094.

Amano, T. & Kato, H. (2010). Appearance control using projection with model predictive control, *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pp. 2832-2835. https://doi.org/10.1109/ICPR.2010.694

LIMIAR: A VR Decision Simulator for Residential Care Professionals

João Matiello

LIMIAR, Berlin, Germany

Corresponding author: joao@constanslab.com

Keywords: Virtual reality; decision simulation; residential care; reflective practice; self-guided debriefing

Residential care professionals often face situations in which there is no single correct response. LIMIAR is a standalone virtual reality decision simulator for practising judgement in those moments. The playable Meta Quest 3 prototype contains S01, “The Wrong Family”, a branching scenario about a communication failure involving a resident’s family. Participants make decisions, see their consequences and complete five structured reflection questions in the headset. The experience provides no score or model answer; its purpose is to help participants examine the priorities and assumptions behind their choices. LIMIAR currently runs as a local, self-guided experience. A planned extension will use the scenario path and reflection responses as limited context for an optional, scenario-bounded AI conversation. This demo submission presents the playable decision-and-reflection journey and its development roadmap.

Context and Aim

Residential care work often requires professionals to balance a resident’s dignity, a family’s expectations, incomplete information and operational limits. Immersive simulation provides a repeatable setting for practice, but the experience still needs to be examined and connected back to work (Fanning and Gaba, 2007; Kolb, 1984). LIMIAR therefore links each decision and consequence to structured reflection. The design aim is a complete standalone journey that can be completed without a facilitator: act, see what follows and reconsider the reasoning behind the choice.

Current Prototype

The playable prototype was built in Unity 6 with the Universal Render Pipeline for Meta Quest 3. A PICO build is in development as a planned second headset target. S01 “The Wrong Family” places the participant in a residential care setting after sensitive information has been shared with the wrong family. The task is not to recall a procedure, but to decide how to respond to those involved.

Decision points alter the sequence and character responses. After seeing the consequence, the participant completes five reflection questions. S01 currently runs locally as a self-contained decision-and-reflection experience on the headset.



Figure 1. *Two views from the playable S01 beta on Meta Quest 3: an interactive decision point (left) and a resident-care interaction (right).*

Experience and Debrief Design

The five single-select questions ask what the participant noticed, prioritised and reconsidered. They form the playable self-guided debrief. A later version may use these answers and the scenario path as limited context for an optional AI-guided conversation.

1. Scenario and decisions	2. Consequence	3. Five reflection prompts	4. AI-guided conversation	5. Session close
PLAYABLE	PLAYABLE	PLAYABLE	PLANNED EXTENSION	PLAYABLE

Figure 2. *Playable journey and optional planned extension. The session can close without step 4.*

Technical Approach and Planned AI Extension

The playable core remains local. The planned AI layer would add an orchestration service between the headset and a hosted language model. It would send only the scenario identifier, decision path and five selected responses under a scenario-specific prompt. The model would ask one question at a time. It would not score the participant, certify competence or provide clinical advice.

Table 1. *Development status and intended conference demonstration.*

Component	Status at submission	EuroXR demo target
S01 branching scenario and consequences	Playable beta	Stable Quest 3 demo build
Five-question self-guided debrief	Playable	Included in core demo
PICO headset build	In development	Targeted for conference demo, subject to validation
Hosted LLM conversation	Planned extension	Optional prototype after privacy and reliability gate

Demo and Evaluation Plan

The EuroXR demonstration is planned as an 8–12 minute standalone VR experience. A participant follows a shortened path through S01, makes a decision, sees the consequence and completes the reflection questions. The current reference build runs on Meta Quest 3, while PICO support is in development and is targeted for the conference demo if validation is completed in time. An external screen can mirror the headset view. The demonstration will focus on the playable decision, consequence and reflection loop.

A formative pilot with a residential care provider is scheduled for 7 October 2026. Care professionals and operational stakeholders will assess interaction clarity, scenario credibility, completion with limited assistance and the relevance of the reflection questions. The pilot takes place after the camera-ready deadline, so this submission does not report results.

Safety, Privacy and Limitations

LIMIAR is designed for professional learning and reflective practice rather than clinical decision support. The current build operates locally without a hosted language model. Before the planned AI layer is introduced, its data flow, retention model and service providers will be defined and evaluated. The October pilot will provide initial evidence on usability, perceived relevance and the value of the decision-and-reflection approach in a real care training context.

Expected Contribution

LIMIAR demonstrates a practical model for connecting immersive decision-making, visible consequences and structured reflection within a standalone VR experience. The planned AI layer extends this model while preserving the decision-and-reflection loop as the core of the experience.

References

Fanning, R. M. & Gaba, D. M. (2007). The role of debriefing in simulation-based learning. Simulation in Healthcare 2(2), 115–125. https://doi.org/10.1097/SIH.0b013e3180315539

Kolb, D. A. (1984). Experiential Learning: Experience as the Source of Learning and Development. *Englewood Cliffs*, NJ: Prentice-Hall.

Liu, K., Zhang, W., Li, W., Wang, T. & Zheng, Y. (2023). Effectiveness of virtual reality in nursing education: a systematic review and meta-analysis. *BMC Medical Education* 23, 710. https://doi.org/10.1186/s12909-023-04662-x

Influence of Material Surfaces on the Quality of Spatial Maps Generated by the Microsoft HoloLens 2

Maximilian Putz, Sandra Böhm, Sven Winkler, Franziska Klimant, and Martin Dix

Chemnitz University of Technology, Institute for Machine Tools and Production Processes, Chemnitz, Germany

Corresponding author: maximilian.putz@mb.tu-chemnitz.de

Keywords: Microsoft HoloLens 2, Spatial Mapping, Accessibility

Introduction

Spatial mapping is a core prerequisite for positioning virtual content and tracking the real-world environment around the user in Augmented and Mixed Reality (AR/MR) applications. The Microsoft HoloLens 2 derives its depth data from an active Time-of-Flight (ToF) sensor that emits modulated infrared signals. This means reconstruction quality strongly depends on surface properties and lighting conditions. This study is set in the HoloRolo project at Chemnitz University of Technology, where spatial mapping data detects room collisions in real time and calculates paths to automatically navigate an electric wheelchair. For the project, the accuracy of the spatial map is directly related to the safety of the application. While reflective and transparent materials are known to cause faulty depth data, systematic comparisons of material surfaces for the HoloLens 2 remain limited. We therefore assess how material surfaces and lighting conditions affect spatial mapping quality, aiming to guide the furnishing of accessible homes for HoloLens-assisted or similar navigation systems.

Related Work

Non-systematic ToF depth errors are attributed in the literature primarily to signal-to-noise ratio, multipath reflections, light scattering, and motion blur. Using a MESA SR4000 ToF camera against a laser-scanner reference, He et al. found a near-linear increase of depth error with distance, a small lighting effect, and material-dependent standard deviations (He, et al. 2017). The measured deviations were roughly 1.5 mm for wood versus 13.4 mm for metal. Hopkins et al. also found glass unreliably detected by ToF sensors due to specular transmission/reflection (Hopkins, et al. 2025). For the HoloLens 2 specifically, Teruggi and Fassi found centimeter-level deviations and characteristic mesh defects (“holes,” “hallucinations” and “bias”) scanning Milan Cathedral (Teruggi and Fassi 2022). Hou et al. reported mean deviations of 3 – 4 cm over a 30 m indoor corridor and Navares-Vázquez et al. reported a mean deviation of 4.7 cm over 110 m (Hou, et al. 2024) (Navares-Vázquez, et al. 2025). None of these studies isolated the effect of individual material surfaces for the HoloLens 2 under otherwise controlled conditions.

Method

Three materials with different reflective and optical properties were applied to a table (1.6 × 0.8 m, height 0.71 m) and a stool (Ø 0.39 m, height 0.50 m). The material selection consisted of wood

(diffuse, dielectric), aluminum foil (specular, metallic), and cling film (semi-transparent, dielectric, with diffuse backscatter from the underlying wood). Objects were placed centrally in a laboratory room of about 7 × 7 m under daylight (overcast, no direct sun) and artificial light (fluorescent tubes).



Figure 67. *Photo of the test scene. From left to right: wood, cling film, aluminum foil*

Data acquisition followed a fixed rectangular two-way route with six floor markers, a sensor-to-object distance of ≈ 0.7 m, and a metronome-paced head-movement sequence (horizontal, vertical, circular) at each stop. Six main scans (3 materials with 2 lighting conditions each) were recorded. The scenario wood/daylight was repeated three times to assess reproducibility. These measurements showed minimal deviation, which led to conducting only one scan per scenario.

All six scanned scenarios were compared against a 3D-modeled ground-truth mesh using CloudCompare. Each scan and the ground-truth model were resampled into point clouds of 1,000,000 points. For each point, the distance between the ground-truth model and the closest scenario scan point was determined.

Results

Table 10 *Quantitative cloud-to-mesh distance (cm) results of the six main scans*

Scan	Mean dev. [cm]	SD [cm]	Max dev. [cm]
Wood, daylight	3.20	1.82	11.91
Wood, artificial light	2.66	1.80	10.45
Aluminum foil, daylight	29.53	23.04	72.43
Aluminum foil, artificial light	32.12	23.56	77.45
Cling film, daylight	2.56	2.04	13.77
Cling film, artificial light	2.68	1.64	11.44

Wood produced consistently low deviations under both lighting conditions, with largely complete reconstructions of table, stool, and floor and residual deviations concentrated at edges. Cling film performed comparably, retaining the general object geometry with slightly rougher surfaces and less defined edges than wood. For both materials, over 99 % of points deviate by less than 10 cm (see **Figure 2**, **Figure 3**).

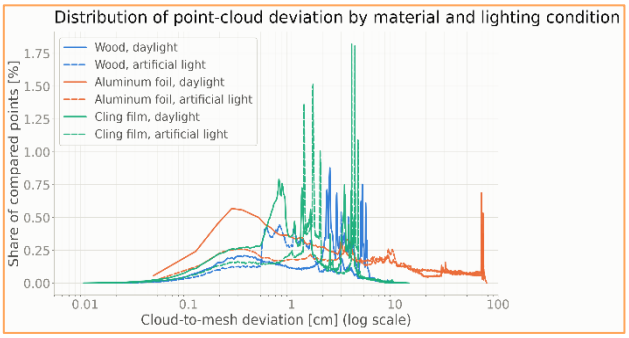


Figure 68. *Distribution of cloud-to-mesh deviation per scan scenario, as percentage share of the 1,000,000 compared points (log scale).*

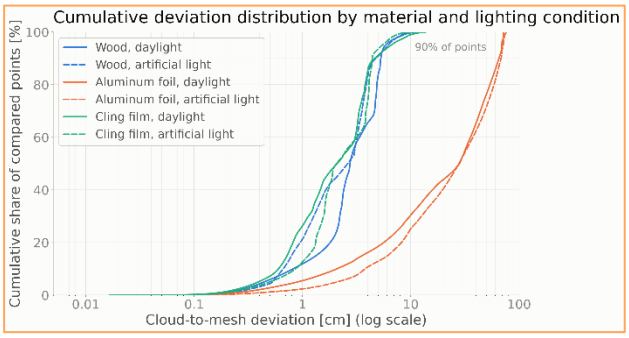


Figure 69. *Cumulative distribution of cloud-to-mesh deviation per scan scenario in cm (log scale)*

Aluminum foil caused near-total reconstruction failure. Mean deviation rose to over 30 cm with a maximum of up to 77.5 cm. 28.34% of points deviating by more than 50 cm. The coated table surface was largely absent from the mesh, with floor geometry substituted in its place and only fragments of the stool captured. The reconstruction was visibly unstable during acquisition.

Discussion

Diffuse reflection scatters the IR signal in many directions, ensuring a usable return, whereas strong specular reflection redirects it away from the sensor, producing the “holes,” “hallucinations” and “bias” errors described by Teruggi et al. and leading to the substantial accuracy loss seen in Table 1 (Teruggi and Fassi 2022). Cling film's result is attributed to partial backscatter from its diffuse wood substrate. Lighting condition had a comparatively small effect relative to material. Mean deviation shifted by less than 0.6 cm for wood and cling film, and by about 2.6 cm for aluminum foil. This is plausibly because the ToF sensor evaluates its own modulated signal and suppresses unmodulated ambient light and only moderate indoor lighting was tested instead of direct sunlight.

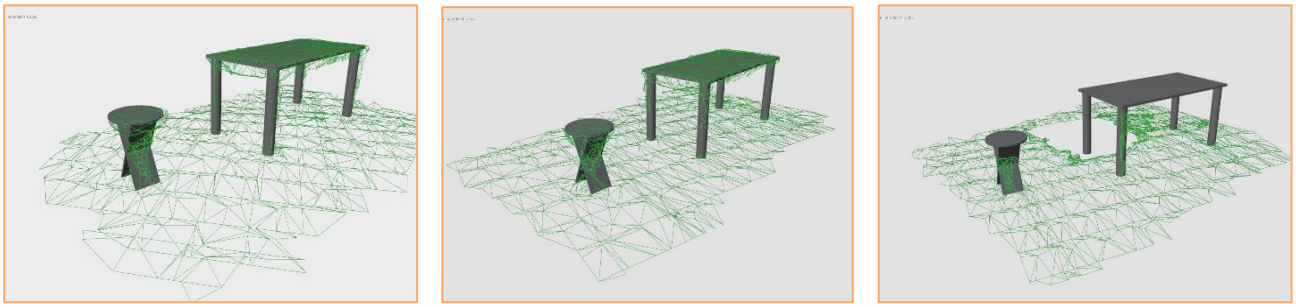


Figure 70. *Scan result (green) vs. ground-truth mesh (grey). From left to right: wood, cling film, aluminum foil.*

Conclusion and Outlook

The data of this study show a significant influence of surface materials on the HoloLens 2 spatial map quality. Diffusely reflecting wood and semi-transparent cling film (with a diffuse backing) yield comparable, reliable reconstructions, with 99.7% of points deviating by less than 10 cm. Strongly specular aluminum foil causes severe failure, with 28,34% of points deviating by more than 50 cm and a maximum approaching 80 cm. These results directly carry over to practical guidance for furnishing accessible homes for use with the HoloLens 2 navigation system. Objects like mirrors and similar reflective or highly glossy metallic surfaces should be mindfully placed and adjusted in order to guarantee safety. Future work includes applying these results to realistic apartment scenarios to

test and compare how the identified problematic surfaces can be set up to reduce spatial mapping failures.

Acknowledgements

This project (InnoTeam HoloRolo) is co-funded by the European Union and co-financed from tax revenues on the basis of the budget adopted by the Saxon State Parliament.



Co-funded by
the European Union



This project is co-financed from tax revenues on the basis of the budget adopted by the Saxon State Parliament.

References

He, Y., Liang, B., Zou, Y., He, J., & Yang, J. (2017). Depth errors analysis and correction for time-of-flight (ToF) cameras. *Sensors*, *17*(1), 92. https://doi.org/10.3390/s17010092 .

Hopkins, M., Murali, V., Kumar, V., & Taylor, C. J. (2026, June). Real-time glass detection and reprojection using sensor fusion onboard aerial robots. In *2026 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 14608-14615). IEEE. https://doi.org/10.48550/arXiv.2510.06518.

Hou, J., Hübner, P., Schmidt, J., & Iwaszczuk, D. (2024). Indoor mapping with entertainment devices: Evaluating the impact of different mapping strategies for Microsoft HoloLens 2 and Apple iPhone 14 Pro. *Sensors*, *24*(4), 1062. https://doi.org/10.3390/s24041062.

Navares-Vázquez, J. C., Qiu, Z., Arias, P., & Balado, J. (2025). HoloLens 2 performance analysis for indoor/outdoor 3D mapping. *Journal of Building Engineering*, *108*, 112826.. https://doi.org/10.1016/j.jobe.2025.112826.

Teruggi, S., & Fassi, F. (2022). Hololens 2 spatial mapping capabilities in vast monumental heritage environments. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 46(2), 489-496. https://doi.org/10.5194/isprs-archives-XLVI-2-W1-2022-489-2022.

Toward Reliable Pupillometry in eXtended Reality: A Video-Based Pipeline for Isolating Cognitive Workload from Pupil Data

Giovanni Pignoni, Fabio Grandi, and Margherita Peruzzini

Department of Industrial Engineering, University of Bologna, Viale Risorgimento 2, Bologna, Italy

Corresponding author: giovanni.pignoni@unibo.it

Keywords: Pupillometry, virtual reality, eye tracking, cognitive workload, pupillary light reflex.

Introduction

Eye tracking is now a standard feature of commercial Virtual Reality (VR) and eXtended Reality (XR) headsets, introduced to support foveated rendering, gaze-based interaction and automatic interpupillary distance adjustment (Adhanom et al. 2023). Pupillometry can therefore be deployed in immersive experiments with no additional instrumentation. The task-evoked pupillary response has been validated as an indicator of effort and memory load since the 1960s (Kahneman and Beatty 1966), but the pupillary light reflex dominates the measured signal, with luminance-driven changes spanning 2 to 8 mm against a cognitive component often below 1 mm (Mathôt 2018). That order-of-magnitude difference has confined the measurement to laboratories and tasks with controlled illumination.

In VR the visual stimulus is generated by the display itself, which is both an opportunity and an obstacle: the rendered frame is known exactly, while the mapping from pixel values to emitted luminance is proprietary, light leakage is uncontrolled, and inter-individual variability in the light response remains a confound for which no established calibration procedure exists (Eckert et al. 2022). Recent multi-sensor work in mobile augmented reality reaches a similar conclusion on the difficulty of isolating load from a single ocular channel (Pluisch et al. 2025).

This extended abstract reports work in progress: a post-hoc pipeline that adapts a published field method for luminance-compensated pupillometry (Pignoni et al. 2021) to the constraints of XR headsets, a per-participant calibration procedure, and a pilot on six participants with a Varjo XR-4 assessing its feasibility. The work is framed within the DTRIP4H project (GA 101188432), which develops a decentralised digital twin environment for health research infrastructures together with the VR interfaces used to explore it. Interfaces of this kind pose new challenges to the quantification of an operator's cognitive workload: content is often dense and variable, viewpoint and gaze are free, and the luminance of the rendered field changes with every user action. Although such studies take place in the laboratory, these conditions closely resemble those of a field experiment. This invites the re-development of the tools that once enabled field studies, so that they can now serve advanced laboratory work.

Pipeline

The original method estimates a dynamic baseline representing the pupil diameter expected from luminance alone and interprets the deviation from it as a cognitive workload (CW) proxy. It combines

pupil diameter, gaze position and a point-of-view scene video, with absolute luminance provided by a photometric sensor on the eye-tracker frame and a gaze-centred area of interest segmented by GrabCut. The luminance estimate feeds the Watson and Yellott (2012) unified formula and the expected diameter is subtracted from the filtered measurement. Three elements had to be revised for the headset (Figure 1).

Luminance from the rendered frame

To avoid the challenge of adding external sensors inside a headset, the display itself may become the light source to be characterised. Per-pixel sRGB values recovered from the scene video recorded by Varjo Base are gamma decoded and mapped linearly between the black and white points of the panel, with an independent gain on each channel, before the BT.709 photopic weighting. Six quantities are left free to the operator: the two photometric endpoints, the decoding gamma and the three color channel gains.

Two-area gaze weighting

GrabCut is computationally expensive and unstable on low-contrast detail. Following Eckert et al. (2022), it is replaced by the weighted average of the mean relative luminance inside a circular region centred on the gaze point and that of the remainder of the frame. A 35:65 ratio was adopted for internal testing in a flighsimulator environment and preferred to the 26:74 weighting reported by Eckert et al. for their scenes. The ratio depends on the angular size and contrast of the visual targets and requires adjustment at experiment level. It has no influence on the calibration below, where the field is uniform.

The unified formula describes the pupil in equilibrium. An **optional pre-filter** approximates the transient behaviour: a fixed latency followed by a one-pole recursive filter with asymmetric time constants, so that constriction after a luminance increase is almost immediate while re-dilation is slow (Mathôt 2018). The default constants are first-pass values rather than identified ones.

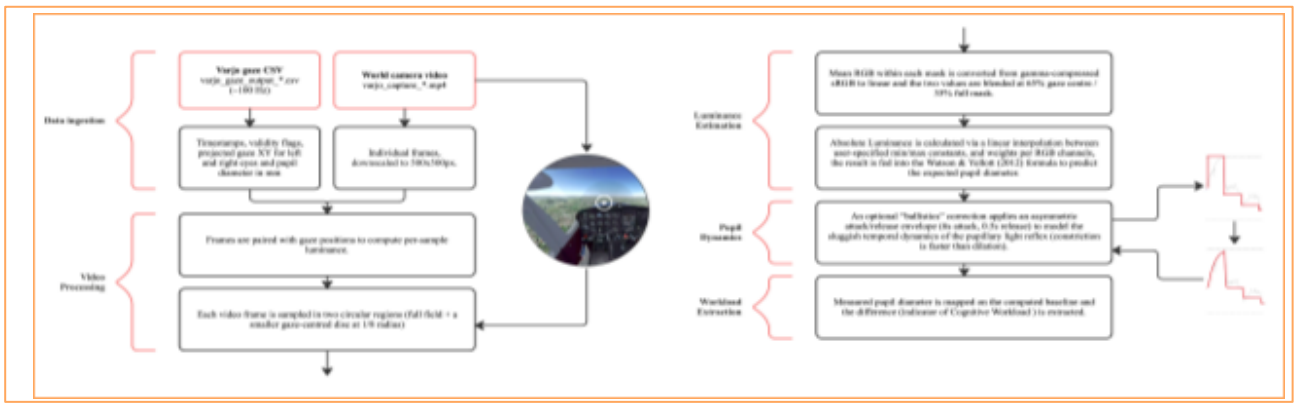


Figure 1. *Simplified pipeline steps.*

Calibration procedure

The six free parameters cannot be fixed a priori, as they absorb the combined effect of headset rendering, the undocumented conversion of pupil size to millimetres performed by the device, and individual physiology. They are per-participant personalisations rather than a generalisable photometric model of the headset.

A calibration environment was developed in Unity on an OpenXR setup. The participant is placed at the centre of an enclosing sphere carrying a double-sided unlit material, whose colour is driven at runtime by a timestamped file, giving abrupt transitions between uniform full-field stimuli. The sequence comprises 20 stimuli of 6 s, for a total of 120 s: an achromatic ramp of eight grey levels uniformly spaced on the 0 to 255 axis, followed by red, green and blue ramps at code values 64, 128, 191 and 255, each block presented from dark to light. The order is fixed and the stimuli are

solid colours, a simplification of the pseudo-random sequences of Eckert et al. (2022) adopted to keep the procedure short and repeatable for manual adjustment.

The operator then compares the measured trace against the expected one and adjusts the photometric endpoints on the achromatic ramp first, followed by the channel gains on the chromatic ramps, iterating until the two curves match to a degree judged acceptable. The resulting parameter set is stored and reused for every recording of the same participant.

Pilot results

Seven participants were recruited within the laboratory; one recording was discarded for insufficient signal quality, leaving six valid cases. Because the calibration sequence contains no workload manipulation, the root-mean-square of the delta pupil diameter over the sequence (CW rms) measures uncompensated luminance error, and a lower value indicates a tighter fit between model and recorded response.

At default settings the mean CW rms was 0.49 mm (SD 0.11). Adjusting the photometric endpoints alone brought it to 0.43 mm (SD 0.12), a reduction of about 13%. Adding per-channel compensation brought it to 0.27 mm (SD 0.10), a total reduction of about 44%, with final values between 0.15 and 0.39 mm and an improvement observed for every participant. The default gamma of 2.2 was retained throughout, indicating that the sRGB non-linearity describes the transfer function adequately within this sample. The endpoints and the gains, in contrast, varied substantially: the white point spanned 1500 to 6500 cd/m² (mean 4250, SD 2208), the green gain stayed close to unity (mean 1.08), the red gain took intermediate values (mean 3.83) and the blue gain was markedly elevated (mean 11.33, SD 5.71). The last result follows from the BT.709 weighting, in which the blue primary carries 7% of the photopic luminance signal, so that a large gain is needed for an equivalent pupillary effect and any difference in short-wavelength sensitivity is amplified.

Discussion and next steps

The reduction is meaningful against the signal to be detected, as the task-evoked response to moderate load is reported in the 0.1 to 0.5 mm range, but the upper bound of the residual error after calibration approaches that range. The pilot therefore shows that the exposed parameters can remove a large part of a dominant confound and that a 120 s procedure is sufficient to do so consistently; it does not show that workload estimation from the residual is reliable in VR. The limitations are explicit. Six participants support a feasibility claim only, not an account of inter-individual variability. The manual fit is operator dependent and inter-operator agreement has not been quantified. The fixed ascending order limits carry-over from the preceding level but does not remove it. The channel gains act as static scalings, whereas wavelength-dependent responses differ in time course as well as in amplitude, a distinction the present formulation does not capture. Work in progress addresses the automated identification of both the photometric and the dynamic parameters from the same 120 s recording, which would remove operator dependence and allow randomised sequences, and the application of the calibrated pipeline to a task-based dataset with a workload manipulation of known direction, as already demonstrated for the field version of the method (Pignoni et al. 2021).

Integration in the DTRIP4H project

Use Case 4 addresses artificial tumour niches for animal-free cancer prognosis and drug screening. Its VR interface spans three scales: the macro scale of whole-brain MRI segmentation and radiomics, the micro scale of cancer core growth and surface expression, and the nano scale of molecular dynamics trajectories. This interface act as navigation and exploration client for complex data, enabling multiscale exploration for teachers and students. UC4 will provide the first application testbed for this work: the calibrated pipeline will be deployed there to assess the cognitive workload of the proposed interaction design, contributing an objective component to the use-case validation.

Acknowledgements.

Funded by the European Union under Grant Agreement No. 101188432 (DTRIP4H). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority. Neither the European Union nor the granting authority can be held responsible for them.

References

Adhanom, I.B., MacNeilage, P. & Folmer, E. (2023). Eye tracking in virtual reality: a broad review of applications and challenges. *Virtual Reality* 27, 1481–1505.

Eckert, M., Robotham, T., Habets, E.A.P. & Rummukainen, O.S. (2022). Pupillary light reflex correction for robust pupillometry in virtual reality. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 5(2), 1–16. https://doi.org/10.1145/3530798

Kahneman, D. & Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154(3756), 1583–1585.

Mathôt, S. (2018). Pupillometry: psychology, physiology, and function. *Journal of Cognition* 1(1), 16. https://doi.org/10.5334/joc.18

Pignoni, G., Komandur, S. & Volden, F. (2021). Accounting for effects of variation in luminance in pupillometry for field measurements of cognitive workload. *IEEE Sensors Journal* 21(5), 6393–6400. https://doi.org/10.1109/JSEN.2020.3038291

Pluisch, M., Gugenheimer, J., Cho, Y., Julier, S. & Kruijff, E. (2025). A multi-sensor approach for cognitive load assessment in mobile augmented reality. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR),* Daejeon, 208–218. https://doi.org/10.1109/ISMAR67309.2025.00033

Watson, A.B. & Yellott, J.I. (2012). A unified formula for light-adapted pupil size. *Journal of Vision* 12(10), 12. https://doi.org/10.1167/12.10.12

Immersive Cultural Heritage Experiences Through an Extended Reality Digital Twin of the Ancient Hydraulis

Georgios Moysiadis, Savvas Koltsakidis, Emmanouil K. Tzimtzimis, and Dimitrios Tzetzis

Digital Manufacturing and Materials Characterization Laboratory, School of Science and Technology, International Hellenic University, 57001 Thessaloniki, Greece

Corresponding author: d.tzetzis@ihu.edu.gr

Keywords: Digital Twin, Extended Reality, Immersive Technologies, Hydraulis Organ, Interactive learning

Abstract

Cultural heritage represents a fundamental connection between past and present, preserving knowledge, technological achievements, artistic expression, and historical identity. However, traditional museum experiences often rely mainly on static displays, limiting the ability of visitors to fully understand the functionality, construction principles, and the individual parts that form a historical artefact (Song & Evans, 2024). Recent developments in digital technologies, particularly Digital Twins and Extended Reality (XR), provide new opportunities to transform the way cultural heritage is documented, interpreted, and experienced (Silva & Teixeira, 2021). This work presents the development and application of a digital twin of the ancient Greek hydraulis, one of the earliest known keyboard instruments and an important milestone in the history of musical technology. The digital representation aims not only to preserve the geometry and appearance of the artefact but also to create an interactive environment where users can explore, understand, and experience the instrument beyond the limitations of a conventional museum exhibition.

The proposed approach is developed within the framework of the Musicsphere programme, which investigates the integration of digital technologies for enhancing musical and cultural heritage experiences. The digital twin acts as a bridge between physical artefact preservation and immersive digital interaction, enabling users to access detailed information, visualize hidden structural elements, and explore the instrument from perspectives that would not be possible through a physical exhibit alone (Dagnav et al., 2026). An overview of the proposed digital twin concept and its integration with XR technologies is presented in Figure 1. Through XR technologies, visitors can interact with the digital hydraulis in an intuitive and engaging manner. Using Virtual Reality (VR), Augmented Reality (AR), or other extended reality environments, users can observe the instrument in three-dimensional space, rotate and examine its components, explore its internal mechanisms, and gain a deeper understanding of its operation.

The development of the digital twin began with the collection and analysis of the available historical and archaeological evidence concerning the construction and operation of the ancient hydraulis. The reconstruction was primarily informed by archaeological findings, including the hydraulis discovered at Dion and documented by Pandermalis, together with historical descriptions of the instrument and its operating principles, particularly those provided by Vitruvius. Since the surviving evidence does not provide a complete engineering definition of the original instrument, the available information was interpreted and combined to reconstruct its geometry and mechanical

arrangement as faithfully as possible. Based on this evidence, a detailed three-dimensional CAD model of the hydraulis was developed, representing both its external form and the internal components required to understand its operating principle.

The resulting CAD model forms the geometric foundation of the digital twin and is directly connected to the physical prototype that will be developed within the Musicsphere project. This common digital representation provides a link between historical documentation, physical reconstruction, and immersive visualization. For museum and educational applications, the model is transferred into an XR environment in which users can inspect the hydraulis from different perspectives and interact with its individual components. In particular, components can be selectively removed or isolated, allowing otherwise hidden internal mechanisms to become visible and help users understand how the different parts of the instrument contribute to its operation. In this way, the digital twin extends beyond a static three-dimensional reconstruction and becomes an interactive tool for the interpretation, communication, and preservation of European musical and technological heritage.

The presented poster demonstrates the digital twin of ancient hydraulis and highlights the potential of XR technologies to redefine cultural heritage experiences. By combining accurate digital reconstruction with immersive interaction, the project explores how emerging technologies can transform museums from passive exhibition spaces into interactive environments of exploration, learning, and engagement.

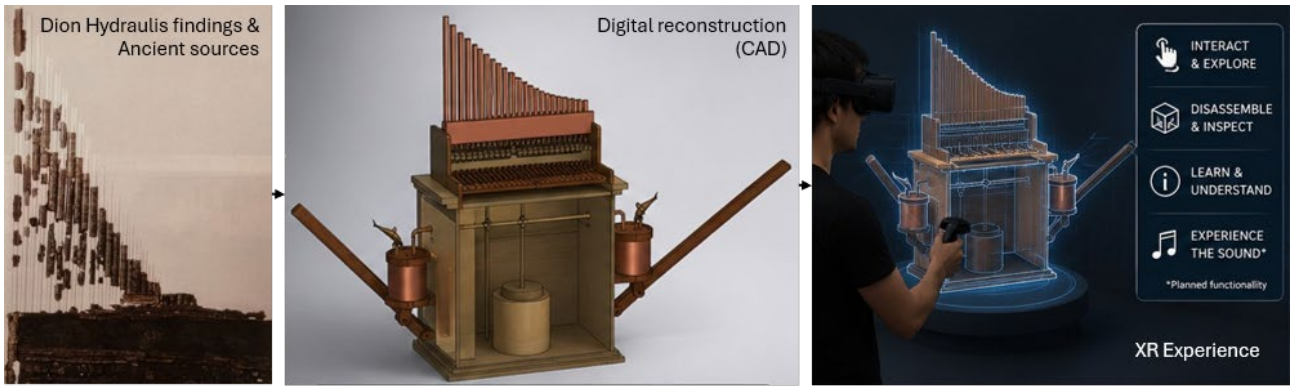


***Figure 1**. Conceptual overview of the digital twin development of the ancient Greek hydraulis From the analysis of archaeological evidence and ancient sources (left) to the CAD reconstruction (center) and its integration into an XR environment (right) that enables immersive exploration, understanding, and interaction with the instrument.*

This research is part of the MusicSphere project that is funded by the European Union under the Horizon Europe Framework Program and the grant agreement N° 101233618. Views and opinions expressed are however, those of the authors only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Dagnaw, G., Capuano, R., & Muccini, H. (2026). Digital twins for cultural heritage: A systematic analysis of the state of the art. ACM Computing Surveys, 58(9), 1-35. https://doi.org/10.1145/3793541

Silva, M., & Teixeira, L. (2021, December). eXtended Reality (XR) experiences in museums for cultural heritage: A systematic review. In International conference on intelligent technologies for interactive entertainment (pp. 58-79). Cham: Springer International Publishing.

Song, Z., & Evans, L. (2024). The museum of digital things: extended reality and museum practices. Frontiers in Virtual Reality, 5, 1396280. https://doi.org/10.3389/frvir.2024.1396280

From Physical Organ Pipes to Immersive Digital Heritage: Reverse Engineering, Digital Libraries, and XR Exploration

Georgios Moysiadis[1], Savvas Koltsakidis[1], Emmanouil K. Tzimtzimis[1], Dimitrios Sagris[2], Constantine David[2], and Dimitrios Tzetzis[1]

[1] Digital Manufacturing and Materials Characterization Laboratory, School of Science and Technology, International Hellenic University, 57001 Thessaloniki, Greece

[2] Manufacturing Technology and Production Systems Laboratory, Department of Mechanical Engineering, School of Engineering, International Hellenic University, 62124 Serres, Greece

Corresponding author: d.tzetzis@ihu.edu.gr

Keywords: Organ pipes, Reverse Engineering, Computed Tomography (CT), Digital Reconstruction, Extended Reality (XR), Digital Heritage

Abstract

Pipe organs represent a unique intersection of musical heritage, craftsmanship, acoustics, and engineering. Their sound is generated through individual pipes whose geometry, dimensions, materials, and construction characteristics directly influence their acoustic behavior (Angster et al., 2017; Rossing & Fletcher, 2021). However, the manufacture of organ pipes has traditionally relied heavily on the empirical knowledge and experience of skilled craftsmen, with limited availability of detailed technical documentation for historical examples. The systematic digitization and documentation of such components can therefore contribute both to the preservation of traditional knowledge and to a better understanding of the relationship between pipe geometry and sound.

Within the framework of the MusicSphere project, this work presents the development of digital reconstructions of organ pipes through a reverse engineering approach, as a first step towards the creation of a comprehensive digital pipe library. Three physical organ pipes were selected and digitized using a combination of manual dimensional measurements and three complementary 3D scanning technologies: the Artec Eva structured-light scanner, the NextEngine Ultra HD laser scanner, and X-ray micro–Computed Tomography (CT) Tomoscope HV Compact 225 kV from Werth Messtechnik GmbH (Gießen, Germany) (Figure 1a–c). The different scanning approaches were evaluated to compare their capabilities and limitations in documenting the complex geometry of the pipes.

The Artec Eva employs structured-light scanning and enables rapid, non-contact acquisition of external surfaces, with a reported 3D point accuracy of up to approximately 0.1 mm and a 3D resolution of up to 0.2 mm. The NextEngine Ultra HD employs laser-based triangulation and provides high-resolution surface digitization, making it suitable for capturing fine external geometrical details. However, both optical techniques are fundamentally limited by line-of-sight accessibility and therefore cannot directly capture enclosed internal surfaces or geometrical features obscured from the scanner. X-ray μCT was consequently employed to overcome this limitation. By reconstructing the object as a three-dimensional volumetric dataset composed of voxels, μCT enables the

simultaneous characterization of both external and internal geometries without requiring physical sectioning of the artefact. For the present measurements, a voxel size of 60μm was achieved, allowing otherwise inaccessible internal features of the pipes to be identified and reconstructed.

Key external and accessible dimensions were additionally measured using precision calipers and subsequently used to support and validate the CAD reconstruction. The final digital models were therefore generated through the complementary use of direct dimensional measurements and 3D scanning data, combining the dimensional reliability of manual measurements and high-resolution external surface acquisition with the ability of μCT imaging to capture complex and inaccessible internal geometrical features. Deviation analysis between the reconstructed geometries and μCT-derived data, demonstrated satisfactory agreement, with mean deviations below 0.3 mm in regions where the original geometry could be reliably captured by the scanning process.

The resulting digital models form the basis of a broader digital library provide dimensional, geometrical, and acoustic information for different types of organ pipes. Rather than functioning solely as a repository of three-dimensional models, the proposed library establish connections between pipe geometry and the corresponding sound characteristics, supporting documentation, comparative analysis, research, and educational applications.

As further development within MusicSphere, Extended Reality (XR) technologies are integrated to transform this digital content into an interactive experience (Silva & Teixera, 2021; Zou et al., 2022). Through Virtual Reality (VR), users are able to inspect and manipulate organ pipes at different scales, explore their geometrical characteristics, and listen to the sounds associated with individual pipe configurations (Figure 1d). By combining reverse engineering, digital documentation, acoustic information, and immersive technologies, the proposed approach aims to preserve specialized organ-building knowledge while providing researchers, musicians, craftsmen, and the wider public with new ways of studying and experiencing this important aspect of musical heritage.

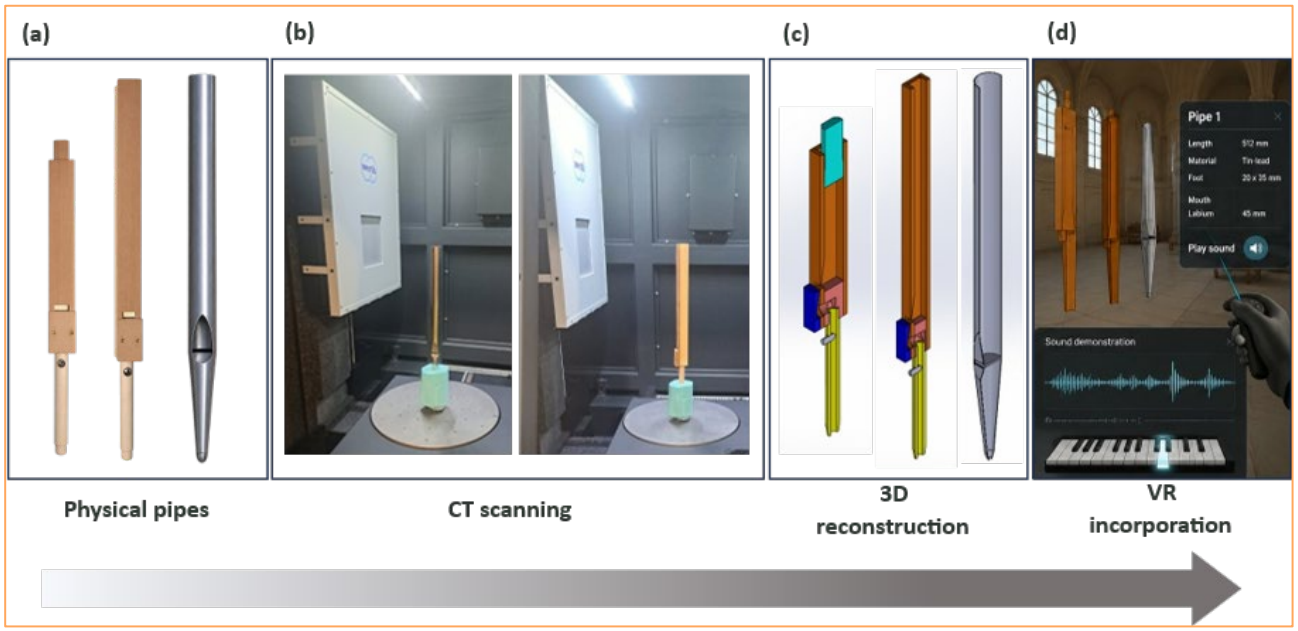


***Figure 1**. Workflow for the digitization and immersive exploration of organ pipes: (a) physical organ pipes, (b) μCT scanning for geometrical acquisition, (c) 3D reconstruction of the digitized pipes, and (d) integration of the digital models into a VR environment for interactive visualization and sound demonstration.*

This research is part of the MusicSphere project that is funded by the European Union under the Horizon Europe Framework Program and the grant agreement N° 101233618. Views and opinions expressed are however, those of the authors only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Angster, J., Rucz, P., & Miklós, A. (2017). Acoustics of organ pipes and future trends in the research. *Acoustics today*, *13*(1), 10-18.

Rossing, T. D., & Fletcher, N. H. (1998). *The physics of musical instruments* (pp. 728-732). Springer.

Silva, M., & Teixeira, L. (2021, December). eXtended Reality (XR) experiences in museums for cultural heritage: A systematic review. In *International conference on intelligent technologies for interactive entertainment* (pp. 58-79). Cham: Springer International Publishing.

Zhou, Y., Chen, J., & Wang, M. (2022). A meta-analytic review on incorporating virtual and augmented reality in museum learning. *Educational Research Review*, *36*, 100454. https://doi.org/10.1016/j.edurev.2022.100454

Development of a 3D Digital Twin for a Filament Winding Industrial Facility: An Immersive User Interface for Industry 4.0

Maria Siafali, Emmanouil K. Tzimtzimis, Savvas Koltsakidis, and Dimitrios Tzetzis

Digital Manufacturing and Materials Characterization Laboratory, School of Science and Technology, International Hellenic University, 57001 Thessaloniki, Greece

Corresponding author: d.tzetzis@ihu.edu.gr

Keywords: Digital Twin, Photogrammetry, Virtual Tour, Industry 4.0, Virtual Reality, UI/UX Design

Abstract

Digital Twins have emerged as a cornerstone technology of Industry 4.0, enabling the virtual representation, monitoring, and management of complex physical assets (Melesse et al., 2020). Facilitated by advanced spatial sensing and cyber-physical links, this technology is increasingly deployed to enhance asset management, remote operations, and situational awareness across industrial environments (Soori et al., 2023). However, traditional facility documentation methods, such as two-dimensional architectural blueprints and static digital photography, often fail to convey the three-dimensional complexity and operational nuances of high-stakes industrial sites, a limitation that is particularly evident in facilities housing specialized equipment where spatial orientation and technical training are critical for safety. This work presents the development of an immersive Digital Twin for a specialized industrial filament winding facility, transitioning the concept from a static 3D model into a dynamic, interactive information portal.

By synthesizing UAV-acquired exterior photogrammetry with high-resolution interior 360° panoramic nodes, a hybrid digital environment was established to provide a 1:1 scale telepresence for remote inspection and safety training. An overview of the development workflow of the digital twin and its integration with an XR interface is presented in Figure 1. Exterior data were captured with a UAV drone flown in a grid pattern across five altitudes and three viewing angles, yielding 388 high-resolution images that were processed in photogrammetry software through alignment, dense reconstruction, and texturing stages to generate an 8K-textured polygonal mesh of approximately 1.1 million triangles. The interior of the facility was documented through eleven strategically placed panoramic nodes, positioned along a line-of-sight navigation path at approximate human eye level. Since geometric modelling remains one of the most resource-intensive stages of the digital twinning process (Agapaki & Brilakin, 2022), this multi-modal acquisition strategy was adopted to balance metrical accuracy with acquisition efficiency across the facility.

The exterior mesh and interior panoramas were integrated within professional multimedia software tool to create interactive 360-degree virtual tour, using a "Portal" architecture, in which the 3D model functions as a volumetric container and high-precision hotspots transport users seamlessly between the macro-scale exterior and the micro-scale interior production zones. The core technical contribution of the study is the design of a bespoke, adaptive User Interface (UI) that acts as a semantic layer over the raw geometric data. An interactive floorplan with a real-time radar indicator

provides continuous spatial orientation across the facility, while a contextual visibility logic automatically hides non-essential interface elements once a Virtual Reality Head-Mounted Display is detected, reducing visual noise during safety-critical training. As shown in related work on virtual industrial and heritage platforms, the usability of such systems depends strongly on functional clarity and scene interactivity (Zhou et al., 2025), principles that informed the hotspot taxonomy developed for this facility.

Beyond static geometry, semantic hotspots embed technical specifications, maintenance logs, and safety protocols directly onto individual machines, while dedicated links provide live access to real-time operational parameters, such as spindle speed, resin temperature, and cycle times, as well as to the operating status of curing ovens and resin storage areas. Embedded 360° video hotspots further replace static panoramas with recordings of the winding process in motion, conveying kinetic and auditory information that a static mesh alone cannot capture. The complete environment was deployed on both desktop and VR (Meta Quest 3) platforms, employing a gaze-and-click interaction technique that improves accessibility and comfort during immersive training, consistent with findings on collaborative virtual reality training systems (Le et al., 2021).

The implementation of a Digital Twin for a high-consequence industrial facility introduces new possibilities for remote management and workforce training. Instead of presenting the facility as a static geometric model, the resulting system provides a dynamic, spatialized database in which stakeholders actively access asset information, monitor live operational data, and rehearse safety procedures. This approach improves accessibility by allowing technical staff, managers, and trainees to navigate the facility and consult localized asset data without entering hazardous production zones. Moreover, the integration of diverse data sources, ranging from 3D meshes and 360° video to live operational data and technical documentation, creates a single source of truth for facility management, supporting future extensions such as Industrial IoT integration and predictive maintenance.

The presented poster demonstrates the Digital Twin of an industrial filament winding facility and highlights the potential of combining photogrammetry, panoramic imaging, and a custom-designed interface to redefine industrial facility management. By combining accurate spatial reconstruction with an adaptive semantic UI, the project explores how emerging XR technologies can transform manufacturing plants from passive exhibition spaces into interactive environments of remote inspection, training, and engagement.

This paper is derived from research conducted as part of the Master's Thesis in the M.Sc. in Strategic Product Design program at the International Hellenic University. The authors gratefully acknowledge the support, guidance, and contribution of the teaching and laboratory staff of the program.

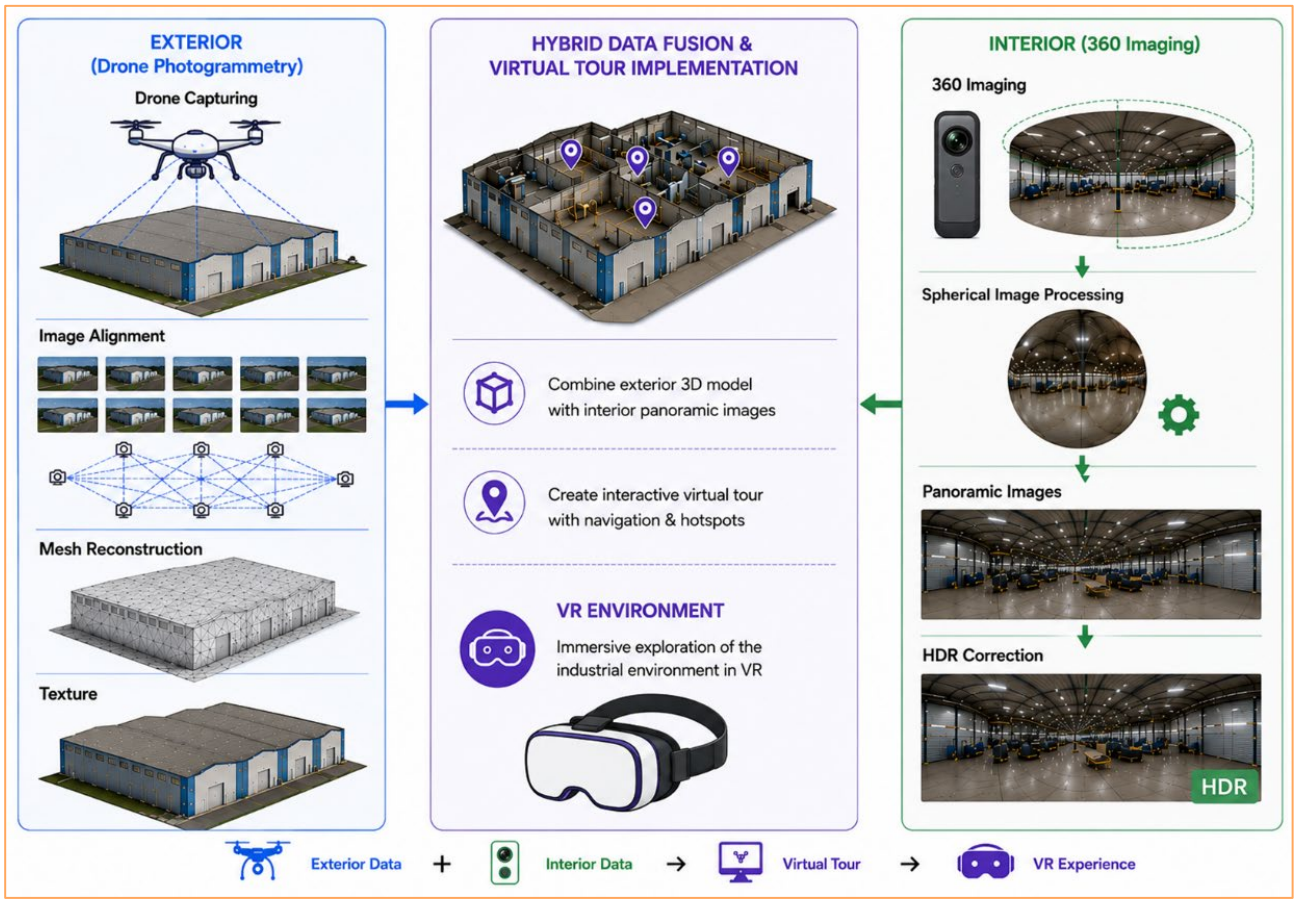


Figure 1. *Integrated development workflow of the industrial Digital Twin, from multi-modal data acquisition (UAV photogrammetry and 360° panoramic imaging) through photogrammetric processing to deployment within the interactive XR interface that enables remote inspection, safety training, and real-time monitoring.*

References

Agapaki, E., & Brilakis, I. (2022). Geometric Digital Twinning of Industrial Facilities: Retrieval of Industrial Shapes. arXiv:2202.04834. https://doi.org/10.48550/arXiv.2202.04834

Le, K.-D., Ly, D.-N., La, T.-T., Nguyen, C., Fjeld, M., Nguyen, T.V., & Tran, M.-T. (2021). Development and Evaluation of a Collaborative Virtual Reality System for Tour Guide Training. *Virtual Reality*, *25*, 911–928.

Melesse, T.Y., Di Pasquale, V., & Riemma, S. (2020). Digital Twin Models in Industrial Operations: A Systematic Literature Review. *Procedia Manufacturing*, *42*, 267–272. https://doi.org/10.1016/j.promfg.2020.02.084

Soori, M., Arezoo, B., & Dastres, R. (2023). Digital Twin for Smart Manufacturing: A Review. Sustainable Manufacturing and Services Economics, 2, 100017. https://doi.org/10.1016/j.smse.2023.100017

Zhou, Q., Wang, S., & Wang, J. (2025). Exploring User Experience in Virtual Industrial Heritage Platforms: Impact of Cultural Identity, Functional Clarity, Scene Interactivity, and Narrative Quality. *Buildings*, *15*, 253. https://doi.org/10.3390/buildings15020253

Exploring Local Dorsolateral Prefrontal Cortex Dynamics During Virtual Navigation Using Virtual Reality and Stereo-Electroencephalography

Teodoro Ors[1], Roberto Llorens[1], Mariano Alcañiz[1], María Palanca[2], Antonio Gutiérrez[2], and Vicente Villanueva[2]

[1]Research in Human-Centered Technology, Universitat Politècnica de València, Valencia, Spain [2]Refractary Epilepsy Unit, Neurology and Neurosurgery Service, Hospital Universitari i Politècnic La Fe, Valencia, Spain

Corresponding author: raorqui@htech.upv.es

Keywords: Virtual navigation, Stereo-electroencephalography, Dorsolateral prefrontal cortex

Introduction

Immersive Virtual Reality (VR) provides a useful experimental framework for studying spatial navigation because active interaction can be preserved while environmental and behavioral variables remain tightly controlled. Navigation research has traditionally emphasized the hippocampus and medial temporal structures, where oscillatory activity is strongly linked to movement and spatial processing (Bush et al., 2017; Watrous et al., 2011). However, successful navigation also requires executive processes such as action monitoring, working memory, behavioral planning, and goal-directed control. These functions involve the dorsolateral prefrontal cortex (DLPFC), a region that is itself structurally and functionally heterogeneous (Jung et al., 2022).

This heterogeneity creates a methodological problem. Scalp electroencephalography commonly represents prefrontal activity using a small number of frontal electrodes and therefore averages signals over broad cortical territories. Functional MRI offers better spatial specificity but lower temporal resolution and restricts natural interaction. Stereo-electroencephalography (sEEG), although available only in patients undergoing invasive presurgical monitoring, records intracranial activity directly from precisely localized contacts with millisecond temporal resolution. Its value here is not simply that it has greater spatial resolution, but that contact-level analysis can reveal local modulation patterns that may disappear when the DLPFC is treated as a single region.

This preliminary study therefore combined immersive VR and sEEG to examine local DLPFC dynamics during virtual navigation. The specific aims were to determine whether spectral activity recorded from individual DLPFC contacts varied as a function of relative navigation speed and whether contacts located within the same anatomical region exhibited distinct spectral modulation profiles. The study was conceived as an exploratory feasibility analysis rather than as a population-level characterization of DLPFC function due to limited spatial coverage.

Materials and Methods

Four patients with drug-resistant epilepsy undergoing sEEG presurgical monitoring participated in the study. Electrode implantation was determined exclusively by clinical requirements, resulting in participant-specific DLPFC coverage. All participants were clinically stable, remained under prescribed antiseizure medication, and provided written informed consent. The study was approved by the Ethics Committee of Hospital Universitari i Politècnic La Fe.

The experimental setup followed the methodological framework for integrating sEEG with immersive VR described by (Alcañiz et al., 2026). Briefly, participants navigated freely through a virtual forest (Figure 1), while seated, using a Meta Quest 2 headset. Each controller-trigger press produced a fixed displacement along a predefined path, while head rotation enabled visual exploration. Neural activity was acquired using a Natus Quantum LTM system, and VR events were temporally synchronized with the intracranial recordings using the photodiode-based triggering system shown only in the running laptop.



Figure 71. *Screenshot of the laptop screen, displaying the virtual forest and the photodiode flashing-square.*

Pre-implantation MRI and post-implantation CT were co-registered in MNI space to localize the electrodes, and only bipolar contacts anatomically assigned to the DLPFC were retained. sEEG signals were resampled to 512 Hz, notch-filtered at 50 and 100 Hz, and segments containing epileptiform activity were excluded after expert review. Navigation activity was normalized relative to the resting baseline using z-scores and segmented into consecutive non-overlapping 3 seconds epochs. Because every trigger press represented a fixed forward displacement, relative navigation speed was estimated from the number of displacement events in each epoch, yielding levels of 0.00, 0.33, 0.67, 1.00, and 1.33 events per s.

Welch PSD was estimated for delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-99 Hz) bands. Contact-level spectral profiles were normalized according to response shape and clustered using candidate solutions from K=2 to K=5. Cluster quality was assessed using silhouette score and inertia, and speed-dependent differences were evaluated using Kruskal–Wallis tests followed by Bonferroni-corrected Mann–Whitney U comparisons stablishing statistical significance at $p<0.05$ after correction.

Results

DLPFC activity was not adequately described by a single regional spectral response. Contact-level analyses revealed distinct modulation trajectories as relative navigation speed increased, indicating coexisting local electrophysiological responses within the sampled DLPFC.

Shape-based clustering supported two recurrent modulation profiles. Among solutions from K=2 to K=5, K=2 yielded the highest silhouette score (0.4590 vs. 0.4477, 0.4154, and 0.3991 for K=3–5), while inertia decreased from 1821.0 at K=2 to 780.3 at K=5. Thus, additional clusters improved within-cluster fit but reduced separation and parsimony, supporting K=2 as the most interpretable solution. Because clustering was performed on normalized profile shape, the groups reflect differences in speed-dependent dynamics rather than absolute PSD magnitude.

The two profiles showed distinct relationships with relative navigation speed. Bonferroni-corrected group-level analyses identified significant within-profile modulation across all five frequency bands. Between-profile comparisons were significant across all bands for the 0–0.67 and 0–1.00 contrasts ($p<0.001$), with particularly persistent differences in beta and gamma: gamma differed in all evaluated speed contrasts ($p\leq0.033$), and beta in five of six.

This divergence is illustrated for gamma in Fig. 2A, while Fig. 2B shows that contacts belonging to both profiles were spatially intermingled within the sampled DLPFC rather than forming clearly segregated anatomical groups. Together, the clustering, statistical, and spatial results support local heterogeneity in navigation-related DLPFC modulation.

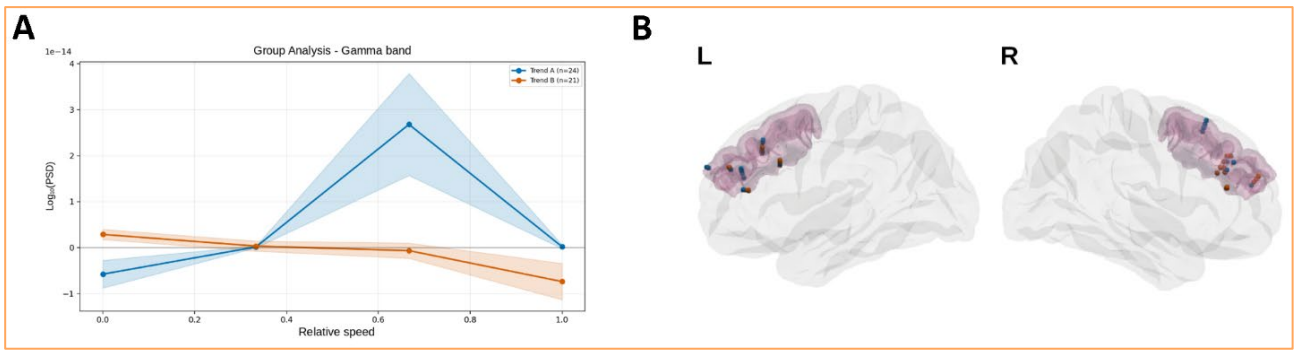


Figure 72 *Local DLPFC modulation profiles during virtual navigation. (A) Group-level gamma-band spectral modulation across relative navigation-speed. (B) Spatial distribution of DLPFC contacts assigned to each gamma-band profile in the left (L) and right (R) hemispheres.*

Discussion and Conclusions

The principal finding of this study is that DLPFC electrophysiological activity during immersive navigation is heterogeneous at the local contact level. Rather than exhibiting a common regional response, contacts within the same anatomically defined DLPFC followed two recurrent spectral modulation profiles as navigation speed changed. The convergence between clustering, corrected statistical comparisons, and the spatial distribution of contacts supports this interpretation: the profiles differed in their speed-dependent dynamics and were not simply separated by absolute spectral power or by clearly segregated anatomical territories.

These findings extend previous evidence of structural and functional heterogeneity within the DLPFC (Jung et al., 2022). Whereas previous work identified differentiated organization across DLPFC subregions, our results reveal heterogeneity at a finer electrophysiological scale: spatially intermingled contacts within the same anatomically defined DLPFC territory exhibited markedly different modulation trajectories during the same VR navigation task. This suggests that DLPFC functional organization may be more locally differentiated than can be captured by conventional subregional parcellations, with distinct neural dynamics coexisting within spatially close cortical sites. Thus, our findings provide intracranial electrophysiological evidence that extends the concept of

DLPFC heterogeneity from large-scale structural and functional subdivisions to local task-dependent dynamics.

The distinction between profiles remained after normalization of absolute PSD magnitude and was supported by between-profile statistical comparisons, indicating that regional averaging could obscure meaningful local dynamics. The stronger differentiation observed in beta and gamma suggests that higher-frequency activity may be particularly sensitive to changes occurring after virtual movement onset, although this interpretation remains exploratory.

The study is limited by the four-patient clinical cohort, participant-specific electrode coverage, epilepsy and antiseizure medication, and the use of controller-derived relative speed rather than physical locomotion. Nevertheless, the recurrence of distinct profiles across participants and frequency bands provides preliminary evidence that navigation-related DLPFC activity is locally heterogeneous and supports the combined use of immersive VR and sEEG to investigate local cortical dynamics.

References

Alcañiz, M., Ors, T., Gutierrez, A., Villanueva, V., & Llorens, R. (2026). Methodological framework for integrating intracranial EEG with immersive virtual reality. *Frontiers in Virtual Reality*, *7*. https://doi.org/10.3389/frvir.2026.1803960

Bush, D., Bisby, J. A., Bird, C. M., Gollwitzer, S., Rodionov, R., Diehl, B., McEvoy, A. W., Walker, M. C., & Burgess, N. (2017). Human hippocampal theta power indicates movement onset and distance travelled. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(46), 12297–12302. https://doi.org/10.1073/pnas.1708716114

Jung, J., Lambon Ralph, M. A., & Jackson, R. L. (2022). Subregions of DLPFC Display Graded yet Distinct Structural and Functional Connectivity. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(15), 3241–3252. https://doi.org/10.1523/JNEUROSCI.1216-21.2022

Watrous, A. J., Fried, I., & Ekstrom, A. D. (2011). Behavioral correlates of human hippocampal delta and theta oscillations during navigation. *Journal of Neurophysiology*, *105*(4), 1747–1755. https://doi.org/10.1152/jn.00921.2010

Perceived Motion Fidelity in Hyper-Realistic VR Avatars: A Preliminary Comparison of IK-Only and IK+MediaPipe Tracking

Maria Eleonora Minissi[1*], Alejandro Trillo[1], Julia Ajenjo Domingo[1], Javier Marín Morales[1], Mariano Alcañiz[1]

Human-Centered Technology Research Institute (HUMAN-Tech), Polytechnic University of Valencia, Valencia, Spain

*Corresponding author: Maria Eleonora Minissi

Keywords: Virtual Reality; Motion Fidelity; Inverse Kinematics; MediaPipe; Hyper-Realistic Avatars

Abstract

Motion fidelity is critical to the credibility of self-avatars in immersive virtual reality (VR), particularly when hyper-realistic appearance increases users' expectations of natural and coherent movement. This preliminary within-subject study compared two upper-body tracking configurations for a seated mirror-view hyper-realistic avatar: an inverse kinematics pipeline driven by headset and hand tracking only (IK-only), and a hybrid pipeline integrating additional webcam-based pose information through MediaPipe (IK+MediaPipe). Ten right-handed participants completed a VR procedure in which they performed predefined upper-body postures, object-oriented arm movements, and spontaneous hand gestures under both conditions. Motion fidelity was evaluated through independent observer ratings and participant self-report measures, including a perceived motion fidelity questionnaire and the Agency domain of the Virtual Embodiment Questionnaire (VEQ). Observer-based fidelity scores showed no statistically significant differences between tracking conditions for any movement. Self-report results revealed a more specific pattern: perceived matching of the avatar's hands and arms during complex gestures was significantly higher in the IK+MediaPipe condition, whereas perceived movement synchrony was significantly higher in the IK-only condition. All remaining self-report comparisons were non-significant. These exploratory findings suggest that adding webcam-based pose estimation may improve perceived anatomical correspondence during complex upper-body configurations, but may also introduce temporal or perceptual costs that reduce perceived synchrony. The results highlight the need to consider tracking accuracy, stability, and movement-specific constraints jointly when designing hyper-realistic self-avatar systems for VR.

Introduction

Embodiment in VR depends not only on seeing a virtual body but on experiencing it as the locus of one's own actions. Body ownership, agency, and self-location are strongly influenced by visuomotor congruence (Slater et al., 2010; kilteni et al., 2012) mismatches in posture, timing, or gesture kinematics can therefore reduce the credibility of a self-avatar. Consumer VR commonly reconstructs body motion from sparse head-and-hand tracking using inverse kinematics (IK). Such approaches are practical and can support embodiment, but previous work shows that adding tracked information does not automatically improve the experience: animation fidelity, latency, drift, and end-

effector coherence also matter (Parger et al., 2018; Eubanks et al., 2020; Yun et al., 2023; Debarba et al., 2022).

These issues are especially relevant for hyper-realistic avatars. Realistic and personalized avatars can strengthen body ownership, presence, and self-identification (Latoschik et al., 2017: Waltemate et al., 2018; Salagean et al., 2023), but their visual fidelity may also make motion anomalies more salient. Camera-based pose estimation offers a low-cost way to enrich sparse VR tracking. MediaPipe/BlazePose estimates 33 body landmarks in real time (Bazarevsky et al., 2020), potentially providing elbow and upper-body constraints that IK must otherwise infer. However, combining headset/controller and camera signals can introduce occlusion, jitter, desynchronization, and coordinate-alignment errors. Direct perceptual comparisons of these alternatives in hyper-realistic self-avatars remain limited. This study therefore examined whether a seated mirror-view avatar is perceived as more faithful with IK-only or IK+MediaPipe tracking, using observer-based and user-reported measures of motion fidelity and agency.

Methods

The experimental study was conducted using the Meta Quest 2 VR headset connected to a laptop with an Intel i7 processor (10th generation), including 16 GB of RAM and an NVIDIA RTX 2060. The hyper-realistic avatars mirroring users' movements were generated from Character Creator (Reallusion Inc.). Two upper-body avatar-tracking configurations were implemented in Unity: an IK-only condition and a hybrid IK+MediaPipe condition.

In the IK-only condition, avatar motion was driven using the tracking data provided by the Meta Quest 2 headset and user's hands through the Meta SDK. The headset position was used to estimate the user's head location, while the hands positions were used as targets for the avatar's hands. To generate plausible upper-body motion from these sparse tracked points, inverse kinematics were computed using the commercial FinalIK plugin for Unity. FinalIK was used to infer the intermediate joint rotations of the upper limbs and spine, including shoulders, elbows, and torso alignment, from the tracked head and hand positions. In this configuration, elbow motion was not directly measured but was generated by the IK solver according to the anatomical constraints and target positions defined in the avatar rig.

In the IK+MediaPipe condition, an additional webcam-based body-tracking pipeline was integrated to obtain complementary upper-body landmark information. A Python application using MediaPipe (Bazarevsky et al., 2020) was adapted to estimate the spatial position of selected upper-body landmarks, specifically focusing on the head, arms, elbows, and hands, since the experimental setup only required upper-body tracking. The estimated landmark coordinates were transmitted from Python to Unity through a UDP communication protocol. Once received in Unity, the coordinates were normalized and transformed to match the three-dimensional coordinate system of the virtual environment. The purpose of this hybrid implementation was to combine the robustness of headset and hand tracking from the Meta Quest 2 with the additional anatomical information provided by MediaPipe, particularly regarding elbow position. In this configuration, the headset and user's hands continued to define the main head and hand targets, while MediaPipe landmarks were explored as additional constraints for improving the realism of elbow motion.

Ten right-handed participants (3 females; 20–30 years) completed both conditions in randomized order. Participants remained seated and viewed a full-body avatar positioned in front of them as a mirror representation (see Figure 1).



Figure 73. *Representation of the virtual environment from participant's perspective.*

For each condition they performed the same sequence of eight movements shown in Figure 2). They then produced spontaneous conversational hand gestures for 20 s. Participants were instructed to perform movements naturally and clearly and to pause briefly between them.

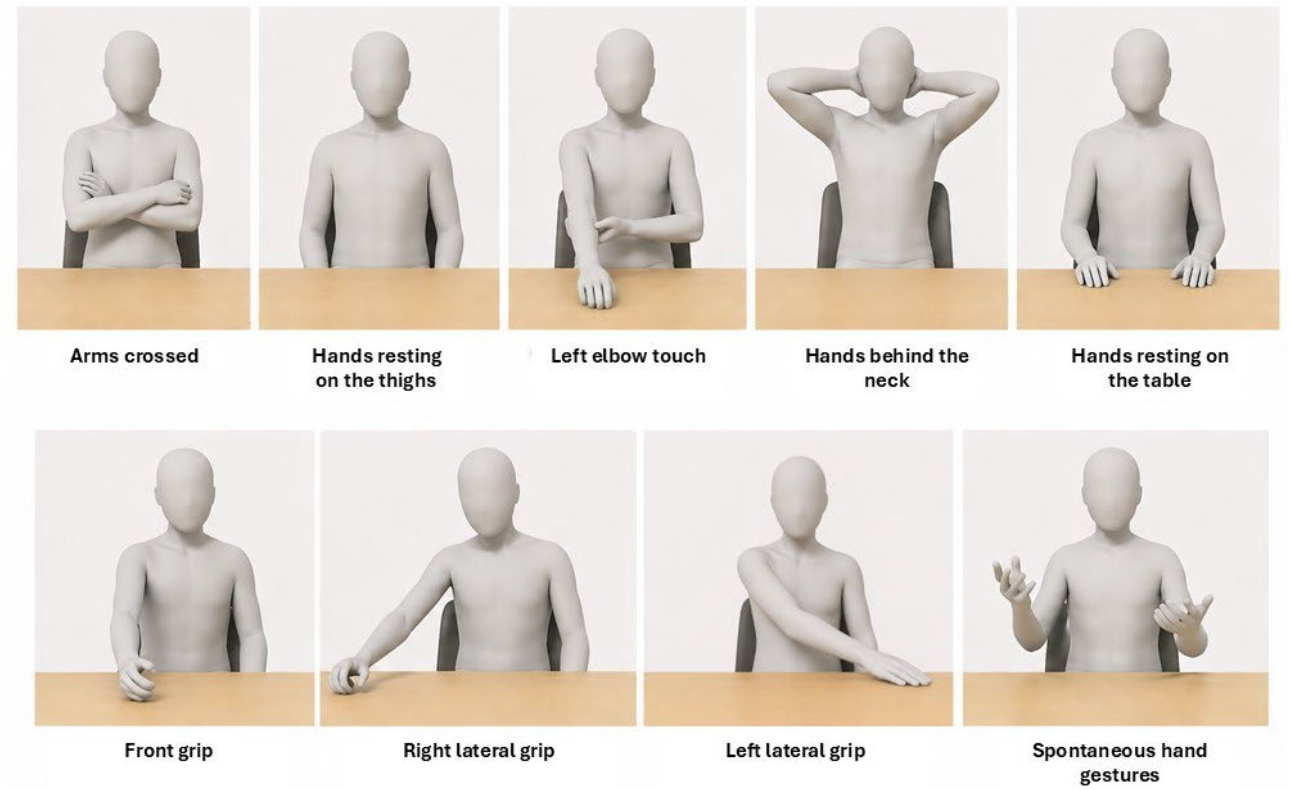


Figure 74*. visual representation of the tested body movements.*

Two experimenters independently rated the fidelity of each movement from the participant's real-time VR view on a 5-point scale (1 = very low precision; 5 = very high precision). After each condition, participants completed a nine-item ad-hoc perceived motion fidelity questionnaire (5-point agreement scale) covering movement replication, complex hand/arm matching, lateral inversion, postural coherence, naturalness, hand similarity, bodily intention, believability, and avatar identification (see Table 1 for the complete item list).

Table 1. *English translations of the self-reported questionnaire on perceived motion fidelity.*

1) The avatar faithfully replicated basic movements such as crossing the arms.
2) The position of the avatar's hands and arms matched mine during more complex gestures (e.g., touching the elbow, simulating grips).
3) I observed that the avatar responded correctly when movements involved lateral inversion.
4) The avatar generally maintained a coherent body posture, without anomalous or unnatural positions.
5) The avatar's movements appeared natural and human-like.
6) The avatar's hand movements were similar to mine.
7) The avatar's gestures conveyed bodily intentions similar to my own.
8) The avatar's representation of my movements seemed believable to me.
9) I felt identified with the avatar during the experience.

They also completed the four-item Agency domain of the VEQ (Peck & Gonzalez-Franco, 2021) on a 7-point scale, assessing motor ownership, control, agency, and movement synchrony. Observer ratings were averaged across the two experimenters for each movement. Wilcoxon signed-rank tests compared conditions separately for every observer-rated movement and questionnaire item. The study was approved by the ethics committee of the Polytechnic University of Valencia (P04_30-06-2025).

Results

No observer-based movement comparison reached statistical significance. Across the nine tested movement categories, p values ranged from .058 to 1.000. Descriptively, both configurations obtained higher fidelity scores for hands on the table, grip simulations, and spontaneous gestures, whereas arms crossed, hands on thighs, elbow touch, and hands behind the neck were more difficult. Although IK-only showed a descriptive advantage for several movements, these differences were not statistically reliable.

Participants' self-report ratings for each condition are presented in Figure 3, showing two significant condition-specific effects.

Participants rated item 2 on the match between their hands/arms and the avatar during complex gestures significantly higher in the IK+MediaPipe condition ($Z = 2.023$, $p = .037$). Conversely, VEQ perceived movement synchrony was significantly higher in the IK-only condition ($Z = -2.521$, $p = .012$). All remaining perceived-motion-fidelity and VEQ comparisons were non-significant; the closest was sense of agency ($Z = 2.023$, $p = .054$), favouring IK-only. Thus, the hybrid system improved one aspect of perceived anatomical correspondence without producing a general advantage in visible fidelity or agency.

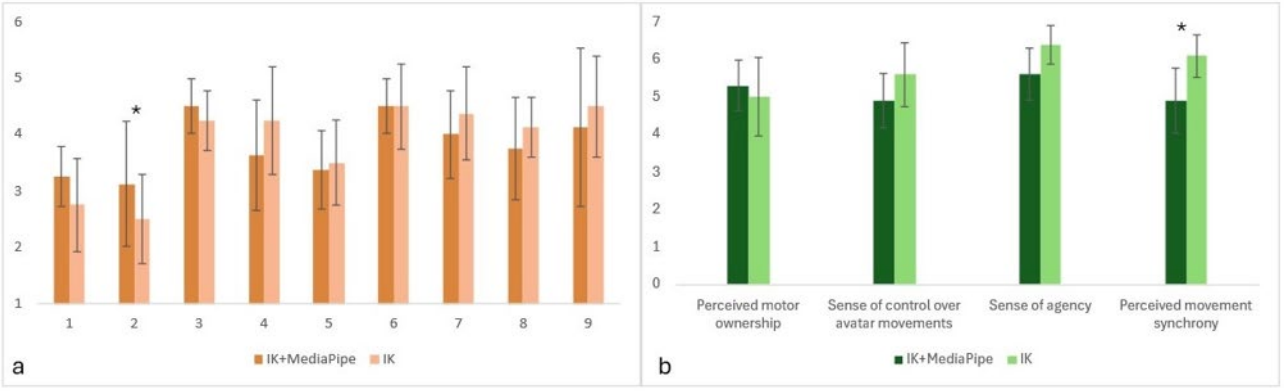


Figure 75. *Mean participant ratings and standard deviations by condition for (a) the perceived motion fidelity questionnaire and (b) the VEQ Agency domain.*

Discussion and Conclusions

The results do not support a general superiority of either tracking configuration. Observer-based fidelity was comparable across conditions, while self-report revealed a trade-off: adding MediaPipe improved perceived matching of hands and arms during complex gestures, but IK-only produced a stronger sense of movement synchrony. The first effect is consistent with the idea that additional anatomical information can help constrain elbow and upper-limb configurations that sparse head-and-hand tracking must otherwise infer [3–5]. For tasks in which the exact configuration of the arms is perceptually important, a hybrid pipeline may therefore provide a useful benefit.

At the same time, additional sensing can create perceptual costs. Visible elbow jitter and tracking artifacts were observed during some IK+MediaPipe movements, although they were not objectively quantified. Previous work indicates that animation plausibility depends on temporal stability, latency, drift, and coherence as well as the amount of available tracking information [5,6]. In a hyper-realistic avatar, small instabilities may be particularly noticeable because appearance establishes strong expectations of natural human motion. This offers a plausible interpretation of why local anatomical matching improved while perceived synchrony decreased.

Performance also depended on movement geometry. Hands-behind-neck, crossed-arm, and elbow-touch configurations were among the most problematic, plausibly because they increase self-occlusion or move targets outside reliable camera or headset-tracking regions. These findings indicate that the choice between IK-only and IK+MediaPipe should be guided by the goals and constraints of the VR application rather than by assuming that more tracking sources are automatically better. If the priority is robust, low-latency, temporally coherent movement in a deployable consumer setup, IK-only may be preferable, particularly for applications where synchrony, agency, and believability are central. If the priority is improving the perceived configuration of arms and elbows during complex gestures, a hybrid approach may be beneficial, provided that the system can control jitter, occlusion, and synchronization issues. This distinction is especially relevant for social VR, training, assessment, and therapeutic applications using realistic or personalized avatars, where different perceptual dimensions may matter depending on the task. The main limitations of the present study are the small sample, the absence of objective kinematic ground truth, and the lack of direct latency/jitter measurements. Future work should include larger samples, movement-specific error metrics, synchronized kinematic ground truth, and confidence-weighted or temporally smoothed sensor fusion. Overall, the study shows that more tracking data are not inherently better: the perceptual value of a tracking pipeline depends on how anatomical accuracy, stability, and movement constraints interact.

Acknowledgements

This work was supported by the Generalitat Valenciana funded project "EASIER", Grant CIPROM/2023/48. MEM received partial funding from the Polytechnic University of Valencia (PAID-10-20).

References

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F. & Grundmann, M. (2020). BlazePose: On-device real-time body pose tracking. arXiv:2006.10204. https://doi.org/10.48550/arXiv.2006.10204

Debarba, H.G., Chague, S. & Charbonnier, C. (2022). On the plausibility of virtual body animation features in virtual reality. *IEEE TVCG* 28(4), 1880–1893. https://doi.org/10.1109/TVCG.2020.3025175

Eubanks, J.C., Moore, A.G., Fishwick, P.A. & McMahan, R.P. (2020). The effects of body tracking fidelity on embodiment of an inverse-kinematic avatar for male participants. *IEEE ISMAR*, 54–63. https://doi.org/10.1109/ISMAR50242.2020.00025

Kilteni, K., Groten, R. & Slater, M. (2012). The sense of embodiment in virtual reality. Presence: *Teleoperators and Virtual Environments* 21(4), 373–387. https://doi.org/10.1162/PRES_a_00124

Latoschik, M.E., Roth, D., Gall, D., Achenbach, J., Waltemate, T. & Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. *Proc. VRST*, 39:1–39:10. https://doi.org/10.1145/3139131.3139156

Parger, M., Mueller, J.H., Schmalstieg, D. & Steinberger, M. (2018). Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality. *Proc. VRST*, 23:1–23:10. https://doi.org/10.1145/3281505.3281529

Peck, T.C. & Gonzalez-Franco, M. (2021). Avatar embodiment: A standardized questionnaire. Frontiers in Virtual Reality 1, 575943. https://doi.org/10.3389/frvir.2020.575943

Salagean, A., Crellin, E., Parsons, M., Cosker, D. & Stanton Fraser, D. (2023). Meeting your virtual twin: Effects of photorealism and personalization on embodiment, self-identification and perception of self-avatars in virtual reality. *Proc. CHI*, 499:1–499:16. https://doi.org/10.1145/3544548.3581182

Slater, M., Spanlang, B., Sanchez-Vives, M.V. & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE* 5(5), e10564. https://doi.org/10.1371/journal.pone.0010564

Waltemate, T., Gall, D., Roth, D., Botsch, M. & Latoschik, M.E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE TVCG* 24(4), 1643–1652. https://doi.org/10.1109/TVCG.2018.2794629

Yun, H., Pontón, J.L., Andújar, C. & Pelechano, N. (2023). Animation fidelity in self-avatars: Impact on user performance and sense of agency. *IEEE VR*, 286–296. https://doi.org/10.1109/VR55154.2023.00044

Behaviourally expressed, behaviourally silent: attention and presence in VR telemetry

Hannah Hayes and Carlo Campagnoli

School of Psychology, University of Leeds, Leeds, United Kingdom

Corresponding author: c.campagnoli@leeds.ac.uk

Keywords: Spatial presence, attention allocation, trait absorption, eye tracking, behavioural telemetry

Introduction and motivation

For two decades XR evaluation has wanted a measure of presence that does not interrupt the experience it measures (Skarbez et al. 2017). The candidate has always been the record the body leaves inside a headset: sway, gaze dispersion (BCEA), head-movement signature. The measure has not arrived and each null has been read as instrumentation failure. The reading has a warrant. Convincing environments elicit behaviour appropriate to them (Slater 2009) and presence forms only when attention is allocated to the mediated space (Wirth et al. 2007, Vorderer et al. 2004), a capacity trait absorption is held to support (Tellegen & Atkinson 1974). If behaviour follows the virtual world and attention builds presence, the motor record should carry a trace of how present its wearer feels. Our own self-report work assumed as much (Bi et al. 2025, Hayes & Campagnoli 2026). Three alternatives exhaust what a behavioural search can return. Either (1) the instrumentation has been too coarse to recover a signal that is there; (2) the environment has not moved behaviour at all; or (3) spatial presence has no behavioural correlate to recover, at any precision these streams support. Deciding between them requires something the literature has not supplied: a second construct, assayed on the same instrument in the same session, to show what the battery recovers. We supplied it. Two reports were scored minutes after one exposure and tested against 69 whole-session indices: spatial presence, the sense of having been located in the mediated space; attention allocation, the devotion of attentional resources to it. If it were true that presence leaves a behavioural trace at the precision these streams support, the presence row should look like the attention row.

Method

Participants and measures. Sixty-one participants took part (mean age 21.3 years, SD 1.9, range 18–28; near-even gender split). Before entering VR they completed demographics, prior VR use and the 34-item Tellegen Absorption Scale (KR-20 = .87), after which they experienced “The Cubicle” on a Meta Quest Pro with integrated eye tracking, with nothing measured during the experience itself. Immediately afterwards they completed four six-item MEC-SPQ subscales (Vorderer et al. 2004): self-location (α = .93) and possible actions (.88), summed into the spatial-presence composite (.91); attention allocation (.88); suspension of disbelief (.81).

The environment. “The Cubicle” (van Beek et al. 2016) stages the alienation of repetitive office work by dismantling a realistic office. In the task phase participants filed eight folders into cabinets, stepping, turning and reaching (median 93 s); filing the eighth folder triggered a dissolution phase

and then a fractal phase inside a Menger sponge, neither of which carried a task, though behaviour was free throughout. The two post-task phases (median 257 s combined) were analysed as the abstract segment. Immersion, in Slater's (2009) sense, was constant; only the content changed, which we treated as an environmental shift.

Indices and analysis. Eye-in-head, head 6-DoF and both hands 6-DoF were logged per rendered frame (sampling rate and collection wave entered as covariates in robustness checks throughout). From these streams we computed 69 whole-session indices in two screens, 32 classical and 37 composite, spanning eight families from movement quantity and stillness to scanning structure (Krejtz et al. 2015), movement spectra (Balasubramanian et al. 2015) and eye–head coordination (Sidenmark & Gellersen 2020). An audit retired fourteen further indices as artefacts. Eye and gaze indices used n = 58, event-based n = 56, head and hand n = 61. Because the battery was redundant by design, we also ran an omnibus permutation test per outcome alongside BH-FDR. Attention allocation was ceiling-limited (11 of 61 at maximum) and neither it nor presence was normally distributed (Shapiro-Wilk $p < .001$ and .005), so associations are Spearman unless an r is given.

Results

Sensitivity

Contrasting the task phase with the abstract segment (n = 60 with hand-coded boundaries, FDR across 17 hypothesis-linked indices), all 17 changed significantly and several by more than a within-subject standard deviation: stationary gaze entropy rose ($dz = +1.30$, the largest effect in the set), gaze transitions became more predictable (−1.11) and participants moved less (−1.04).

The presence null

Not one index reached nominal significance with spatial presence.

Across both whole-session screens (69 indices) there were 0 at $p < .05$; on the composite screen 1.8 were expected by chance, permutation $p = 1.00$ (Fig. 1). The smallest presence p-value across the battery was .076 (classical screen); on the composite screen it was .12, where randomly reshuffled presence scores typically produced .033. The strongest correlate anywhere was $r = −.21$ and gaze BCEA correlated with presence at $r = .04$. The analysis had 80% power for $|r| \geq .35$ at n = 61. Self-location returned 0 of 37, possible actions 2 and suspension of disbelief 0.

The dissociation

Attention allocation behaved quite differently. Across the same screens 12 of the 37 composite indices were associated at $p < .05$, against 1.9 expected by chance (permutation $p = .005$). The signal was diffuse, so we treated the pattern as the result (Fig. 1). Participants reporting more allocated attention made more frequent gaze shifts ($\rho = +.31$) with smaller and less variable saccades within them (−.39) and spent less time with the head still (−.32). More important, collapsing the convergent indices removes the multiplicity problem: six core hand and head indices averaged into a movement-vigour composite ($\alpha = .95$) correlated with attention at $\rho = .37$ ($p = .004$), a single test surviving Bonferroni correction, while correlating essentially zero with presence (.12) and absorption (.03).

Note the asymmetry. Trait absorption predicted spatial presence in self-report substantially ($\rho = .43$, $p < .001$) yet left no behavioural signature: 2 of 37 at $p < .05$, permutation $p = .42$.

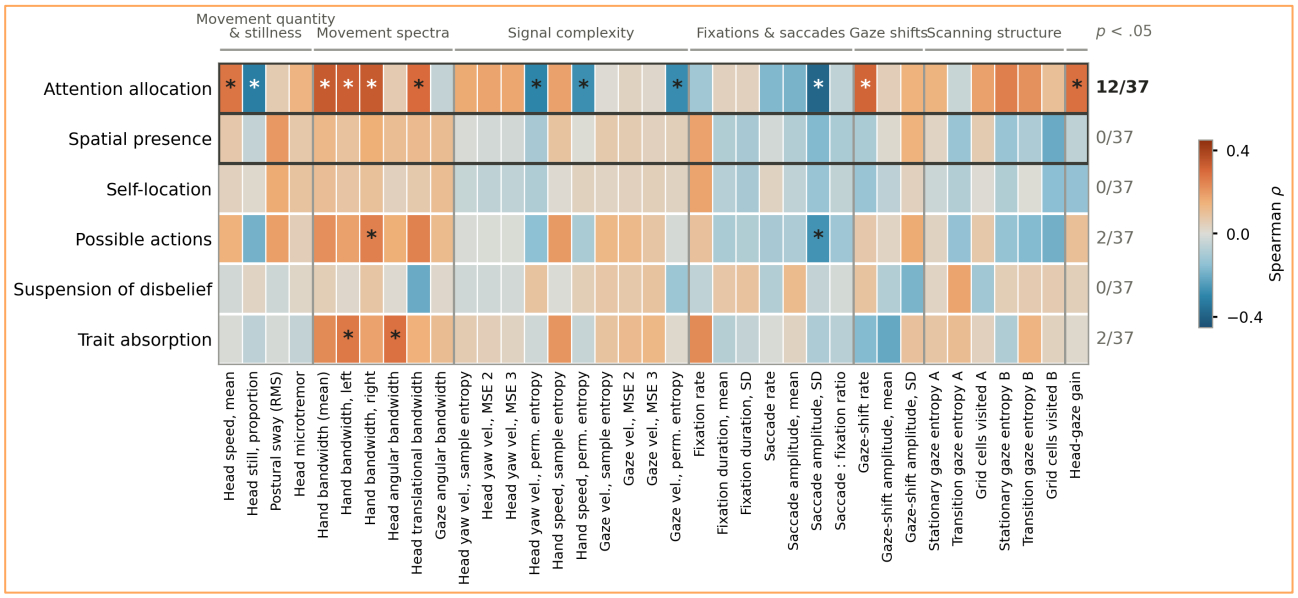


***Figure 1**. Composite behavioural battery × self-report (Spearman ρ; * p < .05, uncorrected; head/hand n = 61, eye/gaze n = 58, event-based n = 56). The rows are the taxonomy: associations cluster in the attention-allocation row (12 of 37 against 1.9 expected) while the spatial-presence row is empty, smallest p = .12.*

Discussion

Three constructs, assayed in one session against one battery, occupy three positions.

Attention allocation: presence-predictive (ρ = .50) and behaviourally expressed, recovered by a single whole-session composite. Trait absorption: presence-predictive (ρ = .43) and behaviourally silent. Spatial presence: behaviourally silent, recovered by no index at the precision these streams support.

These are not degrees of one thing. Both constructs that predict presence do so at a similar strength, yet only one left a record, so what separates them is not the presence link but whether the construct describes something a participant did: a report of allocated attention does, which is why it left a record, whereas a report of presence describes where the participant believed themselves to have been and left nothing this battery could resolve. Attention predicts presence at ρ = .50 while behaviour tracks attention at only ρ = .37, so an association carried entirely through attention would sit near .19, below the |r| = .35 this design resolves at 80% power. What the null rules out is a presence signal of the size behaviour carries for attention, not one attenuated through it.

For evaluation the consequence is definitional: a dashboard fed by gaze dispersion, head motion or movement quantity is an attention assay, so calling it a presence proxy is a category error. That reading is bounded by one environment, one rig, an exploratory design and sixty-one participants, though the presence null itself only firmed up as the hand-coded set expanded from thirty-nine to sixty takes.

We advocate that implicit measures in XR be specified by the construct they recover rather than the one they were built to approximate, which for whole-session telemetry is attention: presence will have to be measured as a report about location until something other than the motor record carries it.

Acknowledgements

We thank Manjari Acharya for assistance with data collection and pre-processing. Ethical approval: University of Leeds School of Psychology Research Ethics Committee, PSCETHS-1666 (30 January 2026).

References

Balasubramanian, S., Melendez-Calderon, A., Roby-Brami, A., & Burdet, E. (2015). On the analysis of movement smoothness. *Journal of neuroengineering and rehabilitation*, *12*(1), 112.

Bi, X., Hayes, H. & Campagnoli, C. (2025). Individual influences in highly immersive VR: two exploratory studies on personality, perception, and presence. In: *Helin, K., Schiavi, B. & Tsaknaki, E. (eds.) EuroXR 2025: Proceedings of the Application, Poster and Demo Tracks of the 22nd EuroXR International Conference*. Espoo: VTT. VTT Technology 440, 225–228.

Hayes, H.R. & Campagnoli, C. (2026). Trait absorption amplifies the path to spatial presence in highly immersive virtual reality: attentional mediation and dose-response effects. *BioRxiv preprint*, March 2026. https://doi.org/10.64898/2026.03.03.709394

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., ... & Villalobos, N. (2015). Gaze transition entropy. *ACM Transactions on Applied Perception (TAP)*, *13*(1), 1-20. https://doi.org/10.1145/2834121

Sidenmark, L. & Gellersen, H. (2020). Eye, head and torso coordination during gaze shifts in virtual reality. *ACM Transactions on Computer-Human Interaction* 27(1), Article 4. https://doi.org/10.1145/3361218

Skarbez, R., Brooks, F.P. & Whitton, M.C. (2017). A survey of presence and related concepts. *ACM Computing Surveys* 50(6), Article 96. https://doi.org/10.1145/3134301

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society* B 364(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Tellegen, A. & Atkinson, G. (1974). Openness to absorbing and self-altering experiences (“absorption”), a trait related to hypnotic susceptibility. *Journal of Abnormal Psychology* 83(3), 268–277.

van Beek, R., Hoogeboom, J., de Graaf, J. et al. (2016). The Cubicle. [VR application]. *Pelican Party Studios*.

Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, L., ... & Jäncke, P. (2004). *MEC spatial presence questionnaire (MEC-SPQ). Short documentation and instructions for application. Report to the European Community, Project Presence: MEC (IST-2001–37661)*.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., ... & Jäncke, P. (2007). A process model of the formation of spatial presence experiences. *Media psychology*, *9*(3), 493-525. https://doi.org/10.1080/15213260701283079

HaptiSense: A Wearable Device for Structured Vibrotactile Interaction toward Semantic Haptics in XR

Ali Najm[1*], Domna Banakou[1], and Despina Michael-Grigoriou[1*]

[1]GET Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, 3036 Limassol, Cyprus

*Corresponding authors: ali.najm@cut.ac.cy, despina.grigoriou@cut.ac.cy

Keywords: Wearable haptics; spatiotemporal vibrotactile cues; semantic haptics; tactile communication; extended reality

Abstract

Wearable haptic feedback in XR is often used to represent reactive events such as contact, collision, navigation, or notification. This work investigates a structured vibrotactile approach for communication-oriented and **semantic haptic** interaction. We consider semantic haptics as a research direction in which deliberately structured spatiotemporal vibrotactile cues are associated with predefined contextual information, including interaction-relevant information. Whether these associations can be reliably perceived, recognized, or interpreted remains an empirical question. Building on our previously published HaptiSense wearable platform, the current implementation constructs cues through combinations of spatial form, direction, spatial extent, movement speed, vibration intensity, and temporal pauses, while visualizing their progression within an XR environment. This demonstration presents the HaptiSense wearable platform (Najm et al., 2024) with selected cue configurations and their corresponding wearable output to illustrate how spatial and temporal parameters can be systematically composed as candidate communication-oriented tactile cues. The contribution of this work is the structured cue-construction approach and its implemented XR realization as a basis for controlled evaluation of communication-oriented vibrotactile interaction.

Motivation and Concept: Toward Semantic Haptics in XR

Haptic feedback in XR is commonly used to support reactive or task-related interaction, including contact, collision, force, navigation, texture, and notification feedback (Wee et al., 2021). Previous research on vibrotactile pattern recognition, Tactons, and tactile icons has also demonstrated that tactile stimuli can be deliberately structured through spatial and temporal properties to represent information (Jones et al., 2009; Azadi & Jones, 2013; Brown et al., 2005; Hoogan et al., 2022). This prior work provides an important foundation for investigating wearable haptics as a communication-oriented interaction modality rather than only as reactive feedback. This direction is also informed by earlier work on affective and social touch, which contributed to the broader investigation of communication-oriented tactile interaction (Najm et al., 2022).

In this work, semantic haptics concept is treated as a research direction in which structured spatiotemporal vibrotactile cues are associated with predefined contextual or interaction-relevant information in XR. The tactile structure and its intended meaning are considered separately: a cue

can be technically defined through properties such as spatial form, direction, extent, speed, intensity, and timing, while its association with information such as direction, urgency, or interaction state remains a candidate mapping. A structured cue is therefore not assumed to be inherently meaningful; whether users can perceive, distinguish, recognize, and interpret these mappings must be established empirically.

This work builds on the previously published HaptiSense wearable platform (Najm et al., 2024) and represents the next stage of the research: constructing, visualizing, and physically rendering parameterized spatiotemporal cues within XR. Selected configurations can be experienced and compared through the wearable system, while a subsequent controlled participant study will evaluate their perceptual distinguishability, recognition, and interpretation. The objective is not to define a universal tactile language, but to investigate how structured wearable vibrotactile interaction can support communication-oriented information presentation in **XR.** The present demo therefore implements this research direction by configuring selected spatiotemporal cues in XR and rendering them through HaptiSense for direct experience.

Structured Spatiotemporal Cue Framework

Each configuration represents a structured tactile cue intended for controlled presentation and later recognition testing. The tactile structure itself is defined independently from its contextual association. A cue may, for example, be associated with a predefined direction, interaction state, or other XR-related information, but this association is treated as a candidate mapping rather than an established meaning. Figure 1 summarizes this progression from XR contextual information and candidate mapping to cue construction, wearable rendering, and subsequent controlled evaluation.

HaptiSense XR Demonstration

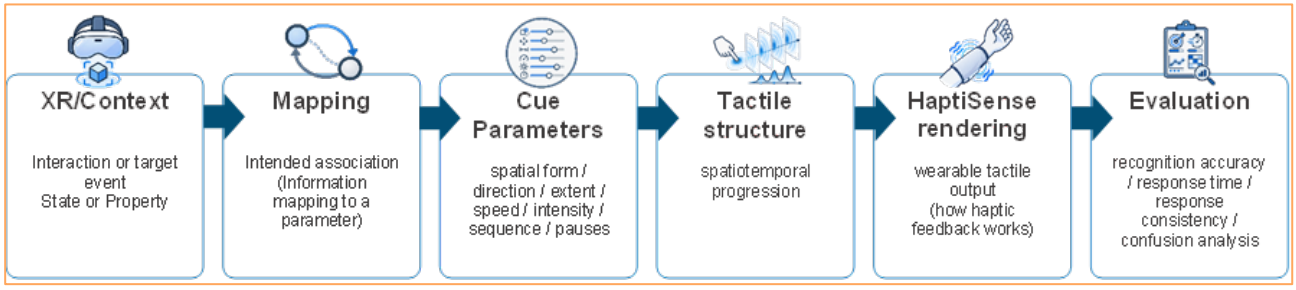


Figure 76. *Conceptual pipeline from XR contextual information to structured vibrotactile cue generation, based on the proposed* ***'semantic haptics'*** *approach and planned evaluation using the HaptiSense platform [1]*

Platform

The cue framework is implemented using the HaptiSense wearable platform, an integrated combination of hardware and software, reported by Najm et al. (2024). The system provides a spatially distributed wearable actuator interface, individually addressable vibrotactile output, Unity-based control, and mapping between virtual interactions and actuator locations. These components provide the rendering infrastructure for the present work.

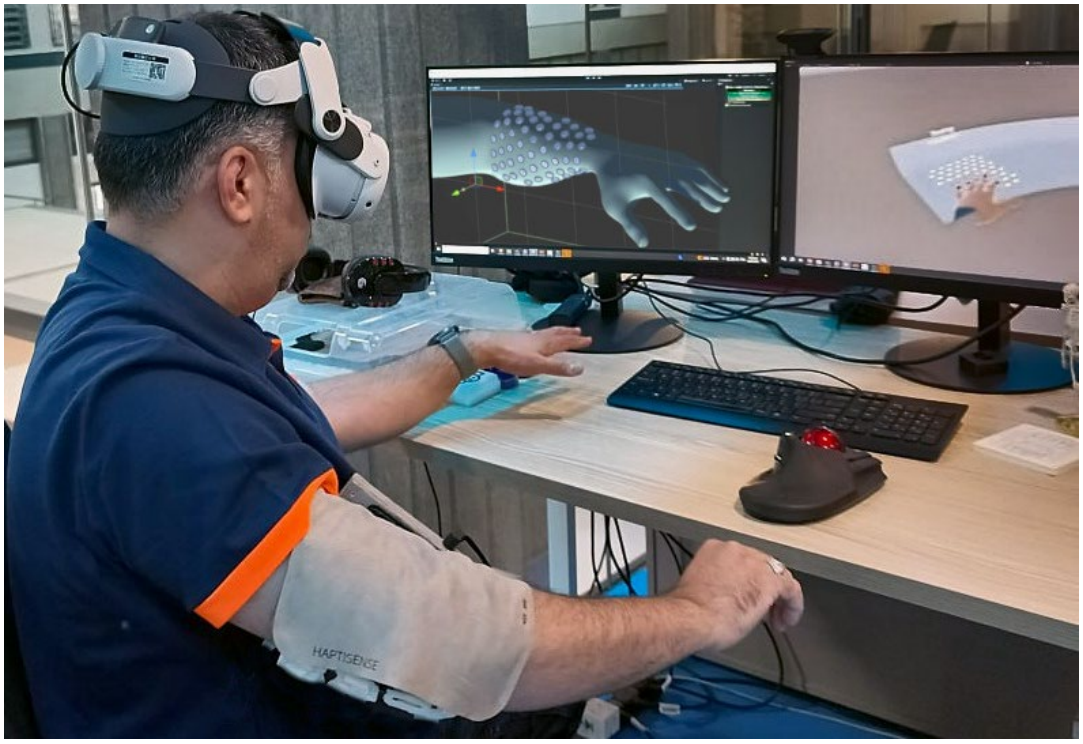


Figure 77. *HaptiSense XR demonstration setup with the wearable device, Meta Quest 3, and Unity visualization.*

XR cue implementation

The XR implementation converts parameterized cue definitions into spatially and temporally varying activation across the wearable interface. In Unity, cue progression is represented through configurable paths and moving virtual stimuli. The HaptiSense mapping layer relates the position of these stimuli to the virtual actuator layout and generates the corresponding actuator commands.

The implementation supports geometric and linear spatial forms together with configurable direction, spatial extent, movement speed, intensity, and temporal pauses. These parameters can be adjusted independently to construct different cue configurations while maintaining the same XR-to-wearable rendering process.

Configured cues

Different cues, varying in shape, size, drawing direction, speed and intensity can be generated. Figure 3 shows examples of spatial forms available in the current implementation, including circular, rectangular, triangular, and linear paths. Direction and other cue parameters can be varied independently of the spatial form, allowing each configuration to represent a different combination of spatial and temporal properties. A smaller subset of these configurations will be selected for the demonstration, with each cue linked to a predefined XR event or interaction state as described in Section 2.

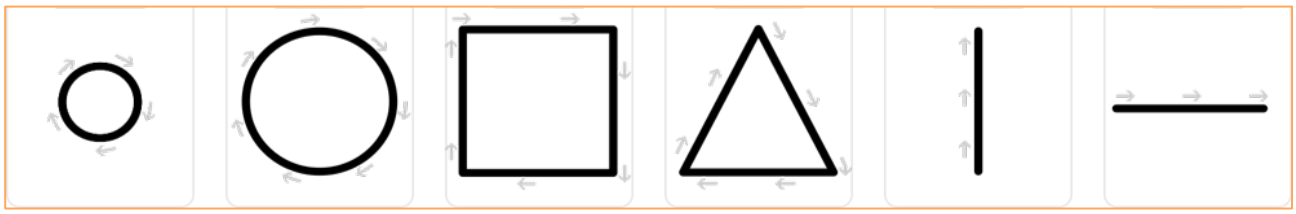


Figure 78. *Examples of configurable spatial cue forms in the current XR implementation.*

Demo Experience and Planned Evaluation

This demo presentation will allow attendees to experience and compare a selected set of structured vibrotactile cues while observing their corresponding progression in XR. Each cue will be linked to a predefined XR event or interaction state, following the mapping approach described in Section 2. The operator will fit the wearable, present the selected cues, vary relevant parameters, and reset the system between demonstrations. A standalone Meta Quest 3 Android build has been deployed to support operation without a host PC.

This demonstration will provide an opportunity to assess the practical operation of the setup and refine the procedure before the controlled experiment. Formal evaluation of perceptual distinguishability, recognition, and semantic interpretation is reserved for a subsequent controlled participant study. Specifically, the subsequent study will evaluate the structured cues using recognition accuracy, response time, response consistency, and confusion analysis to determine whether the predefined cues and their candidate mappings can be reliably differentiated and recognized under controlled XR conditions.

Preparation for the public demonstration includes verifying device operation, fitting and hygiene procedures, safe stopping and reset, and repeatable operation across consecutive demonstrations.

Conclusion

This work presents an implemented XR approach for constructing and rendering parameterized spatiotemporal vibrotactile cues linked to predefined contextual or interaction information. The demonstration allows these cues and their parameter variations to be experienced in practice through the haptic wearable system, showing how the proposed approach can operate within an XR interaction. It also provides a practical foundation for the following controlled experiments of cue differentiation, recognition, and user experience.

References

Azadi, M. & Jones, L. (2013). Identification of vibrotactile patterns: Building blocks for tactons. *2013 World Haptics Conference* (WHC), 347-352. https://doi.org/10.1109/WHC.2013.6548433

Brown, L.M., Brewster, S.A. & Purchase, H.C. (2005). A first investigation into the effectiveness of tactons. *First Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems / World Haptics Conference*, 167-176. https://doi.org/10.1109/WHC.2005.6

Hoggan, E., Raisamo, R. & Brewster, S.A. (2009). Mapping information to audio and tactile icons. *Proceedings of the 2009 International Conference on Multimodal Interfaces*, 327-334. https://doi.org/10.1145/1647314.1647382

Jones, L.A., Kunkel, J. & Piateski, E. 2009. Vibrotactile pattern recognition on the arm and back. *Perception* 38(1), 52-68. https://doi.org/10.1068/p5914

Najm, A., Banakou, D. & Michael-Grigoriou, D. (2024). Development of a Modular Adjustable Wearable Haptic Device for XR Applications. *Virtual Worlds* 3(4), 436-458. https://doi.org/10.3390/virtualworlds3040024

Najm, A., Hadjipanayi, C., Michael-Grigoriou, D., Banakou, D., McGlone, F. & Fairhurst, M. (2022). The Virtual Touch Toolkit: An Interactive Media Mobile Application for Promoting Well-Being through Affective and Social Touch. *2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies* (IMET), 1-4. https://doi.org/10.1109/IMET54801.2022.9929934

Wee, C., Yap, K.M. & Lim, W.N. (2021). Haptic interfaces for virtual reality: Challenges and research directions. IEEE Access 9, 112145-112162. https://doi.org/10.1109/ACCESS.2021.3103598

MIDTOWN: a Model based Immersive Digital Twin Overview With Novice-friendly visualizations

Romain Serres, Charles Bailly, Barbara Schiavi and Julien Castet*

Immersion, Bordeaux, France

*Corresponding author: charles.bailly@immersion.fr

Keywords: Extended Reality, Digital Twin, Information Visualization, Smart City, Collaboration

Introduction

In a rapidly changing world where cities face increasingly complex challenges in design, management and adaptation, how can decision-makers turn vast amounts of urban data into effective actions?

The proposed demonstration focuses on a collaborative XR Digital Twin (DT) of the city of Singapore. While many definitions and concepts related to DTs co-exist (Jones et al., 2020), we consider here the concept of urban digital twins (Ferré-Bigorra et al., 2022), *i.e.* virtual representations of physical systems related to a city (and its associated environment and processes) with exchanges of information between the physical and virtual components. Urban digital twins leverage the data at the heart of smart cities to build and update virtual representations of infrastructures, transport, urban planning or disaster management. Beyond mirroring data, urban DTs can facilitate simulations and decision-making for governments and city stakeholders (Deng et al., 2021) while offering bidirectional communications with physical components.

Extended Reality (XR) offer many benefits for visualizing and interacting with the virtual components of urban DTs. It allows to work directly in a 3D environment while taking advantage of different immersion levels and input/output modalities. Mixing real and virtual environments for Situated Analytics (Lee et al., 2023) is also possible through XR, in addition to co-located and remote collaboration. Nonetheless, urban DT remain complex systems, with a large amount of 2D and 3D heterogeneous data, which raises many Immersive Analytics issues and challenges (Ens et al., 2021).

In this work in progress, we propose the current iteration of an XR demonstrator regrouping a mix of 2D & 3D interactions to explore three main datasets around the Marina Bay of Singapore. We present its design, implementation and the envisioned demo setup. By interacting with a digital twin of a capital city enriched with heterogeneous smart city datasets, including meteorological measurements, thermal maps, and communication network coverage, users can rapidly assess complex situations and evaluate planning scenarios. Participants will experience how immersive analytics enhances spatial understanding and supports urbanism decisions.

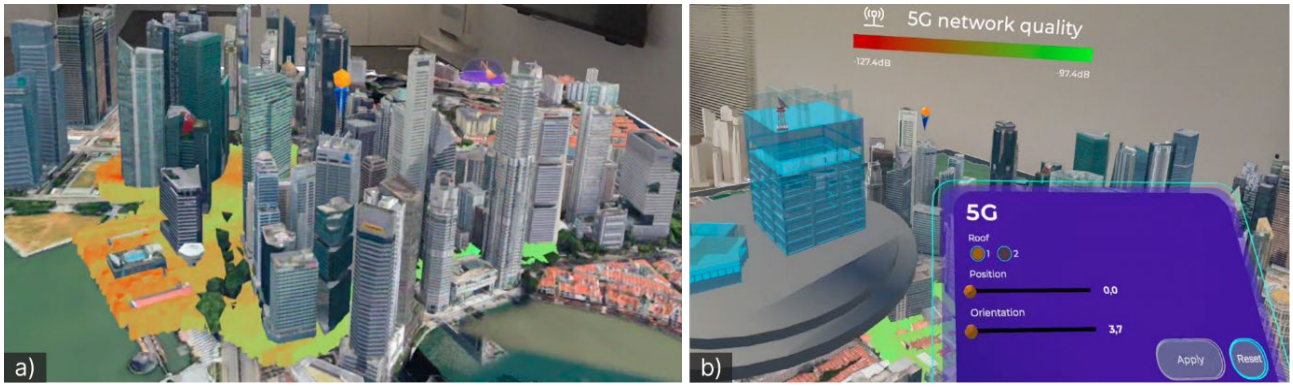


***Figure 79**. Overview of the 5G network coverage a) dataset and b) 2D and 3D interfaces to control antenna parameters.*

Design

MIDTOWN was designed around textured 3D tiles of Singapore from Google, obtained through Cesium. This 3D map is used as shared reference for users, on top of which datasets can be visualized and interacted with. Three main datasets are integrated:

- **Wind dataset**. Wind simulation is represented as streamline trails flowing through the Marina Bay. Users can control the intensity and the orientation of the wind to investigate if some areas have too many turbulences for safe drone flights. The wind speed is represented with a colour gradient. In addition, users can also visualize how air pollution is affected by winds and spreads across the city.
- **Urban heat areas dataset**. Ground temperature is an important factor for the well-being of inhabitants. This dataset as represented as a 3D grid superimposed over the city tile. In addition to the wind parameters, users can filter the range of temperatures they are interested in and the appearance of pillars (transparency, textual description).
- **5G network coverage dataset**. The network coverage is represented as particles covering a neighbourhood. Users can define the position and the orientation of the 5G antenna to visualize the corresponding effects on 5G connectivity.

MIDTOWN includes both 2D UI widgets (sliders, toggles…) and 3D interactions to manipulate the dataset parameters, allowing users to choose between common interfaces and direct manipulation of 3D objects.

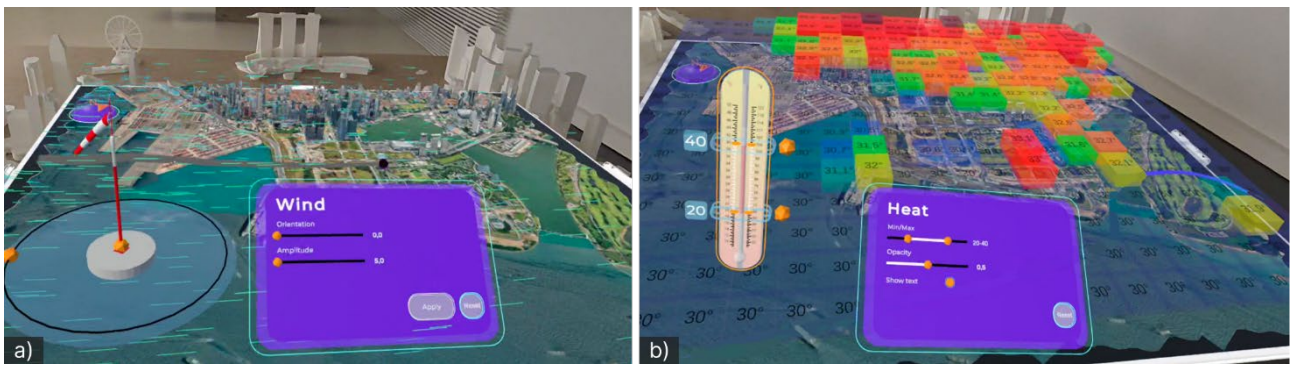


***Figure 80.** Overview of 3D and 2D interfaces for a) wind parameters and b) urban heat areas.*

The application is collaborative: users can join a same session to explore the data together. In this case, one user can take the role of Guide, who control dataset parameters and automatically synchronized them with others. Others then become Spectators, i.e. passive users who can visualize the Guide's datasets but not interact with parameters. Real-time broadcasting of the immersive experience to external displays is also available, allowing non-HMD users to visualize the demonstration. Through this shared workspace, external participants can visualize the XR session, designate locations of interest, highlight virtual objects, and provide navigation cues to the immersed user. This capability facilitates collaborative exploration and decision-making within the digital twin environment.

Finally, the demo also includes a scenario-based storytelling VR experience focused on urban emergency management. Faced with a simulated building fire near the Marina Bay, users must selected simulated intervention resources, including emergency vehicles and response teams. This interactive scenario does not include real data (contrary to the rest of MIDTOWN presented above). Nonetheless, it demonstrates how immersive analytics and collaborative XR technologies can support situational awareness, resource allocation, and decision-making under time pressure in complex urban environments.

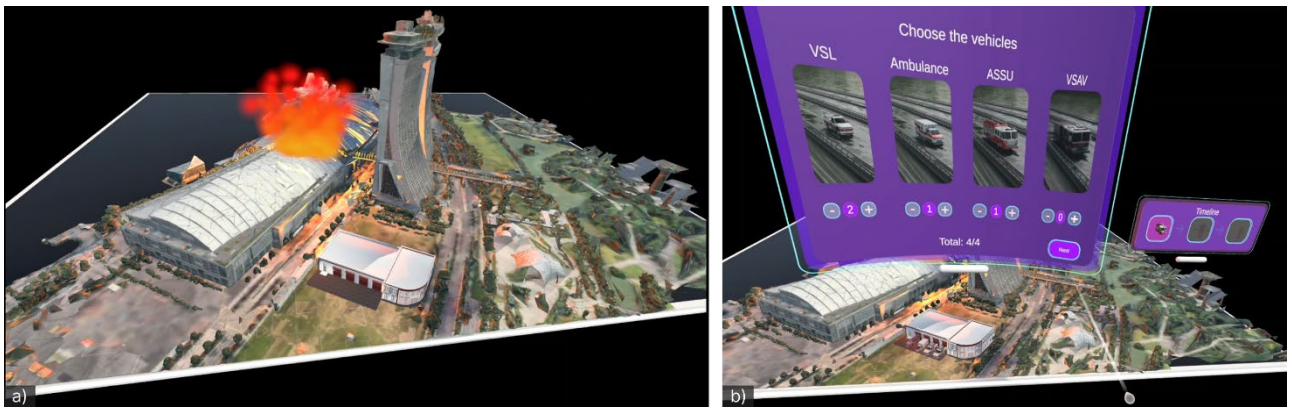


***Figure 81**. Overview of the simulated fire storytelling experience. a) Fire starting. b) Intervention vehicles selection.*

Implementation

Our work is based on Unity (version 2022.3) with the *Cesium for Unity* package (version 1.20.0) to get 3D tiles of Singapore. We used the Photon Fusion module (version 1.1) for the multiplayer features. The application integrates Shariiing XR[6], a plugin from Immersion's Shariiing collaborative software suite, for the streaming to external devices and interaction with external audience.

Datasets were created from city sensors and fed surrogate models directly embarked on the headset. This way, users can adjust in real-time simulation parameters to explore the data, even without internet access. Custom scripts, shaders and VFX graphs were integrated to extract the relevant data from the embarked surrogate models and display them. Specific attention was dedicated to preserving a great Quality of Service (constant 60 fps) adapted for real-time interaction.

The application is compatible with several standalone XR headsets from Meta (Quest Pro, Quest 3) and Pico (Pico 4 and Pico Swan).

[6] With Shariiing XR, your workspaces have never been so immersive | Immersion

Envisioned setup

The demonstration requires Internet connectivity, power outlets, and sufficient space to support safe XR interactions. We can bring our own network router and XR headsets (Quest 3 and Pico 4). To improve the demo, a large tactile screen (tethered to one of our computers) could be valuable to showcase the video streams from the headsets in real time through our ShariiingXR module and foster collaborative analysis. If needed, we may be able to provide our own large tactile display.

References

Deng, T., Zhang, K., & Shen, Z. J. M. (2021). A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of management science and engineering*, *6*(2), 125-134. https://doi.org/10.1016/j.jmse.2021.03.003.

Ens, B., Bach, B., Cordeil, M., Engelke, U., Serrano, M., Willett, W., ... & Yang, Y. (2021, May). Grand challenges in immersive analytics. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-17). https://doi.org/10.1145/3411764.3446866.

Ferré-Bigorra, J., Casals, M., & Gangolells, M. (2022). The adoption of urban digital twins. *Cities*, *131*, 103905. https://doi.org/10.1016/j.cities.2022.103905 .

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP journal of manufacturing science and technology*, *29*, 36-52. https://doi.org/10.1016/j.cirpj.2020.02.002.

Lee, B., Sedlmair, M., & Schmalstieg, D. (2023). Design patterns for situated visualization in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, *30*(1), 1324-1335. https://doi.org/10.1109/TVCG.2023.3327398.

Feasibility of Simultaneous EEG Recording During PCVR-Based Extended Reality Experiences: An Assessment of Signal Quality, Spectral Measures, and Event-Related Potentials

Noemi Passarello[1], Daniela Salvatori[1,2*], and Beerend P. Hierck[1*]
* this two authors have contributed equally

[1]Department of Clinical Sciences - Anatomy and Physiology, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL, Utrecht, The Netherlands

[2]Ombion, Centre for animal-free biomedical translation, Yalelaan 62, 3584 CM, Utrecht, The Netherlands

Corresponding author: n.passarello@uu.nl

Keywords: Extended Reality; EEG; OpenXR; PCVR; Event-related potentials; Validation

Introduction

Extended reality (XR) has become an increasingly important tool in education, healthcare, and cognitive research by enabling immersive and interactive environments for learning and experimentation (Parsons, 2015; Bogomolova et al., 2020). To better understand the cognitive processes underlying XR experiences, electroencephalography (EEG) offers an attractive approach due to its high temporal resolution and ability to capture neural activity associated with attention, perception, memory, and cognitive workload (Luck, 2014). Consequently, the integration of EEG and XR has received growing attention, with previous studies demonstrating the feasibility of recording oscillatory activity and event-related potentials (ERPs) during immersive virtual experiences (Bischof & Boulanger, 2003; Aksoy et al., 2021; Tauscher et al., 2019; Choi et al., 2023; Larsen et al., 2024).

Despite these advances, simultaneous EEG acquisition in XR remains technically challenging. Head-mounted displays may affect electrode contact and increase movement-related artifacts, while accurate synchronization between XR events and EEG recordings is essential for reliable ERP analyses (Tauscher et al., 2019). Furthermore, many existing validation studies rely on hardware and software ecosystems that differ substantially from contemporary OpenXR-based PCVR implementations, limiting the generalizability of previous findings to current XR research environments (Aksoy et al., 2021; Choi et al., 2023; Larsen et al., 2024).

In this work, we evaluated a contemporary OpenXR-compatible PCVR–EEG acquisition pipeline developed to support future neuroscience studies in immersive anatomy education. Specifically, we assessed synchronization stability in wired and wireless configurations, EEG signal quality, spectral characteristics, and the detection of established electrophysiological markers using standard

validation paradigms. Although motivated by educational XR applications, the proposed framework is broadly applicable to cognitive, clinical, and educational XR research.

Methods

Six healthy volunteers participated in this feasibility study. Five participants completed the EEG validation protocol, while one additional participant contributed only to the synchronization assessment. All participants had normal or corrected-to-normal vision and no history of neurological disorders or severe motion sickness.
EEG data were acquired using a 64-channel TMSi SAGA mobile system while participants wore an HTC Vive Focus Vision headset operating in an OpenXR-compatible PCVR environment. The XR application was developed in Unity and synchronized with the EEG system through a dedicated USB TTL trigger interface. Both wired (USB-C) and wireless (Wi-Fi streaming) PCVR configurations were evaluated. The overall system architecture is shown in Figure 1.

The pipeline was evaluated using four cognitive tasks targeting complementary aspects of acquisition quality. First, a timing stability task assessed synchronization accuracy by comparing programmed and observed trigger intervals for both simple and complex visual stimuli under wired and wireless configurations. Second, an eyes-open/eyes-closed paradigm evaluated spectral signal quality through posterior alpha modulation and line-noise measurements. Third, a passive visual oddball paradigm assessed visually evoked ERP responses. Finally, an active auditory oddball paradigm evaluated the detection of canonical N2 and P3 components under task-engaged conditions.

EEG data were preprocessed using MATLAB and FieldTrip. Signal quality was assessed from power spectral density, posterior alpha modulation, high-frequency noise, line-noise contamination, and artifact rejection rates. ERP analyses quantified N2 and P3 amplitude, latency, and signal-to-noise ratio. Given the feasibility nature of the study and the limited sample size, results are presented primarily using descriptive statistics and effect sizes

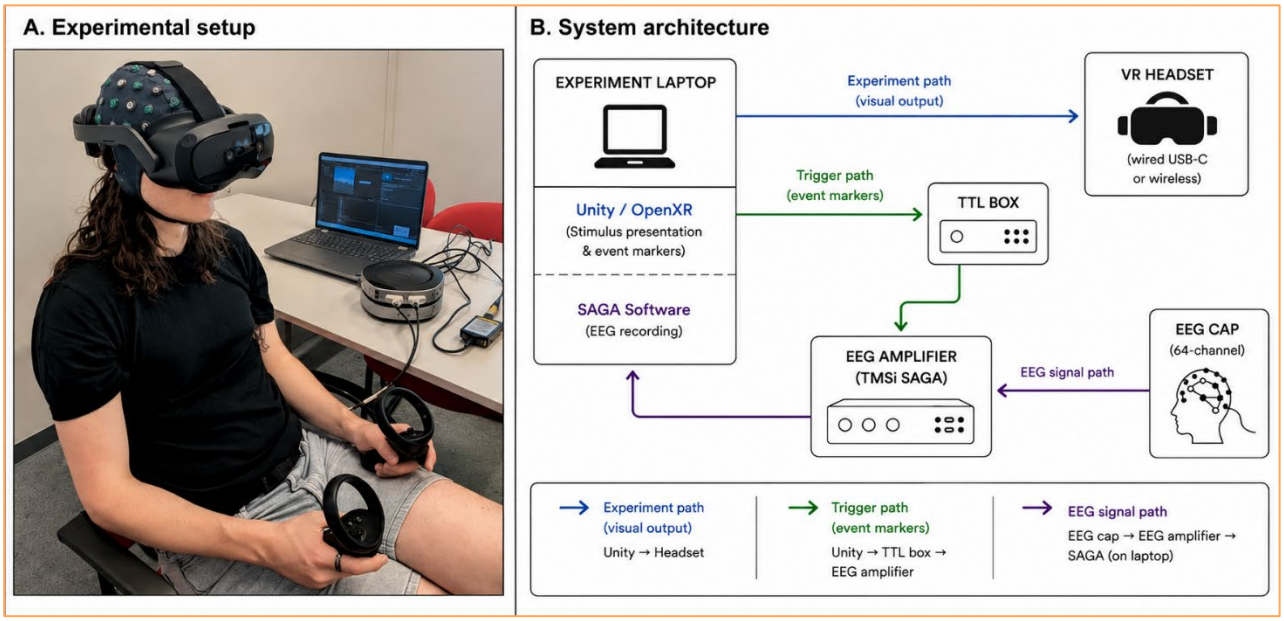


Figure 82. *OpenXR-compatible PCVR–EEG acquisition pipeline.* ***(A)*** *Experimental setup.* ***(B)*** *System architecture illustrating stimulus presentation, TTL synchronization, and EEG acquisition.*

Results

The proposed PCVR–EEG pipeline demonstrated stable synchronization, preserved EEG signal quality, and reliable detection of canonical electrophysiological responses.

Timing stability. Event-marker timing remained stable across both wired and wireless PCVR configurations, with mean timing deviations ranging from 16 ± 7 ms (wireless, simple stimulus) to 47 ± 13 ms (wired, complex stimulus), indicating consistent synchronization performance across streaming configurations and stimulus complexity.

EEG validation. As shown in **Figure 2A**, power spectral density exhibited the expected 1/f profile, with a pronounced increase in posterior alpha power during eyes-closed compared with eyes-open recordings (d = 0.84). The corresponding topography confirmed the expected posterior distribution of alpha modulation, while line-noise contamination, high-frequency noise, and artifact rejection rates remained low, indicating good overall recording quality.

ERP validation. The passive visual oddball elicited only limited ERP modulation. In contrast, the active auditory oddball produced a robust centro-parietal P3 response (difference-wave amplitude: 1.51 ± 1.83 µV, d = 0.82), with the expected scalp distribution (**Figure 2B**), demonstrating reliable detection of canonical event-related potentials during immersive XR.

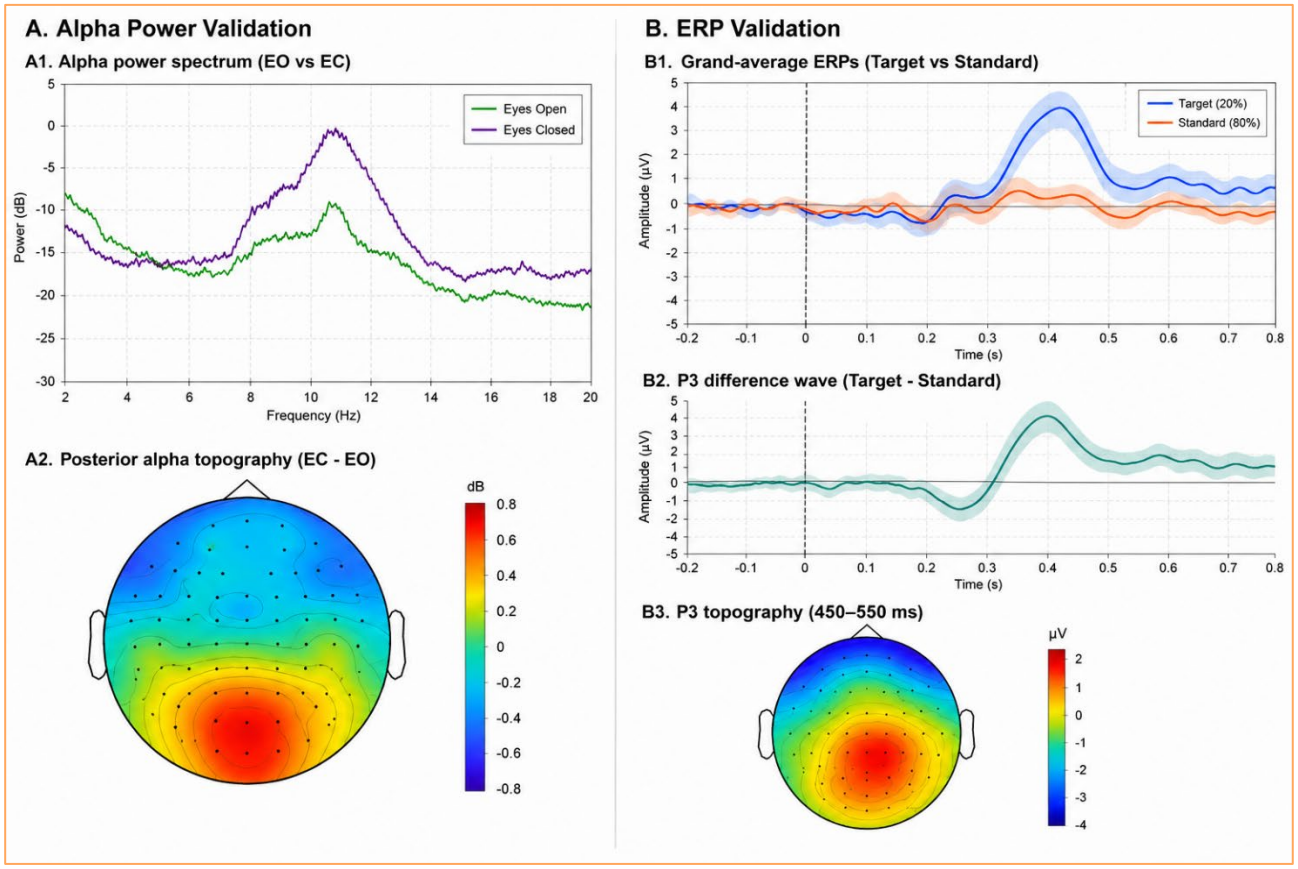


Figure 2. *EEG signal quality and event-related potential measures.* ***(A)*** *Alpha power spectrum (eyes open vs. eyes closed) and posterior alpha topography (EC–EO).* ***(B)*** *Grand-average ERPs, target–standard difference wave, and P3 topography from the active auditory oddball paradigm.*

Discussion

The present study demonstrates the feasibility of simultaneous EEG acquisition during immersive XR using a contemporary OpenXR-compatible PCVR pipeline. Stable synchronization was observed under both wired and wireless configurations, while spectral analyses and the eyes-open/eyes-closed paradigm confirmed the preservation of physiologically meaningful EEG activity.

Furthermore, the successful detection of a canonical P3 response during the active auditory oddball task indicates that the proposed pipeline can reliably capture event-related neural activity during immersive experiences, supporting previous findings obtained in VR environments (Aksoy et al., 2021). Accurate synchronization and preserved signal quality are essential prerequisites for reliable electrophysiological investigations of cognition in XR (Luck, 2014). Together, these findings provide a practical framework for future cognitive, educational, and clinical XR studies using synchronized EEG acquisition.

References

Aksoy, M., Ufodiama, C.E., Bateson, A.D., Martin, S. & Asghar, A.U.R. (2021). A Comparative Experimental Study of Visual Brain Event-Related Potentials to a Working Memory Task: Virtual Reality Head-Mounted Display versus a Desktop Computer Screen. *Experimental Brain Research* 239, 3007–3022.

Bischof, W.F. & Boulanger, P. (2003). Spatial Navigation in Virtual Reality Environments: An EEG Analysis. *CyberPsychology & Behavior* 6, 487–495. https://doi.org/10.1089/109493103769710514

Bogomolova, K., van der Ham, I.J.M., Dankbaar, M.E.W., van den Broek, W.W., Hovius, S.E.R., van der Hage, J.A. & Hierck, B.P. (2020). The Effect of Stereoscopic Augmented Reality Visualization on Learning Anatomy and the Modifying Effect of Visual-Spatial Abilities: A Double-Center Randomized Controlled Trial. *Anatomical Sciences Education* 13, 558–567. https://doi.org/10.1002/ase.1941

Choi, J.W., Kwon, H., Choi, J., Kaongoen, N., Hwang, C., Kim, M., Kim, B.H. & Jo, S. (2023). Neural Applications Using Immersive Virtual Reality: A Review on EEG Studies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31, 1645–1658. https://doi.org/10.1109/TNSRE.2023.3254551

Iwama, S., Takemi, M., Eguchi, R., Hirose, R., Morishige, M. & Ushiba, J. (2024). Two Common Issues in Synchronized Multimodal Recordings with EEG: Jitter and Latency. *Neuroscience Research* 203, 1–7. https://doi.org/10.1016/j.neures.2023.12.003

Larsen, O.F.P., Tresselt, W.G., Lorenz, E.A., Holt, T., Sandstrak, G., Hansen, T.I., Su, X. & Holt, A. (2024). A Method for Synchronized Use of EEG and Eye Tracking in Fully Immersive VR. *Frontiers in Human Neuroscience* 18. https://doi.org/10.3389/fnhum.2024.1347974

Luck, S.J. (2014). An Introduction to the Event-Related Potential Technique. *Cambridge,* MA: MIT Press.

Parsons, T.D. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Frontiers in Human Neuroscience* 9, 660. https://doi.org/10.3389/fnhum.2015.00660

Tauscher, J.-P., Schottky, F.W., Grogorick, S., Bittner, P.M., Mustafa, M. & Magnor, M. (2019). Immersive EEG: Evaluating Electroencephalography in Virtual Reality. In: *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces,* 1794–1800. https://doi.org/10.1109/VR.2019.8797858

DemoXR: A Feasibility Study of Measurement-Grounded AR Instruction Drafting from Robot Demonstrations

Damien Mazeas, Jackson Sunny, Anthony Foulonneau, and Jérémy Lacoche

Orange Research, Cesson-Sévigné, France

Corresponding author: damien.mazeas@orange.com

Keywords: Augmented reality, AR work instructions, authoring by demonstration, RGB-D perception, 3D change detection, vision-language models, industrial maintenance

Introduction

A prior study reported acceptance of mixed-reality technical documentation among experienced industrial workers on the shop floor (Laviola et al. 2023). Such deployments assume the instruction content already exists in writing. Often it does not: many procedures are tacit, and converting a manual into anchored, stepwise AR content remains a per-procedure modelling effort (Gattullo et al. 2022). Generative AI has entered this loop from the document side, spatializing instruction documents (Chen et al. 2023), turning procedural text into in-situ guides (Keelawat & Suzuki 2024) and compiling a query into a vision-grounded task plan (Zhao et al. 2025) or driving just-in-time guidance from a single prompt (Sun et al. 2026). Step correctness is not perfect: Guided Reality reports 65 of 100 steps correct across 15 tasks, and its instructor study found that generative output does not guarantee consistency. On-the-fly guidance tolerates a retry; an instruction that tells a worker where an object goes is judged by being anchored to the right object at the measured pose.

Converting a task executed by a robot into AR guidance for a human worker remains largely underexplored, yet it could support task transfer, operational continuity in case of robot failure, and human–robot collaboration. DemoXR is an offline feasibility prototype that converts an instrumented robot demonstration into a schema-validated AR instruction draft. It deliberately targets a narrow, controlled setting, a cobot performing a simple tabletop pick-and-place task on one workstation, to test the approach end to end and to identify what generalizing it would require. The content to extract is, for each step, which object is moved, from where to where, and with what orientation, so that an AR client can later highlight and label the object and display its measured displacement and destination. Recovering this from a demonstration means segmenting it into steps, localizing what changed in 3D, and naming the object. Multi-view RGB-D change detection and robot telemetry compute the geometric fields (source and destination location, displacement, planar yaw); a vision-language model (VLM) supplies only a semantic object label and measurement-conditioned wording and is never used to infer geometric pose. The design is intended to reduce the degrees of freedom available for spatial errors by deriving geometric fields from measurements rather than VLM inference; this study does not measure instruction-level spatial-error or hallucination rates. Constraining generated plans to measured geometry is an active strategy against spatial hallucination in robot planning (Jung et al. 2026); DemoXR applies it to human-facing instructions instead. The closest prior art also authors from demonstration: AREDA aligns to pre-uploaded CAD by ICP and leaves step text to the author (Bhattacharya & Winer 2019), and Lenssembly uses

programming-by-demonstration in AR (Kosch et al. 2023). Our contribution is this system design rather than new algorithms, and the evaluation assesses perception outputs rather than end-user instruction quality. Across 30 controlled relocation trials, the perception core estimates object-center displacement with a mean error of 1.24 cm, while full-form object-label agreement is 90%.

Human demonstration is a planned extension. The present prototype uses robot replay because repeatable rest states, telemetry-based segmentation and robot-arm masking from the posed URDF enable controlled evaluation of the measurement pipeline; capture from a human demonstrator would additionally require the operator to be segmented and masked.

The DemoXR pipeline

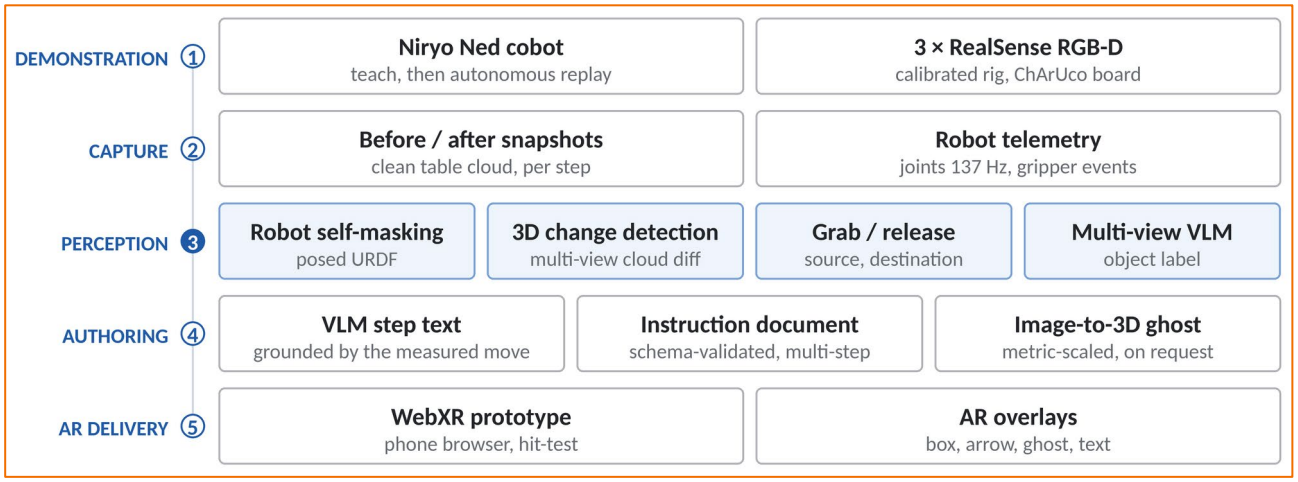


Figure 1. *The five-stage DemoXR pipeline, from multi-camera RGB-D and telemetry capture to a schema-validated instruction document rendered in a WebXR prototype.*

Three Intel RealSense cameras record aligned color and depth streams while a Niryo Ned cobot logs telemetry; capture is the only on-line stage. Calibration places every sensor in one metric table frame, with ChArUco extrinsics (Garrido-Jurado et al. 2014) refined by ICP. The metric core is classical point-cloud geometry on Open3D (Zhou et al. 2018): the cameras are fused into a voxel-downsampled (1 cm) table-frame cloud at the rest states either side of a motion segment and differenced, so the static workstation cancels and only the changed regions remain. Bidirectional nearest-neighbour distances (r = 2.5 cm) yield appeared and disappeared point sets, clustered by DBSCAN into an oriented bounding box for each cluster, so the changed region is localized without a detector gate. The recovered pose is planar translation plus yaw, with roll and pitch assumed zero.

Segmentation is kinematic: end-effector speed is thresholded (0.015 m/s) into motion bursts, and multi-move demonstrations yield linearly chained steps. Per segment the captioner prompts a VLM (Gemini Team 2023) for one imperative sentence constrained to the measured displacement and rotation, with a geometry-only fallback if the call fails, so the VLM is not a dependency of the metric path. Every artefact is checked against a published JSON Schema before it is written and a non-conforming document is refused; this establishes structural validity, not semantic correctness. Each step pairs a textual action with a measured object, anchor and render assets, an Asset Administration Shell (AAS) aligned mapping (Belfadel et al. 2026); the emitted JSON is not yet a conformant submodel. The document carries anchors in the calibrated table frame; a three.js/WebXR prototype renders the highlight box, arrow and scaled ghost mesh in a phone browser by hit test placement, and registering that frame on the device is future work.

A worked feasibility demonstration

The demonstrated task is a two-step object relocation: the cobot moves a white box to a new spot on the table, then places a power adapter inside it. Kinematic segmentation found the two motion bursts and the model generated a measurement-conditioned caption for each (Figure 2). Step 1 moves the box 32 cm with a recovered rotation of −86°; the box stays visible, so its destination

comes from differencing the before and after clouds. Step 2 moves the adapter 21 cm into the box, where it is no longer visible. Its destination is therefore estimated from the end-effector telemetry pose at the end of the placement motion, and this occluded-placement estimate has not been separately evaluated. Grasp and release classification is implemented but not integrated, so the document records object relocations rather than explicit pick and place events. This is a qualitative feasibility case rather than evidence that generated instructions are generally correct.

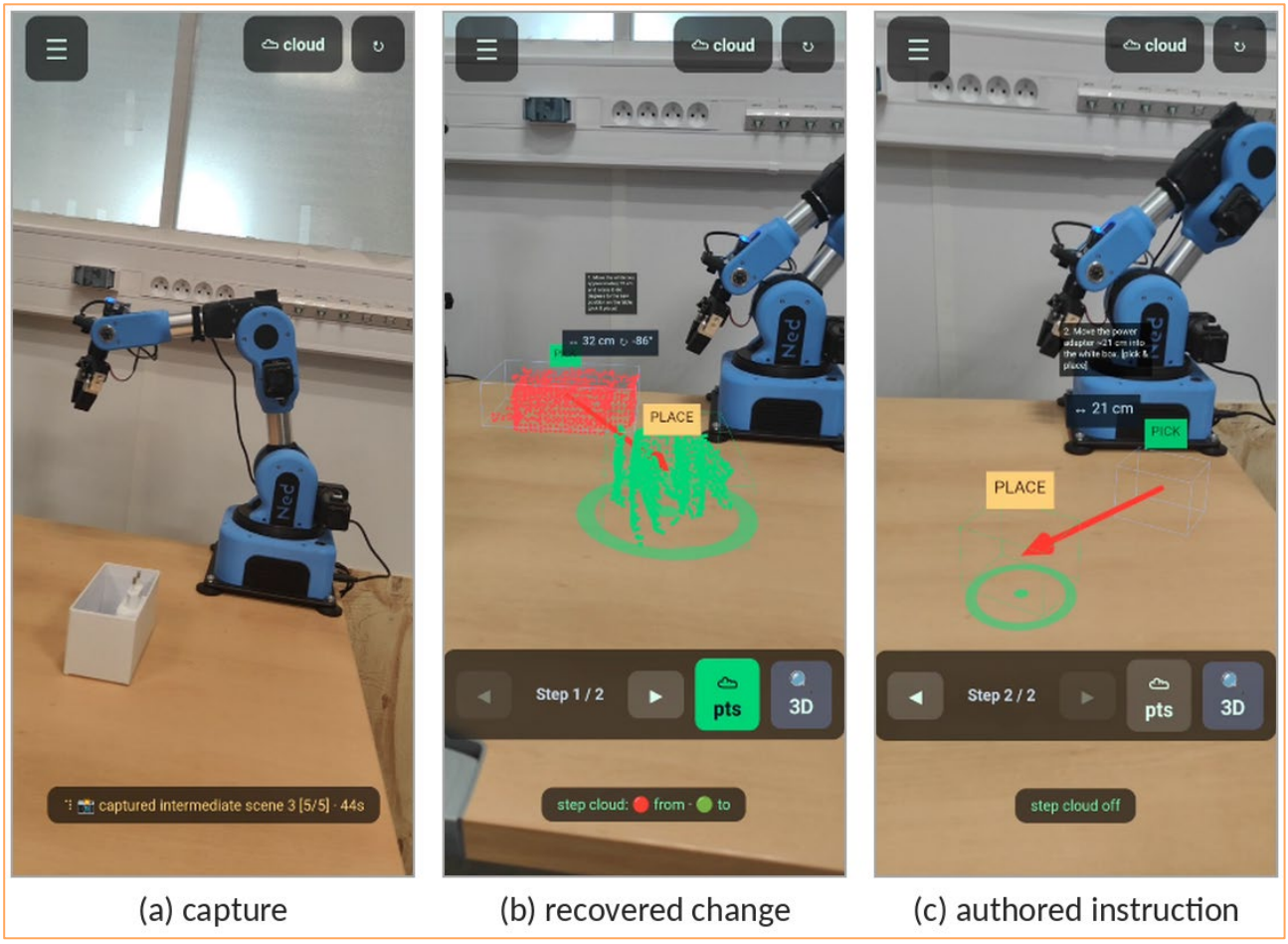


Figure 2. *(a) Capture during the robot replay. (b) Step 1, the recovered change as before (red) and after (green) point clusters with the measured move. (c) Step 2, the instruction in the WebXR prototype, illustrating document playback; physical registration to the table frame is not yet evaluated.*

Preliminary perception accuracy

We evaluated the perception core against hand-measured ground truth over N = 30 controlled single-object relocations (Table 1). Metric placement is the most accurate of the three measures, at 1.24 cm mean and 2.5 cm at worst. The VLM label matches the hand-confirmed name in 90 % of trials on the full colour-plus-noun form and 93.3 % on the category noun alone; all three misses are naming errors on correctly localized geometry, so the failure mode is naming, not localization. For rotation we report the well-defined subset of n = 21, setting aside nine trials whose yaw is unobservable or ill-defined on symmetric objects; the median error is 7.8°, the mean inflated by a single failure at 82.9°.

Table 1. *Perception accuracy over N = 30 controlled single-object relocations, against hand-measured ground truth.*

Measure	Result	Source
Object label, full form	90.0 % (27/30)	VLM, scored by hand
Object label, category noun	93.3 % (28/30)	VLM, scored by hand
Centroid displacement error	1.24 cm mean, 1.15 median, 2.5 max	hand-measured with tape measure
In-plane yaw error (n = 21)	12.75° mean, 7.8° median	hand-measured with protractor

Scope of this prototype and perspectives

- One workstation, controlled single-object relocations (30 recordings)
- Planar translation and yaw only, no full 6-DoF pose recovery
- Robot replay and telemetry, not human-demonstration capture
- Perception-core evaluation only, no instruction-level correctness evaluation
- No user study or evaluation of task performance, safety, or worker comprehension
- WebXR playback prototype, table-frame registration not yet evaluated

An instruction-level evaluation protocol covering step detection and ordering, anchor error, and instruction-grounding errors is under development. The next step is to score generated instructions against manually authored ground truth, followed by marker-based registration of the WebXR client and AAS-conformant submodel export.

The controlled tabletop setting was chosen to make these limits explicit. Extending the approach to larger-scale environments and more diverse robotic systems raises further challenges: recognizing a wider range of tasks, such as assembly operations; detecting objects of varying types, shapes, and sizes; adapting robot-specific trajectories to human-executable motions; handling dynamic or cluttered environments; integrating safety constraints; and evaluating the usability and effectiveness of the generated AR instructions with human users.

The poster shows screenshots of the capture, the recovered changes, and the WebXR playback, together with a link to a video of the demonstration: https://youtu.be/yzsqCRgBCQw.

References

Belfadel, A., Lacheroy, T., Creff, S., Ben Hamida, A. & Didier, J.Y. (2026). XRTwin4Industry: Empowering industrial digital twins with XR and instruction system, a framework for standardization and interoperability. *Future Generation Computer Systems* 182, 108476. https://doi.org/10.1016/j.future.2026.108476

Bhattacharya, B. & Winer, E.H. (2019). Augmented reality via expert demonstration authoring (AREDA). *Computers in Industry* 105, 61–79. https://doi.org/10.1016/j.compind.2018.04.021

Chen, C., Nguyen, C., Hoffswell, J., Healey, J., Bui, T. & Weibel, N. (2023). PaperToPlace: Transforming instruction documents into spatialized and context-aware mixed reality experiences. *Proceedings of UIST* '23. ACM. https://doi.org/10.1145/3586183.3606832

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J. & Marín-Jiménez, M.J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47(6), 2280–2292. https://doi.org/10.1016/j.patcog.2014.01.005

Gattullo, M., Evangelista, A., Uva, A.E., Fiorentino, M. & Gabbard, J.L. (2022). What, how, and why are visual assets used in industrial augmented reality? A systematic review and classification in maintenance, assembly, and training (from 1997 to 2019). *IEEE Transactions on Visualization and Computer Graphics* 28(2), 1443–1456. https://doi.org/10.1109/TVCG.2020.3014614

Gemini Team, Google. (2023). Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:*2312.11805.

Jung, S., Song, H., Kim, D.H., Tan, R., Gao, J., Lee, Y.J. & Kim, D. (2026). Spatially grounded long-horizon task planning in the wild. *arXiv preprint arXiv*:2603.13433.

Keelawat, P. & Suzuki, R. (2024). Transforming procedural instructions into in-situ augmented reality guides with InstructAR. *Adjunct Proceedings of UIST* '24. ACM. https://doi.org/10.1145/3672539.3686321

Kosch, T., Knierim, P., Kritzler, M., Beicht, D. & Michahelles, F. (2023). Lenssembly: Authoring assembly instructions in augmented reality using programming-by-demonstration. In: Human-Technology Interaction: Shaping the Future of Industrial User Interfaces. *Springer,* 199–222. https://doi.org/10.1007/978-3-030-99235-4_8

Laviola, E., Romano, S., Gattullo, M. & Uva, A.E. (2023). Evaluating the worker technology acceptance of a mixed reality technical documentation. Virtual Reality and Mixed Reality (EuroXR 2023). *Lecture Notes in Computer Science 14410. Springer*, 166–180. https://doi.org/10.1007/978-3-031-48495-7_11

Sun, Y., Jiang, Y., Lu, J., Yang, Y., Kuo, Y.H. & Jiang, C. (2026). JARVIS: A Just-in-Time AR Visual Instruction System for Cross-Reality Task Guidance. *arXiv preprint* arXiv:2604.10108.

Zhao, A.Y., Gunturu, A., Do, E.Y.L. & Suzuki, R. (2025). Guided Reality: Generating visually-enriched AR task guidance with LLMs and vision models. *Proceedings of UIST '25*. ACM. https://doi.org/10.1145/3746059.3747784

Zhou, Q.Y., Park, J. & Koltun, V. (2018). Open3D: A modern library for 3D data processing. *arXiv preprint* arXiv:1801.09847.

Externalising and Regulating Emotion through Breath: A Therapist-Supervised Multi-User Mixed Reality Intervention

Jorge Larrosa-Quesada[1 2], Samuel Navas-Medrano[1], Jose Luis Soler-Dominguez[1], Juliana Rojas-Rincon[2], Aikaterini Ampatzoglou[2], Maria Fernandez Capo[2], Maria Gámiz[2], Bruno Porras García[2] and Patricia Pons[1]

[1]HCI research group, Instituto Tecnológico de Informática, Valencia, Spain

[2]BrainXR lab, Universitat International de Catalunya, Barcelona, Spain

Corresponding author: snavas@iti.es

Keywords: Mixed Reality; Therapist-Mediated Interaction; Multi-User Systems; Digital Mental Health

Introduction

Emotion regulation comprises the processes through which people identify, understand, and modify emotional responses (Gross & Thompson, 2006). Difficulties in these processes are relevant across mental-health conditions and are particularly important in preventive work with adolescents and young adults at risk of eating disorders. Dialectical behaviour therapy (DBT) and related third-wave interventions teach skills such as mindful attention, emotion labelling, acceptance, distress tolerance, and response modulation. Evidence suggests that these approaches can improve disordered-eating and body-image outcomes, while dropout and practical delivery challenges remain (Linardon et al. 2019, Dixon & Linardon, 2020).

Extended reality (XR) can make otherwise intangible processes visible, interactive, and situated in the user's physical environment. In the proposed intervention, this capacity is used to transform an internal and subjective state into a manipulable visual representation and then connect that representation to a concrete self-regulation exercise. The activity is not intended to infer or classify a “true” emotion from physiology. Instead, it provides a personally authored metaphor whose visual behaviour is driven by respiration and interpreted in dialogue with a therapist.

This work presents an MR intervention design consisting of (1) externalising emotion through a customisable mannequin; (2) a breath-responsive transition that makes regulation progress perceptible without relying on numerical interfaces; (3) and an asymmetric multi-user design in which one therapist can supervise up to five private, individual activities from a shared MR view.

Related Work

XR has increasingly been explored as a medium for mental-health assessment, skills training, exposure, and therapeutic support (Pons et a. 2022). Its value is not limited to immersion: virtual

content can be spatially anchored, embodied, manipulated, and adapted in real time. In the eating-disorder domain, virtual bodies and embodiment-based exposure have been used to address body-image disturbance and fear of weight gain (Gutierrez-Maldonado et al. 2026, Porras-Garcia et al. 2021). MR extends these possibilities by preserving awareness of the real room and enabling co-present interaction between patients and professionals.

Existing MR systems have also investigated emotional regulation through gamified and customisable activities. ARCADIA, for example, explored modular MR activities for emotional regulation and self-compassion, highlighting the importance of therapist involvement and adaptable interaction (Soler-Dominguez et al. 2024). Also, a DBT-informed VR intervention combining nature exposure and guided breathing increased relaxation in adolescents regardless of their rumination levels, supported the feasibility of immersive environments for training distress tolerance and emotion regulation (Rojas-Rincón et al., 2026). Bioresponsive avatars and visualisations provide another relevant direction: changes in physiological signals can be mapped onto a virtual body or object to support awareness and reflection (Skiers et al. 2022). Such mappings, however, require careful framing because physiological arousal is not equivalent to a specific emotion or clinical state.

The proposed MR intervention builds on these strands but addresses a different interaction configuration. The emotional representation is authored by the participant rather than automatically generated, respiration is used as feedback for a guided exercise rather than as an emotion detector, and the system explicitly supports concurrent supervision. Many XR interventions are designed either for a single user or for symmetric collaboration in which participants share the same virtual content. Our design instead uses asymmetric visibility: patients remain perceptually isolated from one another's therapeutic material, while the therapist receives a combined overview. This supports group deployment without turning an individual emotional exercise into a shared or socially exposed one.

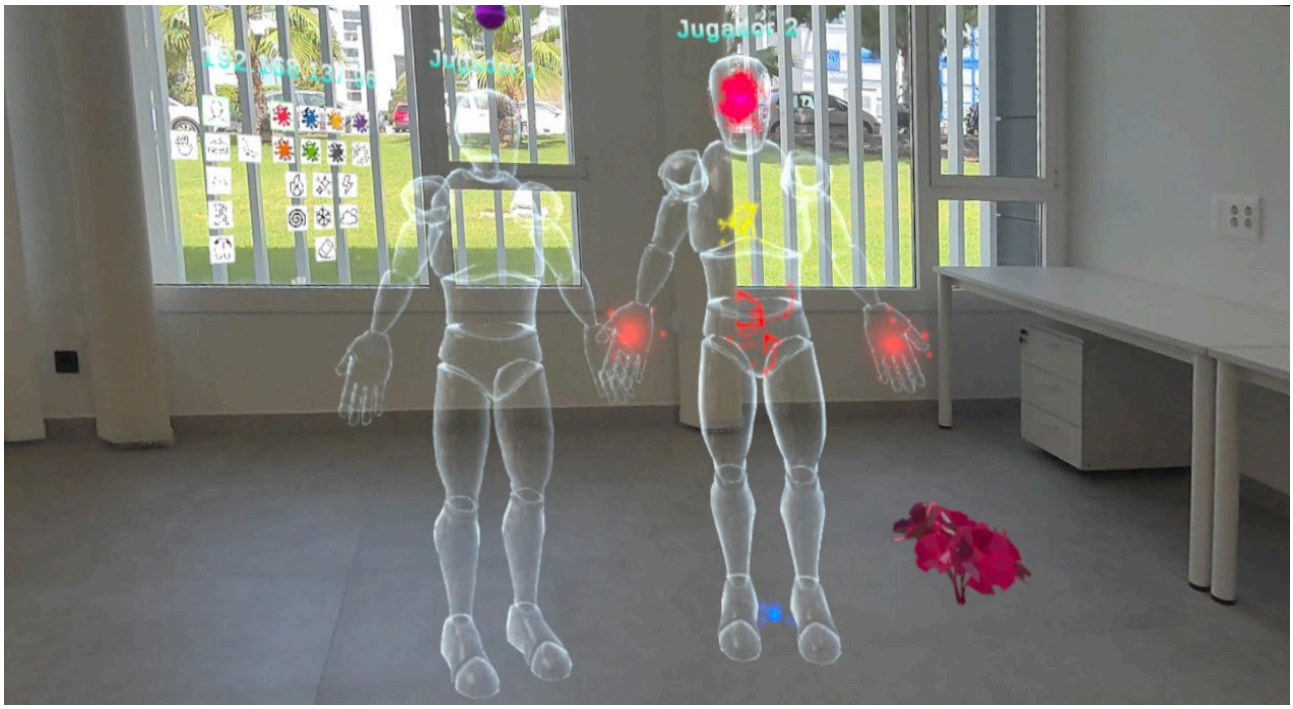


Figure 83. *Therapist's perspective seeing the bioresponsive avatars (mannequins) of two users.*

Design of the intervention

Co-designed by technologists and psychologists and informed by focus groups with patients with eating disorders, relatives, and therapists; the intervention is structured as a transition from externalisation to observation, regulation, and reflection. It is intended to be conducted under the guidance of a trained therapist rather than used as a standalone self-help activity. The intervention was implemented using Meta Quest 3[7] headsets for participants and the therapist. Respiration was captured through a BITalino (r)evolution[8] with a piezoelectric (PZT) respiration sensor[9] and used to control the progressive fading of the visual effects.

Emotional externalisation

At the beginning of the activity, the therapist elicits emotional responses by inviting participants to visualize preselected film scenes designed to evoke emotions relevant to the session. A humanoid mannequin appears in the participant's physical environment through the mixed reality headset.

Participants are provided with a predefined collection of visual effects and colours. They can apply and combine these elements on the mannequin to create a personally meaningful representation. Effects may communicate qualities such as movement, pressure, agitation, heaviness, fragmentation, intensity, or diffusion, although the system does not assign fixed emotional definitions to them.

This stage encourages participants to make active representational decisions. The resulting mannequin is not presented as a computational assessment. It is a participant-authored artefact that provides a basis for discussing where the emotion is experienced, how intense it feels, and which visual qualities appear most closely connected to it.

Guided observation

Once the emotional representation has been created, the therapist initiates a guided observation phase informed by mirror body-exposure procedures. Participants are invited to maintain their attention on the representation without immediately removing, attenuating, or regulating it. Although the mannequin does not reproduce the participant's physical appearance and is therefore not equivalent to conventional mirror exposure, it provides a human-shaped representation through which embodied emotions and sensations can be observed from a manageable distance. This sustained observation aims to reduce experiential avoidance and promote a non-judgemental stance towards the emotional experience (Delinsky & Wilson, 2006; Griffen et al., 2018).

Because the therapist can see the participant's mixed-reality content, guidance can refer directly to the displayed elements. Participants may first be asked to describe the colours, shapes, movements, intensity, and bodily locations in neutral terms and then explore their emotional meaning. For example, the therapist may ask what a particular colour represents, why an effect has been placed in a specific area, or what feelings arise while observing it. This combines elements of guided non-judgemental exposure with approaches that encourage awareness of the emotions elicited by body-related stimuli (Moreno-Domínguez et al., 2012). The visualisation may be refined to represent the participant's experience more accurately; however, removing elements solely to reduce discomfort is postponed until the observation phase is complete, as this could reinforce avoidance. The regulation exercise is subsequently introduced after the emotional experience has first been acknowledged and tolerated.

Breathing-based regulation

Participants subsequently connect to a respiration sensor and complete a guided diaphragmatic breathing exercise. The application processes the respiratory signal to identify periods that meet the

[7] https://www.meta.com/es/quest/quest-3
[8] https://www.pluxbiosignals.com/products/assembled-bitalino-core-ble-bt
[9] https://www.pluxbiosignals.com/products/respiration-pzt

parameters established for calm breathing within the activity. When a calmer pattern is sustained, the selected visual effects progressively increase in transparency.

The transformation is gradual rather than instantaneous. This creates a visible relationship between sustained breathing practice and change in the representation. If calm breathing continues, the effects eventually disappear, leaving the mannequin without the previously selected overlays.

The disappearance of the effects is designed as a metaphor for the modulation of emotional intensity rather than the elimination of emotion itself. It represents the completion of the exercise and gives participants an observable outcome connected to their breathing. Importantly, the application does not infer that the participant has become emotionally calm merely because a respiratory threshold has been reached. The participant and therapist remain responsible for interpreting the experience.

Therapeutic debriefing

After the breathing exercise, the therapist can guide a short reflection on the process. Relevant topics may include whether the visual transformation corresponded with the participant's subjective experience, whether some effects felt more difficult to release, and how breathing influenced bodily sensations or attention.

This closing stage prevents the interaction from being reduced to a simple success/failure mechanism. Reaching the respiratory target is treated as practice in using a regulation strategy, not as evidence that an emotion was incorrect or that it should always disappear.

Discussion and Conclusion

The presented intervention explores how MR can connect subjective emotional representation, physiological interaction, and therapist-led regulation training. Its central interaction loop begins with a participant-authored visualisation and ends with a respiratory-driven transformation of that visualisation. This preserves a distinction between the subjective meaning of an emotion and the measurable properties of breathing.

The multi-user design also illustrates how immersive group interventions do not need to rely on completely shared virtual content. Asymmetric visibility can support simultaneous participation while preserving private individual experiences and giving therapists the broader awareness required to facilitate the session.

Currently, the system is being evaluated at schools and universities with adolescents, young adults and clinicians, to examine whether users understand the relationship between emotional representation and respiratory feedback, whether the metaphor is experienced as supportive rather than invalidating, and how therapists distribute their attention across several participants. Within the scope of this ongoing work, we will also investigate usability, comfort, perceived privacy, therapist workload, and the extent to which the activity supports emotional awareness and the practice of regulation strategies. Rather than claiming that MR or biofeedback directly regulates emotion, this work presents them as resources through which participants and therapists can jointly observe, discuss, and rehearse an emotion-regulation process.

Acknowledgments

This research was funded by MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES, PID2023-146629NA-I00 MCIU /AEI /10.13039/501100011033 / FEDER, UE and the Valencian Institute for the Enterprise Competitiveness-Generalitat Valenciana (IMAMCA/2025/11, IMAMCA/2026/11).

References

Gross, J. J. & Thompson, R. A. (2006). Emotion regulation: Conceptual foundations. In *J. J. Gross (Ed.), Handbook of Emotion Regulation*. New York: Guilford Press.

Linardon, J., Gleeson, J., Yap, K., Murphy, K. & Brennan, L. (2019). Meta-analysis of the effects of third-wave behavioural interventions on disordered eating and body image concerns. *Cognitive Behaviour Therapy* 48(1), 15-38. https://doi.org/10.1080/16506073.2018.1517389

Delinsky, S. S., & Wilson, G. T. (2006). Mirror exposure for the treatment of body image disturbance. *International Journal of Eating Disorders*, 39(2), 108–116. https://doi.org/10.1002/eat.20207

Dixon, L. J. & Linardon, J. (2020). A systematic review and meta-analysis of dropout rates from dialectical behaviour therapy in randomized controlled trials. *Cognitive Behaviour Therapy* 49(3), 181-196. https://doi.org/10.1080/16506073.2019.1620324

Pons, P., Navas-Medrano, S. & Soler-Dominguez, J. L. (2022). Extended reality for mental health: Current trends and future challenges. *Frontiers in Computer Science* 4, 1034307. https://doi.org/10.3389/fcomp.2022.1034307

Griffen, T. C., Naumann, E., & Hildebrandt, T. (2018). Mirror exposure therapy for body image disturbances and eating disorders: A review. *Clinical Psychology Review*, 65, 163–174. https://doi.org/10.1016/j.cpr.2018.08.006

Gutiérrez-Maldonado, J., Wiederhold, B. K. & Riva, G. (2016). Future directions: How virtual reality can further improve the assessment and treatment of eating disorders and obesity. *Cyberpsychology, Behavior, and Social Networking* 19, 148-153. https://doi.org/10.1089/cyber.2015.0412

Moreno-Domínguez, S., Rodríguez-Ruiz, S., Fernández-Santaella, M. C., Jansen, A., & Tuschen-Caffier, B. (2012). Pure versus guided mirror exposure to reduce body dissatisfaction: A preliminary study with university women. *Body Image*, 9(2), 285–288. https://doi.org/10.1016/j.bodyim.2011.12.001

Porras-Garcia, B., Ferrer-Garcia, M., Serrano-Troncoso, E., Carulla-Roig, M., Soto-Usera, P., Miquel-Nabau, H., ... & Gutiérrez-Maldonado, J. (2021). AN-VR-BE. A randomized controlled trial for reducing fear of gaining weight and other eating disorder symptoms in anorexia nervosa through virtual reality-based body exposure. *Journal of clinical medicine*, *10*(4), 682. . https://doi.org/10.3390/jcm10040682

Rojas-Rincón, J., Ampatzoglou, A., Gamiz-Sanfeliu, M., Fernandez-Cardellach, C., Navas-Medrano, S., Soler-Dominguez, J. L., Carballo-Marquez A., Fernandez-Capo M., Garolera M., Pons-Tomas P. & Porras-Garcia, B. (2026). Assessing the Influence of Rumination and Eating Disorder Symptoms on Feasibility Outcomes of a VR-Based Distress Tolerance Training. In *International Conference on Human-Computer Interaction* (pp. 380-396). Springer.

Soler-Dominguez, J. L., Navas, M. & Pons, P. (2024). ARCADIA: A gamified mixed reality system for emotional regulation and self-compassion. *Presented at the International Conference on Human Factors in Computing Systems*.

Skiers, K., Suen Pai, Y. & Minamizawa, K. (2022). Transcendental avatar: Experiencing a bioresponsive avatar of the self for improved cognition. *SIGGRAPH Asia 2022 XR*, 1-2. https://doi.org/10.1145/3550472.3558417

Enabling Natural Interaction with Everyday Physical Objects in a Mixed Reality Animal-Assisted Therapy Scenario

Maria Perez-Caballero, Marta Garcia-Ballesteros, Jose Luis Soler-Dominguez, Samuel Navas-Medrano, and Patricia Pons

HCI Research Group, Instituto Tecnológico de Informática (ITI), Valencia, Spain

Corresponding author: mperez@iti.es

Keywords: Mixed Reality; Object Detection; Tangible Interaction; 3D user interaction; Physical-virtual interaction

Introduction

Mixed Reality (MR) allows digital content to be integrated into the physical world, enabling users to interact with both real and virtual objects. However, creating seamless and natural interactions between these two worlds remains challenging. Systems must recognize everyday physical objects accurately and in real time, estimate their position in 3D space, determine when users are interacting with or manipulating them, and maintain stable and fluid interactions despite occlusions and tracking errors (Mine and Brooks, 1997; Dogan et al., 2024). In addition, interacting directly with physical objects can be difficult when they are partially occluded or hard to reach (Liu et al., 2025). These challenges are particularly relevant in sensitive contexts or when working with vulnerable population. This is the case for therapeutic applications, where users should be able to focus their cognitive and motor resources on the activity rather than on the interaction technique.

To address these challenges, we present a MR approach that enables users to care for a virtual dog using everyday physical objects. The proposed system combines on-device (Meta Quest 3) object detection, depth estimation, and hand tracking to automatically recognize everyday physical objects, estimate their position in the MR scene, infer when they are being manipulated, and support natural interactions with a virtual companion without requiring controllers, fiducial markers, or explicit calibration.

Related Work

Several works have explored the integration of physical objects into Extended Reality (XR) experiences to enable more natural interactions between the real and virtual worlds. Teachable Reality allows users to create tangible Augmented Reality (AR) experiences by teaching the system to recognize everyday objects and associate them with custom interactions (Monteiro et al., 2023). Similarly, Real-Virtual Objects investigates bidirectional tangible interactions in Virtual Reality (VR) through shared objects spanning the physical and virtual worlds, reporting improvements in presence and interaction experience (Bozgeyikli et al., 2024).

Computer vision has also been employed to enhance object-aware MR applications. XR-Objects automatically segments and recognizes everyday objects to provide contextual digital information

around them and trigger object-related actions (Dogan et al., 2024). However, allowed actions are based on selecting options on an augmented textual interface, rather than the manipulation of the object itself.

While these works demonstrate different ways of incorporating physical objects into XR experiences, they either require users to register objects manually beforehand (Monteiro et al., 2023), rely on physical proxies purpose-built to stand for their virtual counterparts (Bozgeyikli et al., 2024), or mediate object-related actions through augmented menus at a considerable segmentation cost, rather than through manipulation of the object itself (Dogan et al., 2024). Our proposed system leverages the colour passthrough of the Meta Quest 3 and the efficiency of a pre-trained YOLO11 detector to automatically recognize a subset of everyday physical objects in a MR experience and enable their direct interaction with virtual scene elements in real time. Because detection operates at the category level, any instance of a supported class can be used without prior registration or instrumentation, yielding a responsive MR system capable of interacting with everyday physical objects through predefined behaviours, without relying on external setups. A further consequence of this design is that, while the system encodes only atomic object-region responses and imposes no ordering upon them, users remain free to compose those responses into the broader routines in which such objects are ordinarily embedded, importing action schemas that the application itself does not represent.

Design of the approach

The proposed approach aims to complement animal-assisted therapy sessions by providing a novel MR experience in which users interact with a virtual dog using real-world objects. Rather than replacing the therapeutic role of real animals, the system offers an accessible and repeatable environment aimed to overcome logistical or financial limitations that might limit patients' access to traditional animal-assisted interventions. In this experience, which has been co-designed carefully with professional therapists, users can perform common dog-care activities, including feeding, giving water, brushing its fur, brushing its teeth, bathing, petting, and playing with the digital dog. The interaction within MR with physical objects provides the tactile feedback and motor skill engagement crucial for occupational therapy, while ensuring a highly natural an intuitive user experience.

Hence, a key design objective was to make these interactions as natural as possible. Users manipulate everyday physical objects directly, without requiring markers, controllers, or additional calibration steps. The system automatically recognizes the objects and enables their interaction with the virtual dog, allowing users to focus on the therapeutic activity itself instead of the underlying technology. This design minimizes interaction overhead, allowing users to focus on caring for the virtual dog rather than on the underlying interaction technology.

This design decision is grounded in the framework of Reality-Based Interaction (Jacob et al., 2008), which argues that interfaces leveraging users' pre-existing knowledge of the real world, such as naïve physics and body awareness, reduce cognitive overhead and produce more intuitive interactions. It also aligns with the theory of Embodied Interaction (Dourish, 2001), which holds that meaning and usability emerge from users' direct, embodied engagement with physical artifacts rather than through abstract or symbolic controls.

Object Detection

The application relies on a YOLO11n object detection model from Ultralytics (Jocher et al., 2024), executed locally on the Meta Quest 3 using Unity Sentis (Figure 1). YOLO11n was selected as it provides a suitable trade-off between inference speed and detection accuracy for on-device execution. The model was fine-tuned to recognize the everyday objects required for the intervention. The training dataset combines images from the LVIS dataset (Gupta et al., 2019), Open Images V7 (Kuznetsova et al., 2020), and a custom dataset collected specifically for this application, containing the dog-care objects used during the intervention. The fine-tuned model achieved an average precision of 99.63% across the eight target object classes. During execution, RGB frames captured through the Meta Quest 3 passthrough cameras are processed by the detector to obtain 2D object

detections. The object depth is estimated using the Meta XR Environment Depth Manager, enabling the detected objects to be projected into Unity world coordinates. These 3D coordinates are subsequently used to support interactions between the detected physical objects and the virtual dog.

Interaction with Physical Objects

Interaction with physical objects is carried out using hand tracking, without the need for controllers. Users interact with real-world objects by reaching them, picking them up, moving them and releasing them as part of the MR experience. To determine when a user is attempting to grab an object, the system combines multiple input signals. The hand tracking capabilities provided by the Meta XR SDK are used to continuously track the user's hands and estimate their pose and position, while additional information, such as the distance between the hand and the object and the stability of the movement, is computed by the interaction system. These signals are combined with the hand pose and temporary loss of visibility that may occur when the hand covers the object. Using these cues together helps to reduce accidental interactions and makes the picking-up process more tolerant to brief tracking interruptions. To increase robustness and reduce false detections, the conditions for a potential grasp must remain valid for a minimum number of consecutive frames before an interaction is confirmed. Similarly, short periods in which the object or the hand cannot be reliably detected, for example due to occlusion during grasping, are tolerated by maintaining the interaction state for a limited amount of time. These temporal thresholds make the picking-up process less sensitive to tracking noise and brief visibility losses while reducing accidental interactions.

Once the object has been picked up, an invisible virtual counterpart within the MR experience is linked to the user's hand and follows its movement. This virtual counterpart enables the real object to interact with the virtual dog (Figure 2). When the user releases the object, the virtual representation is removed.

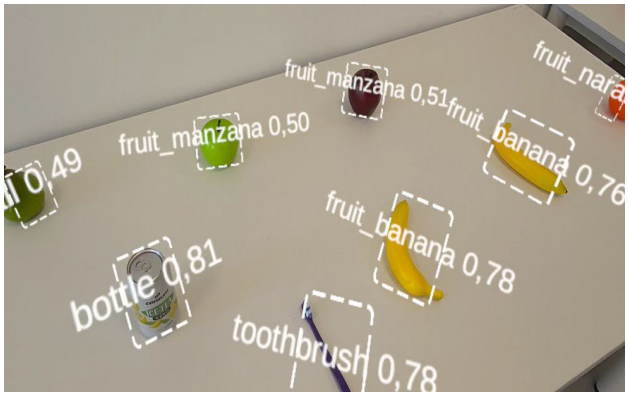


***Figure 84**. Object detection with YOLO11n running on Meta Quest 3*



***Figure 85**. User interacting with real bottle and virtual dog*

Interaction with Virtual Dog

Each tracked physical object is linked to a pre-defined set of actions that determinates how it can be used to interact with the virtual dog. For example, a brush can be used for grooming, fruit can be used to feed the dog, and a toothbrush can trigger a toothbrushing interaction. Usars can also interact directly with the animal by petting it with an empty hand.

The interaction is spatially contextual, meaning that the system considers not only which object is being used, but also where it is applied to the dog's body. The dog is divided into different interaction areas, each of which supports a specific set of actions. An interaction is therefore triggered only when the action associated with the object is compatible with the area being contacted. This location-based behaviour also allows a single object to be used in more than one manner. For instance, when a bottle or glass is brought close to the dog's mouth, the dog performs a drinking

action, while moving the same object towards the dog's body produces water-droplet effects that simulate wetting the animal. When a valid interaction is detected, multimodal feedback is triggered. Depending on the action, this response may include an animation of the 3D model, a sound and specific particle effects, such as water splashes or foam. The dog performs one interaction at a time. While an action is being executed, new actions are temporarily prevented from interrupting it. Once the animation and associated feedback have finished, the dog returns to an idle state and becomes available for another interaction.

In addition to reacting to user-initiated actions, the dog exhibits a limited range of autonomous behaviours. It can follow the user around the physical environment, sit when nearby and bark occasionally. These behaviours help the animal maintain an active presence in the scene, even when the user is not directly interacting with it.

Discussion and Conclusion

This work presents a MR system that combines on-device object detection, depth estimation, and hand tracking to enable natural interactions between everyday physical objects and a virtual therapy dog. Rather than introducing new interaction devices or requiring modifications to physical objects, the proposed approach allows users to interact with the virtual companion using familiar objects that are already part of the therapeutic activities. This allows users to focus on interacting with the virtual dog naturally, rather than on how to perform the interaction. The proposed system demonstrates how everyday physical objects can be incorporated into Mixed Reality experiences without requiring markers, dedicated controllers, or instrumented objects. By automatically recognizing and tracking real objects, the interaction becomes part of the activity itself rather than an additional task that users must learn.

The current implementation is limited to a predefined set of everyday objects and interactions, and its performance depends on the accuracy of object detection, depth estimation, and hand tracking. Future work will evaluate the robustness of the interaction under different environmental conditions and with different user populations. In addition, user studies will investigate the usability, perceived naturalness, and therapeutic value of the proposed approach, as well as its potential to complement existing animal-assisted therapy sessions. Building upon this work, the proposed approach will be extended towards more adaptive and user-aware XR interaction. The multimodal integration of enhanced hand and gesture tracking, eye-tracking, and physiological information will provide the system with a richer understanding of the user's state and interaction patterns. These additional sensing modalities will enable the experience to evolve from a primarily reactive interaction paradigm towards a more adaptive one, in which the behaviour of the virtual companion and the interaction dynamics can be progressively tailored to the user and to the way they engage with the MR environment.

Acknowledgments

We would like to thank the mental health professionals and participants from Fundación SASM and Fundación ACAVALL for their valuable insights and feedback. This work has been funded by the IVACE/FEDER funds through the project CUIDAR (IMDEEA/2025/36) and by the European Union under the Horizon Europe research and innovation programme (IRIS-XR, Grant Agreement no. 101298672).

References

Bozgeyikli, L., Zhao, Y., Kim, M. & Bae, S. (2024). Real-Virtual Objects: Exploring Bidirectional Embodied Tangible Interaction with a Virtual Human in World-Fixed Virtual Reality. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2024).

Dogan, M. D., Wilson, G., Izadi, S. et al. 2024. Augmented Object Intelligence with XR-Objects. In Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST 2024).

Dourish, P. (2001). Where the action is: The foundations of embodied interaction. *MIT Press*

Gupta, A., Dollar, P. & Girshick, R. (2019). LVIS: A Dataset for Large Vocabulary Instance Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR 2019), 5356–5364. https://doi.org/10.1109/CVPR.2019.00550

Jacob, R. J. K., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., & Zigelbaum, J. (2008). Reality-based interaction: A framework for post-WIMP interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 201–210). ACM. https://doi.org/10.1145/1357054.1357089.

Jocher, G., Chaurasia, A. & Qiu, J. et al. (2024). Ultralytics YOLO11. Ultralytics. Available: https://github.com/ultralytics/ultralytics (accessed July 2026).

Kuznetsova, A., Rom, H., Alldrin, N. et al. (2020). The Open Images Dataset V7. *International Journal of Computer Vision*, 128(7), 1956–1981.

Liu, Z., He, Z., Lee, G. A., Billinghurst, M. & Bai, H. (2025). Reality Proxy: Fluid Interactions with Real-World Objects in Mixed Reality via Abstract Representations. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (UIST 2025). https://doi.org/10.1145/3746059.3747709

Mine, M. R. & Brooks Jr., F. P. (1997). Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1997)*, 19–26. https://doi.org/10.1145/258734.258747

Monteiro, K., Vatsal, R., Chulpongsatorn, N., Parnami, A. & Suzuki, R. (2023). Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI 2023), Article 633, 1–15. https://doi.org/10.1145/3544548.3581449

Mosquera, J., Hernández, J., García, D. et al. (2025). A Study of Performance and Interaction Patterns in Hand and Tangible Interaction for Mixed Reality. arXiv:2511.11962.

Synthetic Environment for Navigation, Simulation and Operations for Robotics

Alona Stulii, and Maryna Tymchenko

Sensorama Lab

Corresponding author: ast@sensoramalab.com

Keywords: Training simulator, extended reality (XR), UAV, digital-twin, AI.

Abstract

This paper introduces SENSOR, a developing high-fidelity simulator that combines desktop-based virtual reality, environment-specific digital twins, and operational fleet management for Unmanned Aerial Vehicle (UAV) autonomy training and coordinated mission rehearsal. Positioned within the broader extended reality (XR) landscape, SENSOR enables users to operate virtual UAVs, explore realistic 3D environments, and supervise autonomous missions through an interactive desktop interface. The platform supports both direct human control and supervisory interaction, allowing users to inspect UAV behaviours under configurable environmental and operational conditions. Future development will focus on performance optimisation and extending the existing virtual environment to additional XR interfaces.

SENSOR addresses the fragmentation of civil Unmanned Aerial Vehicle simulation and operational capabilities across research simulators, commercial fleet, and digital-twin tools, enabling coordinated multi-UAV mission execution. This fragmentation occurs at both architectural and functional levels, Airspace management systems are often implemented in incompatible ways across platforms, such as one provider's fixed-route design versus another's dynamic, grid-based airspace structure, which limits transferability of techniques and slows progress toward a standardized architecture, as mentioned by Li et al., (2026).

SENSOR is a physical-AI suite for UAVs that combines a high-fidelity autonomy training simulator with an operational fleet-management system. UAVs are trained on digital twins of specific environments. The use of digital twins can connect virtual system models with control development and hardware-in-the-loop (HITL) validation, supporting the transition from simulation to operational implementation according to Viola & Chen (2023). We build realistic 3D models of logistics corridors and industrial sites, together with vehicle and sensor models for multirotor drones and simple but explicit models of risk factors such as wind, GPS loss, link loss and moving obstacles. Instead of building one-off solutions, we combine a high-fidelity simulator, HITL setups and onboard AI to “teach” UAVs to detect relevant objects, fly to them and perform defined actions, all coordinated by a distributed fleet management system that keeps operators in a supervisory role rather than manually flying each mission.

We systematically evaluate guidance, perception, and safety behaviours under realistic environmental conditions — including wind, dynamic obstacles, and communication failures — prior

to real-world deployment. SENSOR pre-trains object recognition, safe-landing-zone detection, path tracking and behaviour trees. We adopt a simulation-first methodology, as Perez-Segui et al. (2023) noted, this staged approach makes testing safer and more efficient, since problems can be caught and fixed in simulation before any real-world flights take place. This also reduces the number of physical test flights needed, and thus lowers energy use, emissions, noise and disturbance compared to only field-based testing.

The key contributions are to treat autonomy as a configurable product, connect this autonomy directly to operations, and integrate it with industrial workflows. SENSOR allows the end users to pre-train UAVs to perform specific missions with precision while keeping humans in control of training and validation decisions. Through an XR approach based on desktop virtual reality, users can inspect and refine autonomous behaviour patterns, supervise tests, and rehearse missions in a high-fidelity digital-twin environment before progressing to real-world field testing and operational deployment

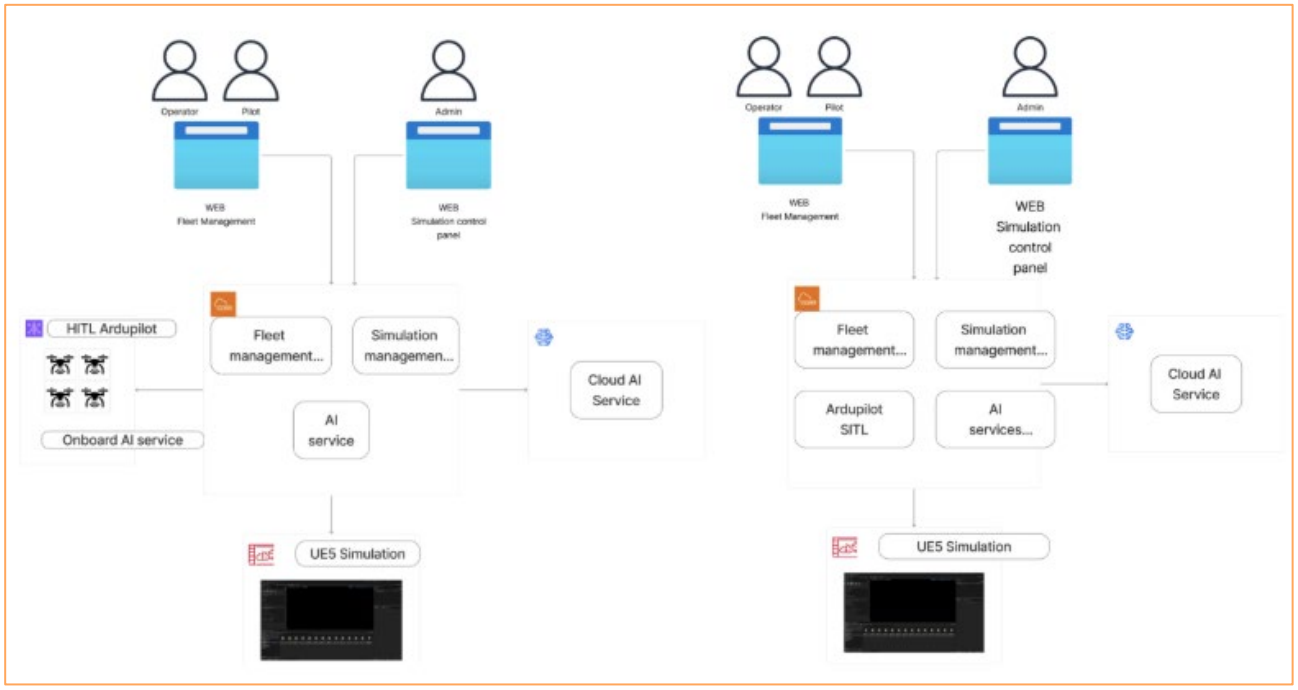
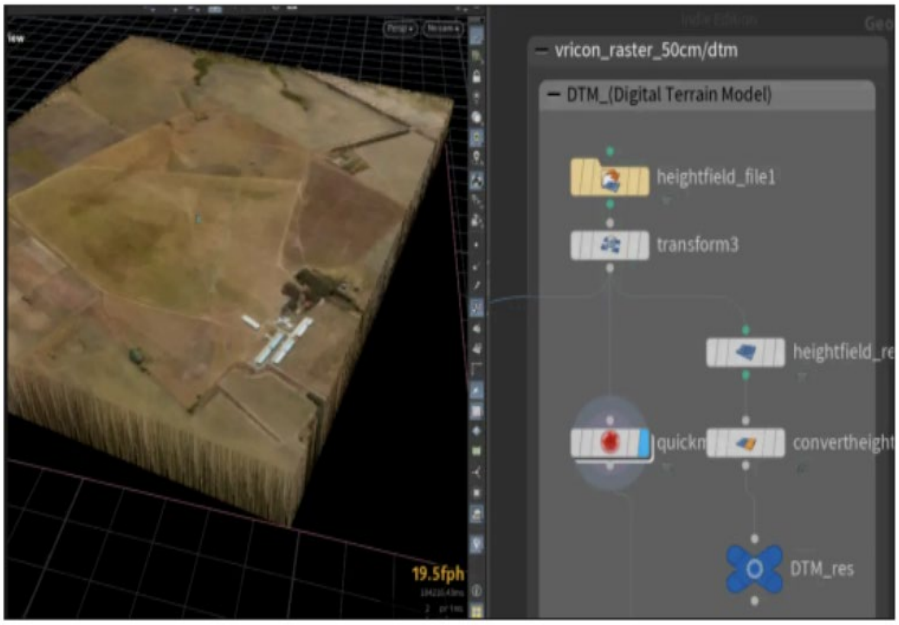


Figure 1. *Draft scheme of SENSOR with connections between simulator, fleet management, AI and UAVs (HITL - on the left side and SITL - on the right side).*



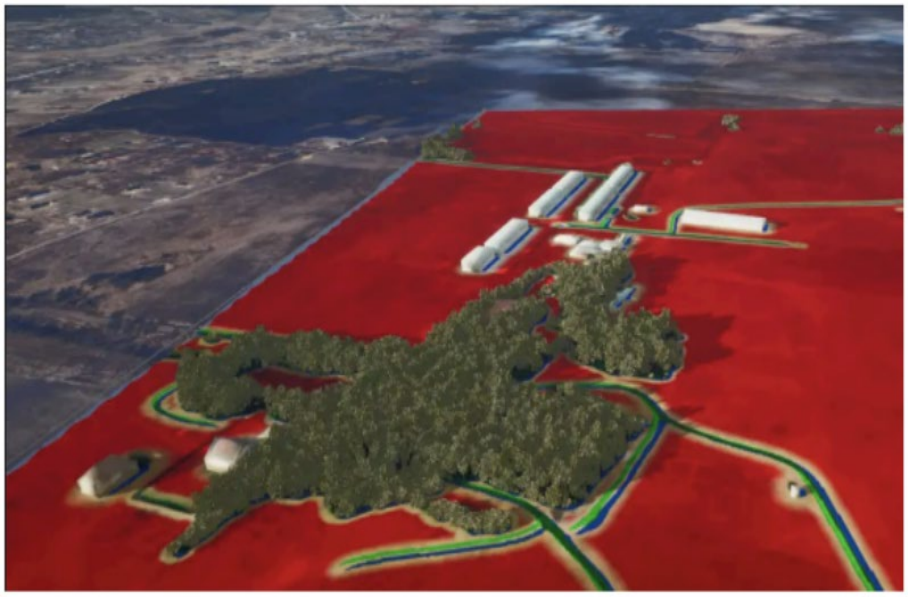


Figure 2 *and* ***Figure 3****. Import of satellite images (left side) and grass and forestry generation in Unreal Engine through mask (right side) to create up-to-date realistic environments for training.*

References

Li, G., Han, X., Shi, Y., & Wang, L. (2026). Bridging the Fragmentation in Unmanned Aircraft System Traffic Management (UTM): A Systematic Survey on UTM. *Drones 2026*, 10(5), 377, (pp.2-3) https://doi.org/10.3390/drones10050377

Viola, J., & Chen, Y. Q. (2023). Digital-Twin-Enabled Smart Control Engineering. *Synth. Lect. Eng. Sci. Technol. Springer 2023. (pp. 74-76)*

Perez-Segui, R., Arias-Perez, P., Melero-Deza, J., Fernandez-Cortizas, M., Perez-Saura, D., & Campoy, P. (2023). Bridging the gap between simulation and real autonomous uav flights in industrial applications. *Aerospace*, *10*(9), 814. https://doi.org/10.3390/aerospace10090814

Perceptual Limits of Avatar Movement Distortion in Embodied VR

Tony Donegan, Beñat Amestoy-Alonso, Edgar Rodriguez de la Calle, and Maria V. Sanchez-Vives

IDIBAPS, Barcelona, Spain

Corresponding author: donegan@recerca.clinic.cat

Keywords: Virtual reality; embodiment; visuomotor gain; psychophysics; agency; social VR

Abstract

Virtual reality systems frequently alter the mapping between physical and avatar movement to extend reach, reduce effort, or shape motor behaviour. Such manipulation should remain sufficiently undetectable to preserve agency. We quantified perceptual limits for gain-modified shoulder abduction in an embodied avatar. Twenty-four adults completed a two-alternative forced-choice task while avatar movement was modified across ten gain levels from −50% to +50%. A logistic psychometric function was fitted to 2,199 recorded trials, with uncertainty estimated by participant-level bootstrapping. The point of subjective equality (PSE) occurred at +3.53% amplification (95% CI −0.04% to 7.58%), suggesting a tendency for slightly amplified avatar movement to be experienced as veridical. The 25% and 75% points defined absolute operational boundaries of −6.09% and +13.15% relative to a mathematically veridical mapping. These findings identify a visuomotor safe zone shifted toward amplification. The limits may inform ergonomics, rehabilitation and control of embodied avatars in social VR experiments.

Introduction

A convincing sense of embodiment in virtual reality (VR) depends partly on congruence between physical movement and the movement seen in an avatar. When visual, proprioceptive and motor signals align, users generally experience the avatar as their body and its actions as self-generated (Kilteni et al., 2012). Nevertheless, many useful VR techniques intentionally alter this mapping. Control-display gain can extend virtual reach, reduce physical effort, redirect movements away from obstacles, simulate weight or resistance, and influence motor performance (Asmandian et al., 2016; Samad et al., 2019; Bourdin et al., 2019; Kilteni et al., 2012).

The practical challenge is to determine how far the mapping can depart from mathematical veridicality before the discrepancy becomes noticeable. Detection limits have been established for redirected walking, hand redirection and fine-motor interaction (Samad et al., 2019; Steinicke et aö., 2010), but evidence for gross upper-limb movements remains limited. Shoulder movement is particularly relevant to large-scale interaction, exercise and rehabilitation, and to embodied social situations in which spontaneous gestures may form part of the behavioural outcome.

We therefore estimated the point of subjective equality (PSE) and operational detection boundaries for gain-modified shoulder abduction. We hypothesised that the perceptually veridical

point might be shifted away from a strict 1:1 mapping, creating a safe operating zone that is displaced toward amplification.

Materials and methods

Participants and design. Twenty-four healthy adults aged 18-45 years took part after providing informed consent. A priori calculation for a repeated-measures medium effect (f=0.25, alpha=0.05, power=0.90) indicated at least 17 participants. The study followed institutional ethical requirements.

Apparatus and embodiment. Participants stood while wearing an Oculus Quest head-mounted display (1440 × 1600 pixels per eye; 72 Hz) and holding the controllers. They were embodied in a gender-matched Rocketbox avatar in Unity 2019.4.10f1 using congruent visuomotor and visuotactile stimulation. A mirror supported observation of the virtual body.

Task and stimuli. Participants repeatedly abducted one shoulder toward 90° while observing the avatar. Avatar displacement was modified using ten gain conditions (−50% to +50% in 10% increments, excluding 0%). Gain was introduced gradually: no distortion was applied below 20°, it increased linearly to the target value at 90° and then remained constant. The protocol comprised 100 randomised trials per participant; 2,199 recorded trials available in the supplied dataset were analysed. After each movement, participants indicated whether the displayed movement was greater or smaller than their physical movement.



Figure 1. *Experimental environment. Participants observed the embodied avatar in a virtual mirror while abducting the shoulder and judging the direction of the movement discrepancy.*

Analysis. For each gain level, the proportion of trials in which physical movement was judged greater than avatar movement was calculated. A two-parameter logistic psychometric function was fitted by nonlinear least squares. The PSE was the gain corresponding to a response probability of 0.50. The absolute operational limits were the gains corresponding to probabilities of 0.25 and 0.75. These values describe boundaries relative to the mathematically veridical 0% mapping; distances from the fitted PSE index perceptual precision. Confidence intervals were estimated using 5,000 participant-level bootstrap resamples, preserving the clustering of trials within individuals.

Results

The logistic function closely described the relationship between movement gain and forced-choice responses. The estimated PSE was +3.53% (participant-bootstrap 95% CI −0.04% to 7.58%). Thus, on average, participants required modest amplification of avatar movement before physical and displayed movements were judged equally often. Because the confidence interval narrowly included zero, this should be interpreted as an estimated positive bias or tendency rather than definitive evidence that the population PSE differs from veridical mapping.

The 25% and 75% response points occurred at −6.09% (95% CI −10.82% to −1.55%) and +13.15% (95% CI 8.70% to 17.91%), respectively. These values delimit an operationally useful zone around the 0% mapping (but do not demonstrate greater perceptual sensitivity to reductions).

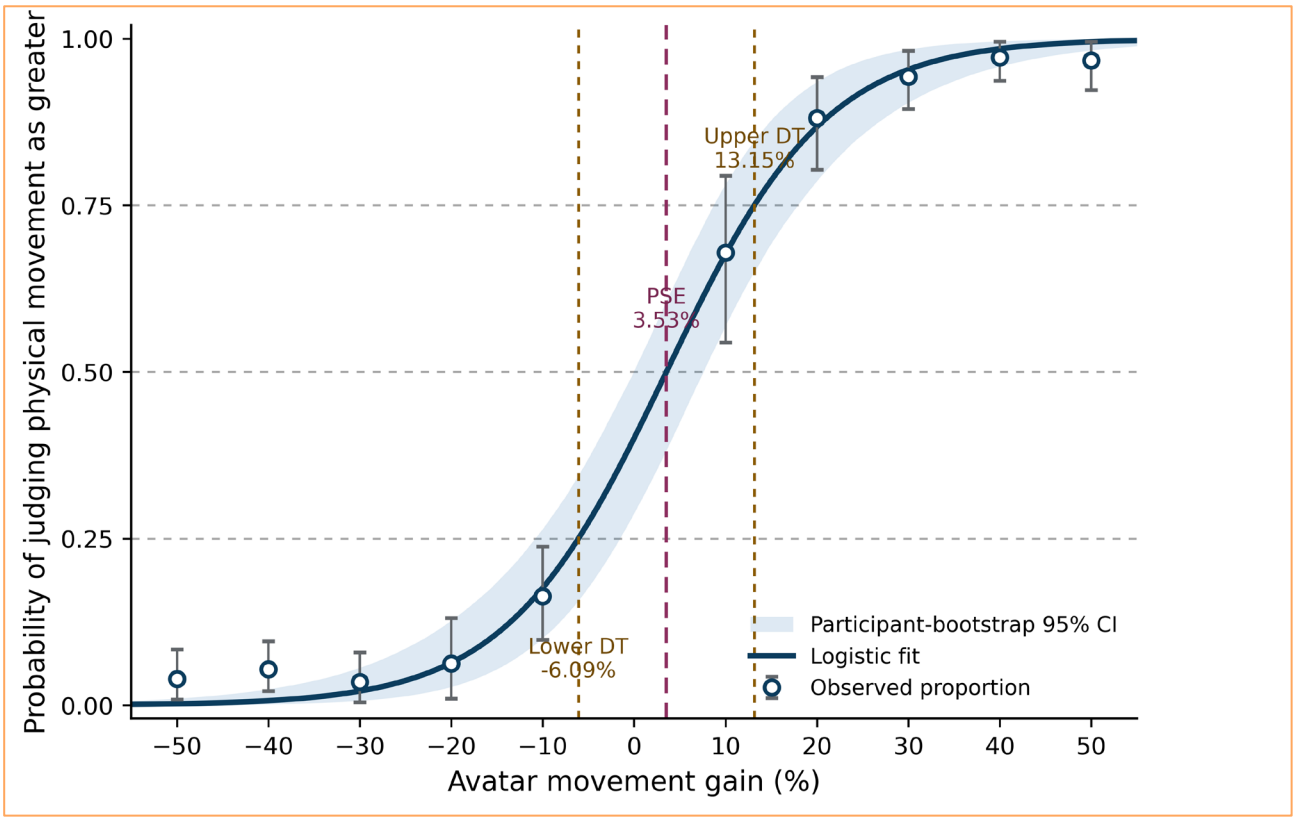


Figure 2. *Psychometric function for avatar movement gain. Points show pooled observed proportions. Error bars and the shaded band are participant-level bootstrap 95% confidence intervals. Vertical lines mark the PSE and the absolute 25% and 75% operational boundaries.*

Discussion

The principal finding is a shifted perceptual equivalence point. A mathematically veridical mapping was not the centre of the fitted response function; instead, approximately 3.5% avatar amplification was judged equivalent to physical movement. This tendency is compatible with evidence that agency attribution can be shaped by favourable action outcomes (Yoshie & Haggard, 2013) a slightly enhanced virtual action may be readily incorporated into the user's motor prediction and experienced as self-generated. It may also reflect properties of embodied VR, including imperfect proprioceptive matching, avatar geometry or reduced visual access to the physical limb. These explanations remain theoretical because agency, preference and mechanism were not measured directly.

The results define a shifted safe zone rather than uneven sensory acuity. Relative to a strict 0% mapping, the fitted operational boundaries extended from approximately 6% reduction to 13% amplification. Designers therefore have more room to introduce amplification before reaching the upper boundary, but this follows from the displaced PSE rather than superior or inferior discrimination on either side. Centering an application near the estimated PSE could permit approximately 9.6% modification in either direction before reaching 25% 75%.

For ergonomic and rehabilitation applications, modest amplification within this zone could extend reach or reduce effort while maintaining a credible relationship between physical and avatar movement. Conversely, reductions relative to veridical mapping approach the lower boundary sooner and should be introduced conservatively when remaining unnoticed is important. The reported values are task-specific design estimates, not universal limits: movement direction, speed, joint configuration, avatar appearance, exposure duration and user population could all alter the PSE or slope.

Implications for social VR

Embodied social-VR studies often treat spontaneous movement, interpersonal distance, gesture or response latency as behavioural outcomes. A noticeable mismatch between physical and avatar movement could reduce agency or presence and thereby confound those measures. Calibrating avatar control within an empirically defined visuomotor zone may therefore strengthen the validity of experiments involving sensitive social interactions.

This issue is relevant to META-TOO, a Horizon Europe project investigating gender-based inappropriate social interactions and bystander behaviour in the metaverse (MeTa-TOO. 2026). In scenarios where participants witness harassment and may intervene through an embodied avatar, credible control is important for interpreting whether response timing and action reflect the social situation rather than properties of the interface. The present limits offer a starting point for avatar calibration; future related work could test whether visuomotor congruence directly influences intervention latency, emotional engagement and perceived responsibility.

Conclusion

Shoulder movement gain in embodied VR showed an estimated PSE of +3.53%, defining absolute operational limits of −6.09% and +13.15% relative to veridical mapping. The key design implication is a safe zone shifted toward amplification, with similar fitted precision around the PSE. Participant-level uncertainty and task specificity should be considered when applying these values to rehabilitation, ergonomic interaction or social VR.

References

Azmandian, M., Hancock, M., Benko, H., Ofek, E., & Wilson, A. D. (2016, May). Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In Proceedings of the 2016 chi conference on human factors in computing systems (pp. 1968-1979). https://doi.org/10.1145/2858036.2858226

Bourdin, P., Martini, M., & Sanchez-Vives, M. V. (2019). Altered visual feedback from an embodied avatar unconsciously influences movement amplitude and muscle activity. Scientific reports, 9(1), 19747.

Kilteni, K., Groten, R. & Slater, M. (2012). The sense of embodiment in virtual reality. Presence 21(4), 373-387. https://doi.org/10.1162/PRES_a_00124

Kilteni, K., Normand, J. M., Sanchez-Vives, M. V., & Slater, M. (2012). Extending body space in immersive virtual reality: a very long arm illusion. PloS one, 7(7), e40867. https://doi.org/10.1371/journal.pone.0040867

META-TOO Consortium. (2026). A transfer of knowledge and technology for investigating gender-based inappropriate social interactions in the Metaverse. https://meta-too.eu/

Samad, M., Gatti, E., Hermes, A., Benko, H., & Parise, C. (2019, May). Pseudo-haptic weight: Changing the perceived weight of virtual objects by manipulating control-display ratio. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-13). https://doi.org/10.1145/3290605.3300550.

Steinicke, F., Bruder, G., Jerald, J., Frenz, H., & Lappe, M. (2009). Estimation of detection thresholds for redirected walking techniques. IEEE transactions on visualization and computer graphics, 16(1), 17-27. https://doi.org/10.1371/journal.pone.0040867

Yoshie, M., & Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. Current Biology, 23(20), 2028-2032.

Floor-projected UI for hands-free interaction with AGVs in intralogistics processes

Sebastian Knopp[1], David Reese[2], and Martin Dix[1]

[1]*Chemnitz University of Technology, Professorship for Production Systems and Processes, Chemnitz, Germany,*

[2]*Zentrum für Bild- und Signalverarbeitung e. V, Ilmenau, Germany*

Corresponding author: sebastian.knopp@mb.tu-chemnitz.de

Keywords: Spatial augmented reality, floor projection, human–robot interaction, automated guided vehicle, intralogistics, projection surface selection

Introduction

Automated guided vehicles (AGVs) are established in intralogistics, but interaction between a picker and the vehicle is still mediated by hardware that sits somewhere other than where the work happens: a terminal on the vehicle, a handheld scanner, or a station at the end of the aisle. Touch interfaces require the worker to approach the vehicle and free a hand; head-worn AR displays avoid this but add donning time, hygiene concerns and per-worker cost. Floor projection offers a third path: the interface appears in the shared floor space where the interaction takes place, needs no worn or handheld device, and can be operated contactlessly.

The difficulty is that a projector mounted on a moving vehicle has no guarantee about the surface it projects onto. A vehicle stopping wherever its transport task ends may face a pallet, a rack leg or a wall, leaving the interface truncated or cast onto an obstacle. This paper reports the design and implementation of *Hubert*, a retrofit of an existing intralogistics AGV with a floor projection subsystem and a positioning algorithm that guarantees an unoccluded, reachable and correctly oriented projection area.

Related Work

Projector-based spatial augmented reality is a long-established alternative to worn displays (Bimber & Raskar 2005), applied in industry mainly at static workstations such as projected assembly instructions (Sand et al. 2016), where surface geometry and illumination are known at design time. On mobile robots, it has served chiefly to communicate intent rather than to provide an interactive interface: Matsumaru (2006) projected forthcoming motion onto the floor ahead of a robot, and Chadalavada et al. (2015, 2020) projected planned paths and safety zones onto shared floor space, reporting improved subjective ratings of reliability, predictability and transparency. Across this work, the projection target is treated as given: defensible for intent signalling, but not for an interactive interface, where a truncated button is a functional

failure. Selecting the projection surface as an explicit, continuously solved sub-problem of the navigation stack has, to our knowledge, not been addressed, yet.

System design

The base is a pre-existing AGV developed in a preceding project for intralogistics transport (Allmacher et al. 2018), able to follow a person carrying a UWB transmitter, travel autonomously to given coordinates, or be remote controlled. Localisation and mapping use an implementation of Hector SLAM (Kohlbrecher et al. 2011) adapted to the platform, based on the vehicle's existing SICK S300 LIDAR at 20 cm height with a 270° field of view. The surface selection described below reuses this scanner instead of adding one, keeping the retrofit inexpensive and transferable to other vehicles of this class (Table 1). The additionally implemented Azure Kinect is used solely for foot detection to enable the UI interaction.

On this existing system an Epson EB-L200SW (3,800 lm, WXGA 1280 × 800) and an Azure Kinect were added. The projector retrofit constrains the electrical system: the vehicle supplies 24 V from two 12 V 40 Ah $LiFePO_4$ batteries while the projector requires 230 V mains, so an inverter stage was added alongside the existing rails; dimensioning this supply chain is the critical design point of such a retrofit.

The Azure Kinect is used exclusively to enable the UI interaction. The projection area is extrinsically calibrated relative to the Kinect, which enables a perspective-corrected (keystone) projection of UI elements onto the surface and establishes a common spatial reference between the Kinect's depth data and the projected image (Figure 1). The calibration procedure as well as a custom image-processing pipeline for user detection, tracking, and feet segmentation are implemented by Zentrum für Bild- und Signalverarbeitung (ZBS) e. V. The UI design and development were carried out by the compl3te GmbH. All modules communicate over ZeroMQ in a publish–subscribe pattern, which keeps the projection subsystem, the interface renderer and the vehicle control loosely coupled.

Occlusion-free projection surface selection

An early design question was whether the projection area should be steered independently of the vehicle. Shifting the projected image within the projector's reachable area moves the interface out of the Kinect's sensing volume, and rotating the projector and camera together requires a superstructure that obstructs loading. Neither offers an advantage over rotating the vehicle on the spot, so the projection area is adjusted through vehicle orientation.

Selection runs continuously during travel rather than only at the stopping point, so a valid target is available the moment the vehicle stops. Scanner data are accumulated into a rolling 2D occupancy grid (Elfes 1989) centred on the vehicle. Candidate poses are sampled on concentric shells around the current position in 10° increments with 30 cm radial spacing, yielding 20 candidates per cycle. For each, a virtual footprint representing the vehicle together with its projection area is tested against the grid, and the candidate is discarded if any cell under the projection footprint is occupied — the condition guaranteeing an unoccluded interface. Survivors are ranked by an additive cost over distance and orientation relative to the user. The area directly ahead is the preferred default; only when it is invalid does the algorithm select from the wider neighbourhood and plan a route, which in the large majority of cases is a rotation on the spot. A full cycle executes in approximately 45 ms, well below the vehicle's control cycle.

Two exceptions proved necessary in operation. First, once the surface has been finally selected, subsequent blocking does not trigger a further corrective motion: it is almost always caused by the user, and moving would produce an interface that retreats from the person using it. Second, collisions with the legs of nearby people are ignored, since the user necessarily stands beside the interface they operate; the leg_detector module from the ROS ecosystem, based on boosted features in 2D range data (Arras et al. 2007), was adapted so that leg-shaped returns are excluded from the test.

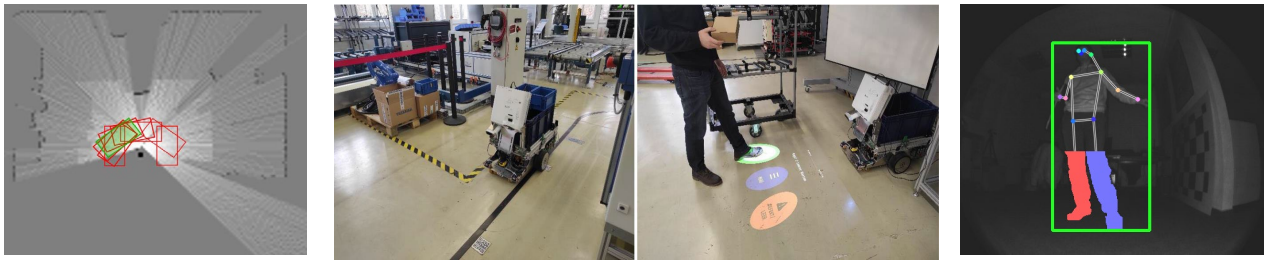


Figure 86. *Rolling 2D occupancy grid with candidate projection areas. Rejected candidates in red, selected area in green(left). The demonstrator in an experimental factory environment (mid-left) and the projected interface in use (mid-right). Detection view of the user and their leg gestures through the Azure Kinect (right).*

Interface design. The floor is an uncontrolled display surface, and constrains the interface more than the projector specification does: finishes range from matte concrete to glossy epoxy and painted markings, and hall illuminance may reach 500–750 lx, so the interface uses large, high-contrast colour fields and generously separated interaction zones. Assessment of the substrate itself remained out of scope, as a 2D scanner at 20 cm height cannot distinguish a suitable floor from a reflective or patterned one. Occlusion by three-dimensional obstacles is handled, radiometric suitability is not; a photometric layer in the grid would close this gap, and the ranking function already accepts an additional penalty term.

Validation

The surface selection was evaluated in a virtual and hybrid commissioning environment (Allmacher et al. 2019) in which the real control software drives a simulated vehicle through reconstructed intralogistics environments. Twelve scenarios were used, covering long rack aisles, aisle intersections and free-standing obstacles. Across 240 runs, a valid unoccluded projection area was found in 97.5% of cases; the remaining failures occurred in the narrowest aisle configuration, where no pose admits the full projection footprint and the vehicle correctly reports that the interface cannot be displayed. Mean corrective rotation was 16° and mean latency from reaching the target pose to display was 0.8 s, dominated by the corrective motion rather than by computation.

All modules were also integrated on the physical vehicle and validated in the laboratory, and the complete interaction sequence was demonstrated in an experimental factory environment (Figure 1). These results establish functional feasibility; quantitative evaluation under real operating conditions and a user study are the immediate next steps.

Conclusion and outlook

We presented a projection retrofit for an existing intralogistics AGV for hands-free human-machine interaction, in which the projection surface is treated as an explicit, continuously solved sub-problem rather than an assumption. The approach reuses the vehicle's existing LIDAR scanner, runs within the existing control cycle and resolves surface conflicts by a minimal rotation of the vehicle rather than by additional actuated hardware. The exceptions built into the occlusion criterion proved as important to usable behaviour as the criterion itself. Future work targets radiometric suitability of the substrate, evaluation in a production environment, and a user study against terminal-based and head-worn alternatives.

Acknowledgements

This work was carried out in the joint project KARIFF ("Kontextsensitive Augmented-Reality-Interaktion mit Fahrerlosen Transportfahrzeugen"), funded by the German Federal Ministry for Economic Affairs and Energy within the Central Innovation Programme for small and medium-sized

enterprises (ZIM) under grant no. 16KN055775, on the basis of a decision by the German Bundestag.

References

Allmacher, C., Dudczig, M., Klimant, P. & Putz, M. (2018). Virtual prototyping technologies enabling resource-efficient and human-centered product development. *Procedia Manufacturing* 21, 749–756. https://doi.org/10.1016/j.promfg.2018.02.180

Allmacher, C., Dudczig, M., Knopp, S. & Klimant, P. (2019). Virtual reality for virtual commissioning of automated guided vehicles. *IEEE Conference on Virtual Reality and 3D User Interfaces* (VR), 838–839. https://doi.org/10.1109/VR.2019.8797981

Arras, K.O., Mozos, O.M. & Burgard, W. (2007). Using boosted features for the detection of people in 2D range data. *IEEE International Conference on Robotics and Automation* (ICRA), 3402–3407. https://doi.org/10.1109/ROBOT.2007.363998

Bimber, O. & Raskar, R. (2005). Spatial Augmented Reality: Merging Real and Virtual Worlds. *Wellesley: A K Peters*.

Chadalavada, R.T., Andreasson, H., Krug, R. & Lilienthal, A.J. (2015). That's on my mind! Robot to human intention communication through on-board projection on shared floor space. *European Conference on Mobile Robots* (ECMR), 1–6. https://doi.org/10.1109/ECMR.2015.7403771

Chadalavada, R.T., Andreasson, H., Schindler, M., Palm, R. & Lilienthal, A.J. (2020). Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human–robot interaction. *Robotics and Computer-Integrated Manufacturing* 61, 101830. https://doi.org/10.1016/j.rcim.2019.101830

Elfes, A. (1989). Using occupancy grids for mobile robot perception and navigation. *Computer* 22(6), 46–57. https://doi.org/10.1109/2.30720

Kohlbrecher, S., von Stryk, O., Meyer, J. & Klingauf, U. (2011). A flexible and scalable SLAM system with full 3D motion estimation. *IEEE International Symposium on Safety, Security, and Rescue Robotics* (SSRR), 155–160. https://doi.org/10.1109/SSRR.2011.6106777

Matsumaru, T. (2006). Mobile robot with preliminary-announcement and display function of forthcoming motion using projection equipment. *IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN), 443–450. https://doi.org/10.1109/ROMAN.2006.314368

Sand, O., Büttner, S., Paelke, V. & Röcker, C. (2016). smARt.Assembly – Projection-based augmented reality for supporting assembly workers. *International Conference on Virtual, Augmented and Mixed Reality* (VAMR), LNCS 9740, 643–652.

From Automated Guided Vehicle to Telepresence Robot: Retrofitting an In-House AGV for VR-Controlled Telepresence

Sebastian Knopp, and Martin Dix

Chemnitz University of Technology, Professorship Production Systems and Processes, 09126 Chemnitz, Germany

Corresponding author: Sebastian.knopp@mb.tu-chemnitz.de

Keywords: Telepresence robot, virtual reality, automated guided vehicle, teleoperation, system integration, retrofitting

Introduction

Telepresence robots (TPR) allow a remote person to move through a distant physical space and take part in what happens there (Kristoffersson et al. 2013). Studying how that experience arises means varying the things that influence it, such as camera position, information presented on the the displays the vehicle's steering behaviour , and the level of control available to the person on site . Commercially available platforms do not permit this. They are black boxes: the operating interface is fixed, the camera position is given, and additional hardware cannot be integrated or instrumented.

The project TeleVeRbundenheit (2023–2026) investigates how virtually extended telepresence can create a sense of social connectedness across distance (Franzkowiak et al. 2024; Legler et al. 2025). It needed a platform on which every subsystem is accessible for modification, measurement and validation — a research instrument, not a product. We therefore converted an automated guided vehicle (AGV) developed in-house in an earlier project into a telepresence robot driven entirely from a virtual reality (VR) environment (Figure 1). This contribution reports that conversion. The work was shared between the Professorship Production Systems and Processes at Chemnitz University of Technology (vehicle, hardware, on-board software) and Mobile-Etage GmbH (VR client, backend, video and audio transmission).



Figure 1. *The demonstrator during a laboratory test: the remote person appears as an avatar on the forward-facing display, with the 360° camera mounted directly above it.*

The base platform

The vehicle is an in-house development from the earlier FOLLOWme ILS project (Allmacher et al. 2019; Allmacher and Klimant 2019): a differential-drive platform with trailing castors, an aluminium profile frame with sheet-metal cladding and a thermoformed housing, powered by a 24 V LiFePO4 battery. It was specified for intralogistics duty at up to 2.6 m/s and 100 kg payload, and perceives its surroundings through a SICK S300 safety laser scanner with a 270° field and a SICK Visionary-T time-of-flight camera.

Its software is organised in three layers — motor control, including the safety functions, direct control towards a target point, and path planning. Localisation uses Hector-SLAM (Kohlbrecher et al. 2011) embedded in a Kalman filter fed by drive-resolver odometry; obstacle avoidance combines a geometric algorithm with a P-controller for angular velocity and a fuzzy controller for forward speed.

Two driving modes existed: autonomous point-to-point travel, and a follow mode tracking a person from the time-of-flight camera.

Reusing this platform meant mobility, the safety layer and the autonomy stack were already proven, and that we kept the drawings, the source code and the wiring — access down to the level of individual components.

From AGV to TPR: driving and safety

The vehicle was first given a remote-control interface accepting keyboard input or a VR/game controller; the Meta Quest 3 controller was adopted as the permanent input device because it maps direction of travel intuitively. Initially restricted to the local network, remote control was later coupled to the call of the VR application, so that the vehicle is driven directly from the VR client.

Obstacle detection and the safety stop stay active while the robot is teleoperated: the remote person steers, but the vehicle will not collide. The follow mode was kept as well, so an on-site companion can simply walk while the robot keeps up — useful whenever the remote person wants to attend rather than navigate. A bidirectional client–server interface provides synchronous transfer of control commands, video and audio.

Output and input components

Two 27-inch displays in portrait orientation were mounted on the vehicle, one with touch input. The forward-facing display shows the avatar of the remote person, the second carries controls and views for the on-site person. Their in-house mount, verified analytically and experimentally, was initially height-adjustable and later fixed at a lower position. For audio output, a bone-conduction headset serves quiet environments and a Bluetooth/Aux speaker serves group situations. An inverter in the 24 V system supplies the added 230 V components.

For visual capture, a stereo camera (Stereolabs Zed 2) combined with a monoscopic 360° camera (Ricoh Theta V) was compared against a full 360° 3D camera (Insta360 Pro). The latter was excluded early: its latency and bandwidth demands proved incompatible with live remote operation. The two remaining cameras were mounted directly above one another, the 360° camera covering everything outside the stereo field of view. The capture point was placed above the front display so that an on-site person looking at the screen is seen from a near eye-contact perspective; a position beside the display produced a distinctly stronger impression of looking past the remote person. In the final configuration, the stereo camera was dropped in favour of a single consistent 360° stream at 30 fps on a more powerful on-board computer, with audio in mono.

VR client and integration

The VR side is a Unity3D application running on Meta Quest 3 and Pico head-mounted displays. Inside VR, a virtual robot with a seated body is rendered beneath the remote person's avatar: scaled to the real vehicle, it strengthens embodiment and eases steering, with drive commands and their direction visualised on the floor. Hand tracking allows driving without physical controllers.

Because the platform is our own, the interface on the robot could be designed rather than merely accepted: audio device selection, volume, a call button, a camera preview and the strength of the wireless connection, mirrored in VR and extended there by a battery indicator, a self-view and error messages. Video and audio run through a streaming SDK, with an alternative local path over OBS and the NDI protocol for operation independent of internet connectivity. An interactive VR tutorial lets users learn the controls before driving the real robot.

Lessons learned

Camera height dominates the spatial impression. Lowering the 360° camera to roughly the eye height of the on-site person markedly improved the remote person's perception of the scene; a

few centimetres higher already produced the impression of floating above everything. The ideal would be a lens at eye height of the displayed avatar, but camera and display occlude each other — of the two compromises, above the display was clearly better than beside it.

Bandwidth is the binding constraint. Under varying wireless network quality, one consistent stream beats a higher-fidelity multi-stream setup. Dropping the stereo camera removed both the resolution collapse under load and the visible seam between camera streams under changing lighting.

Display height is a genuine trade-off. A lower assembly serves the remote person's perspective, a higher one the ergonomics of the on-site person. Adjustability did not justify its mechanical cost; a fixed lower position was an acceptable compromise.

Existing autonomy is an asset, not overhead. Keeping the AGV's safety stack active during teleoperation and retaining the follow mode cost little and left both available as independent experimental conditions.

Findings of this kind require an open platform. Each observation above came from changing one part of the system and observing the effect — camera height, stream configuration, mount geometry, interface layout. On a closed platform, none of these variables is reachable.

Conclusion

Converting an existing AGV yielded a laboratory-validated telepresence platform whose operating interface, sensor positions, mounting geometry and on-board software are all open to inspection and change. That accessibility is the point: it is what turns a demonstrator into a platform on which design questions about telepresence can be answered.

Acknowledgement

This work was funded by the German Federal Ministry of Research, Technology and Space (BMFTR) under grant no. 16SV9138. The responsibility for the content of this publication lies with the authors.

References

Allmacher, C., Dudczig, M., Knopp, S. & Klimant, P. (2019). Virtual Reality for Virtual Commissioning of Automated Guided Vehicles. *IEEE Conference on Virtual Reality and 3D User Interfaces* (VR), 838–839. https://doi.org/10.1109/VR.2019.8797981

Allmacher, C. & Klimant, P. (2019). Test Course Creation for Automated Guided Vehicles. In: Putz, M. & Schlegel, A. (eds.) 18. *ASIM Fachtagung Simulation in Produktion und Logistik. Auerbach: Wissenschaftliche Scripten*, 305–314.

Franzkowiak, L., Birke, J., Langer, D., Schwarzkopf, M. & Bullinger, A. C. (2024). Connectedness, Closeness oder Togetherness – Konzeption des Begriffs „Verbundenheit" im deutschen und englischen Sprachraum. *70. Arbeitswissenschaftlicher Kongress*, Vol. 42, Paper E.3.1. GfA Press.

Kohlbrecher, S., von Stryk, O., Meyer, J. & Klingauf, U. (2011). A flexible and scalable SLAM system with full 3D motion estimation. *IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*, Kyoto, 155–160. https://doi.org/10.1109/SSRR.2011.6106777

Kristoffersson, A., Coradeschi, S. & Loutfi, A. (2013). A Review of Mobile Robotic Telepresence. *Advances in Human-Computer Interaction* 2013, Article ID 902316. https://doi.org/10.1155/2013/902316

Legler, F., Franzkowiak, L., Hawer, H., Fedelheimer, A., Schwarzkopf, M. & Bullinger, A. C. (2025). Virtually Extended Telepresence: Overcoming obstacles of using telepresence robots in practice. *71. Arbeitswissenschaftlicher Kongress*, Vol. 44, 374–379. GfA Press. https://doi.org/10.61063/FK2025

A chance to act: Bystander behaviour after an embodied experience of harassment

Kátia dos Santos Estevães[1,2], Isaac Calvis[1,3], Abigya Abayneh Melese[1,2], Akrivi Katifori[4], Tony Donegan[1], Maria Roussou[4], Maria V. Sanchez-Vives[1,5], Mel Slater[2], and Justyna Świdrak[1*]

[1] *Instituto de Investigaciones Biomédicas August Pi i Sunyer (IDIBAPS), Barcelona, Spain*

[2] *Event Lab, Faculty of Psychology, Universitat de Barcelona, Barcelona, Spain*

[3] *Universitat Politècnica de Catalunya, Barcelona, Spain*

[4] *Department of Informatics and Telecommunications, NKUA, Athens, Greece*

[5] *ICREA, Spain*

*Corresponding author: Justyna Świdrak, justyna.swidrak@gmail.com

Keywords: Virtual reality; embodiment; sexual harassment; bystander behaviour; social VR

Abstract

Social virtual reality (VR) can intensify harassment through presence and embodiment, but can conversely support perspective-taking and prosocial action. We investigated whether men who experienced a harassing, rather than professional, virtual job interview from a female avatar's perspective subsequently showed reduced gender bias or greater bystander intervention. Forty men were randomized to harassing or neutral interview content within an otherwise identical Golden Rule Embodiment Paradigm (GREP). One week later, participants encountered harassment in a new virtual cafeteria and could intervene. Embodiment and presence were high, and harassment elicited more emotional distress, perspective-taking and reflection on gender-based violence. However, implicit and explicit attitudes did not change and intervention rates did not differ significantly between conditions. The study demonstrates the feasibility of replaying participants' own behaviour from a target's embodied perspective, while highlighting floor effects, behavioural ceiling effects and the tension between experimental control, social plausibility and participant safety.

Introduction

Harassment is increasingly reported in social VR, where embodied interaction can make abusive encounters feel immediate and personally consequential (Blackwell et al., 2019). Virtual embodiment can also alter social cognition by allowing users to experience events while embodied in another body (Maister et al., 2015). The Golden Rule Embodiment Paradigm (GREP) formalizes this possibility: a participant witnesses harm (or participates passively), then re-experiences the event from the target's embodied perspective, then receives an opportunity to help in a later situation (Slater & Banakou, 2021). Prior work suggests that experiencing sexual harassment from an

embodied victim perspective may reduce later conformity (Neyret et al., 2020), but effects on attitudes and bystander behaviour remain uncertain.

We tested whether the addition of harassment to an otherwise identical GREP experience influenced subsequent gender attitudes and bystander action. Personalized look-alike avatars were used to strengthen self-identification (Salagean et al., 2023). The manipulated factor was interview content (harassing versus professional/neutral); observation, embodiment as the female interviewee, replay and the later transfer scenario were held constant. This design therefore isolated the added effect of harassment exposure within GREP (not the effect of embodiment itself).

Methods

Design and participants. The between-group study comprised two VR sessions one week apart. Forty male participants aged 20-52 years (M = 29.9±9.49) were randomly assigned to the harassment or neutral condition (n=20 each). Groups were similar in age, education and prior VR experience. Participants provided informed consent, and the study received approval from the participating universities' ethics committees.

Session 1: job interview. Participants first embodied a personalized male avatar and passively observed a female avatar being interviewed. The experimental condition included gendered, demeaning and harassing conduct; the control condition used professional, neutral dialogue. Participants' movements and voices were recorded in VR. They then embodied the female interviewee and replayed the same scene, seeing and hearing their previous avatar behaviour from her first-person perspective. All other procedures were identical between groups.

Session 2: transfer scenario. One week later, all participants entered a virtual workplace cafeteria in their personalized avatar. The same male avatar harassed the female avatar, while the participant was free to respond. Changing the setting provided a transfer test in a novel context and reduced immediate carryover and demand effects.

Measures. Implicit gender bias was assessed using an adapted lexical decision task. Explicit attitudes were assessed using the Sexual Harassment Attitude Scale and Ambivalent Sexism Inventory (Exposito et al., 1998; Mazer & Percival, 1989). Embodiment (ownership, agency and self-location), presence and affect were measured in VR. Cafeteria recordings were coded as no intervention, verbal intervention or behavioural/physical intervention. Semi-structured interviews were examined using theme-based content analysis (Neale & Nichols, 2001).

Analysis. Descriptive statistics summarized experience measures and coded themes. Pre-post implicit and explicit measures were compared across conditions. The proportion intervening in Session 2 was compared using Fisher's exact test, with a risk ratio and 95% CI.

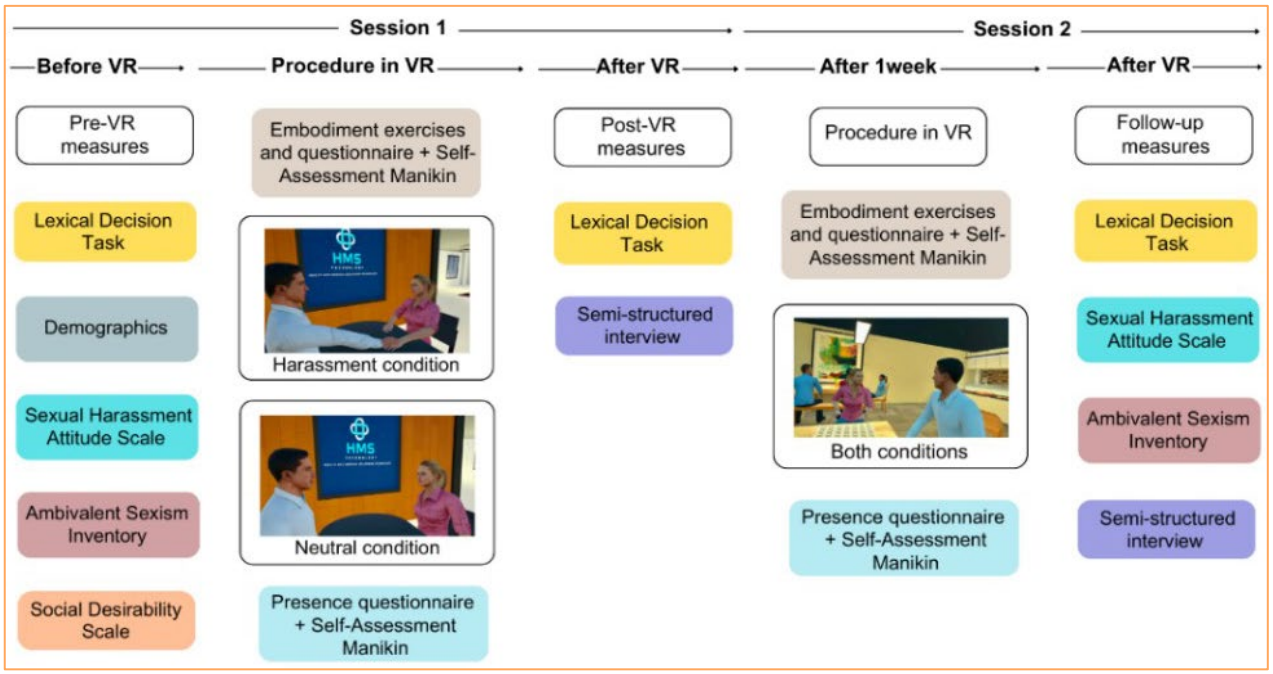


Figure 1. *Experimental procedure across the job-interview and cafeteria sessions.*

Results

Bystander behaviour. In the Session 2 cafeteria, 32/40 participants (80%) intervened: 17/20 in the harassment condition and 15/20 in the control condition. The difference was not statistically significant (p=0.695; risk ratio = 1.13, 95% CI [0.83, 1.55]). Combined verbal and behavioural intervention was the most frequent response in the harassment condition, while verbal-only intervention was most frequent in the control condition.

Attitudes. Baseline lexical-decision, sexual-harassment attitude and ambivalent-sexism scores were low. No significant condition-related pre-post change was found in the implicit or explicit measures, consistent with limited room for improvement and a potential floor effect.

Interview themes. Embodiment and realism were discussed in both groups. Emotional distress (n=17), perspective-taking (n=14), gender-based-violence awareness (n=17) and spontaneous intentions to take action (n=8) were concentrated in the harassment condition.

VR experience. Participants reported high embodiment in both avatars and conditions. Median personalized-avatar embodiment was 5/6 in both sessions (IQR = 2 control; IQR = 1 experimental). Female-avatar embodiment was also high (median = 4, IQR = 2 experimental; median = 5, IQR = 3.5 control), with the strongest ratings for agency and self-location. Place, copresence and plausibility were generally above 4/6. No dizziness was reported.

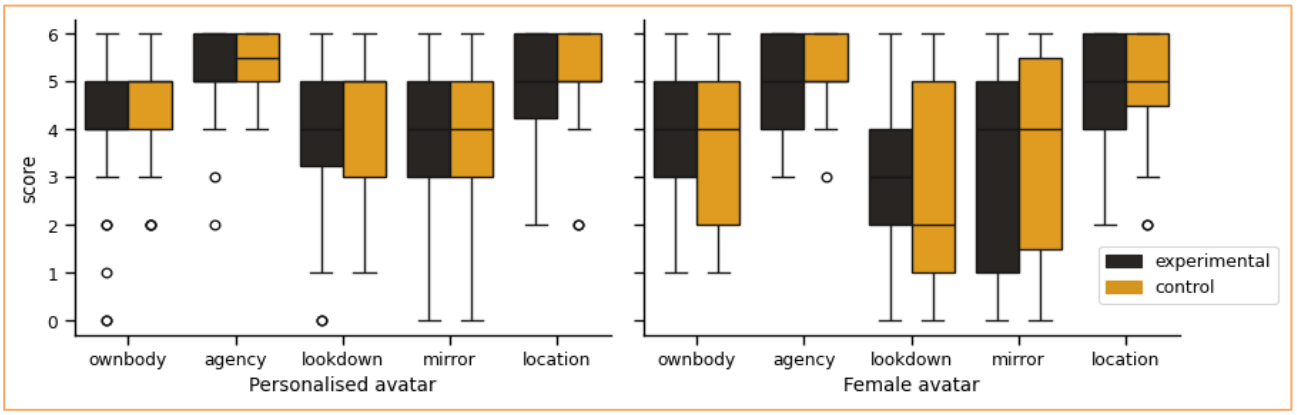


Figure 2. *Embodiment scores for personalized and female avatars by condition.*

Discussion

The harassing interview increased emotional engagement, perspective-taking and reflection on gender-based violence, but did not produce detectable changes in attitudes or significantly increase bystander intervention one week later. Low baseline bias limited the scope for improvement on attitude measures, while the 80% intervention rate limited the binary behavioural measure's ability to distinguish conditions. Measures of latency, escalation, persistence, proximity and gaze may better capture how participants respond.

Realism and emotional intensity can improve the plausibility of sensitive VR scenarios while increasing psychological risk (Slater et al., 2020). Scripted dialogue and replay offer control and participant protection, but may reduce perceived agency when the exchange feels unresponsive. High embodiment alone therefore does not establish behavioural change.

The replay allowed participants to encounter their own recorded movements and voice from the female avatar's perspective, rather than simply observe someone else's experience. Future studies should test more sensitive behavioural outcomes, recruit participants with greater baseline variability and examine whether more interactive scenarios improve social plausibility while retaining experimental control.

Conclusions

The embodied harassment experience elicited strong subjective responses but did not improve gender attitudes or binary bystander intervention relative to a neutral GREP experience. Floor and ceiling effects limited these outcomes. Future evaluations should measure the timing, intensity and quality of action.

References

Blackwell, L., Ellison, N., Elliott-Deflo, N. & Schwartz, R. (2019). Harassment in social virtual reality: Challenges for platform governance. *Proceedings of the ACM on HCI 3, CSCW*, 1-25.

Expósito, F., Moya, M.C. & Glick, P. (1998). Sexismo ambivalente: medicion y correlatos. *International Journal of Social Psychology* 13(2), 159-169. https://doi.org/10.1174/021347498760350641

Maister, L., Slater, M., Sanchez-Vives, M.V. & Tsakiris, M. (2015). Changing bodies changes minds: Owning another body affects social cognition. *Trends in Cognitive Sciences* 19(1), 6-12.

Mazer, D.B. & Percival, E.F. (1989). Ideology or experience? Perceptions, attitudes and experiences of sexual harassment in university students. Sex Roles 20(3), 135-147.

Neale, H. & Nichols, S. (2001). Theme-based content analysis: A flexible method for virtual environment evaluation. *International Journal of Human-Computer Studies* 55(2), 167-189. https://doi.org/10.1006/ijhc.2001.0475

Neyret, S., Navarro, X., Beacco, A., Oliva, R., Bourdin, P., Valenzuela, J., ... & Slater, M. (2020). An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later milgram obedience scenario. *Scientific reports*, *10*(1), 6207.

Salagean, A., Crellin, E., Parsons, M., Cosker, D. & Stanton Fraser, D. (2023). Meeting your virtual twin: Effects of photorealism and personalization on embodiment, self-identification and perception of self-avatars in virtual reality. *Proceedings of CHI 2023*, 1-16. https://doi.org/10.1145/3544548.3581182

Slater, M. & Banakou, D. (2021). The Golden Rule as a paradigm for fostering prosocial behavior with virtual reality. *Current Directions in Psychological Science* 30(6), 503-509.

Slater, M., Gonzalez-Liencres, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., ... & Silver, J. (2020). The ethics of realism in virtual and augmented reality. *Frontiers in Virtual Reality*, *1*, 1. https://doi.org/10.3389/frvir.2020.00001

.

Assessment methodologies for the perceived quality of split-domain light field visualization in single- and simultaneous-participant studies

Peter A. Kara[1*], Tamas K. Tottszer[1], and Tibor Balogh[2]

[1]*Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3, Budapest, 1111, Hungary*

[2]*Holografika, Baross str. 3, Budapest, 1192, Hungary*

Corresponding author: kara@hit.bme.hu

Keywords: Light field, subjective study, quality assessment, user experience

Introduction

Unlike stereoscopic 3D and the various implementations of extended reality, projection-based light field displays offer a 3D visual experience without the need for any viewing device. Such glasses-free 3D displays provide the viewing-angle-appropriate perspective of the visualized content – aiming to replicate the real-life parallax effect. The number of simultaneous observers or users is thus not limited by the availability of viewing apparatus, and no user tracking is needed for operation.

In contrast to multiview displays that repeat the same narrow perspectives over designated viewing positions, light field displays employ the entire field of view (FOV) to create the smooth and continuous parallax effect. The FOV angle is measured from the screen of the display and determines the size of valid viewing area (VVA). The greater the VVA, the more simultaneous observers may be accommodated.

While the general practice is to use the entire VVA for a single 3D model or scene, it is possible to deviate from this common approach (i.e., different observers may simultaneously perceive different contents on the same light field display). An application of such split-domain light field visualization may be the adaptation of conventional split-screen gaming. While split-screen gaming dedicates a portion of the screen to a given player – and trivially constrains competitiveness by providing visual access to the gaming perspectives of all the players – split-domain gaming does not compromise the screen (i.e., each player effectively uses the entire screen for their own gaming perspectives), but only the associated viewing area.

In this paper, we present the assessment methodologies of our ongoing and future empirical research efforts on split-domain light field visualization. The paper discusses the general considerations that are applicable to all of the subjective studies, and highlights the differences between single-participant (i.e., having only one test participant complete the test at one given time) and simultaneous-participant (i.e., having at least two test participants complete the test at the same time) studies.

Prior research efforts

When adjacent viewing domains are created for a light field display via the fusion of content, there is an evident loss of effective FOV due to the interference of neighboring light rays belonging to the different domains, resulting in invalid light field. The affected region between the domains is a separation zone, the characteristics of which were investigated by the work of Javidi et al. (2024). Camera-captured 3D models were fused on a Looking Glass light field display to address the degraded separation zone between the domains. The degradation itself is analogous to the well-known crosstalk effect (Seuntiëns et al., 2005; Kovács et al., 2014), since adjacent perspectives interfere with each other. The authors used a pinhole camera at two capture distances to study the interference.

Split-domain visualization was also enabled by the work of Pan et al. (2020) on resource optimization. The web-camera-based solution processed head position information to provide light field projection only in directions from which the screen could be perceived by observers – as rendering light field in unoccupied regions may be considered a waste of resources. Similarly, a Looking Glass light field display was used in the research effort. Since the solution was designed to accommodate multiple simultaneous observers and different 3D contents, the authors implemented dynamic split-domain light field visualization.

In a recent work of Kara & Balogh (2026), a feasibility study was carried out for split-domain light field visualization. The authors followed the approach of Javidi et al. (2024), but focused on the actual size of the resulting domains. Figure 1 depicts the results from a two- and three-domain setup on the HoloVizio 80WLT light field display. In both research efforts, the size of the separation zone was minimized through fusion without any gradual change.

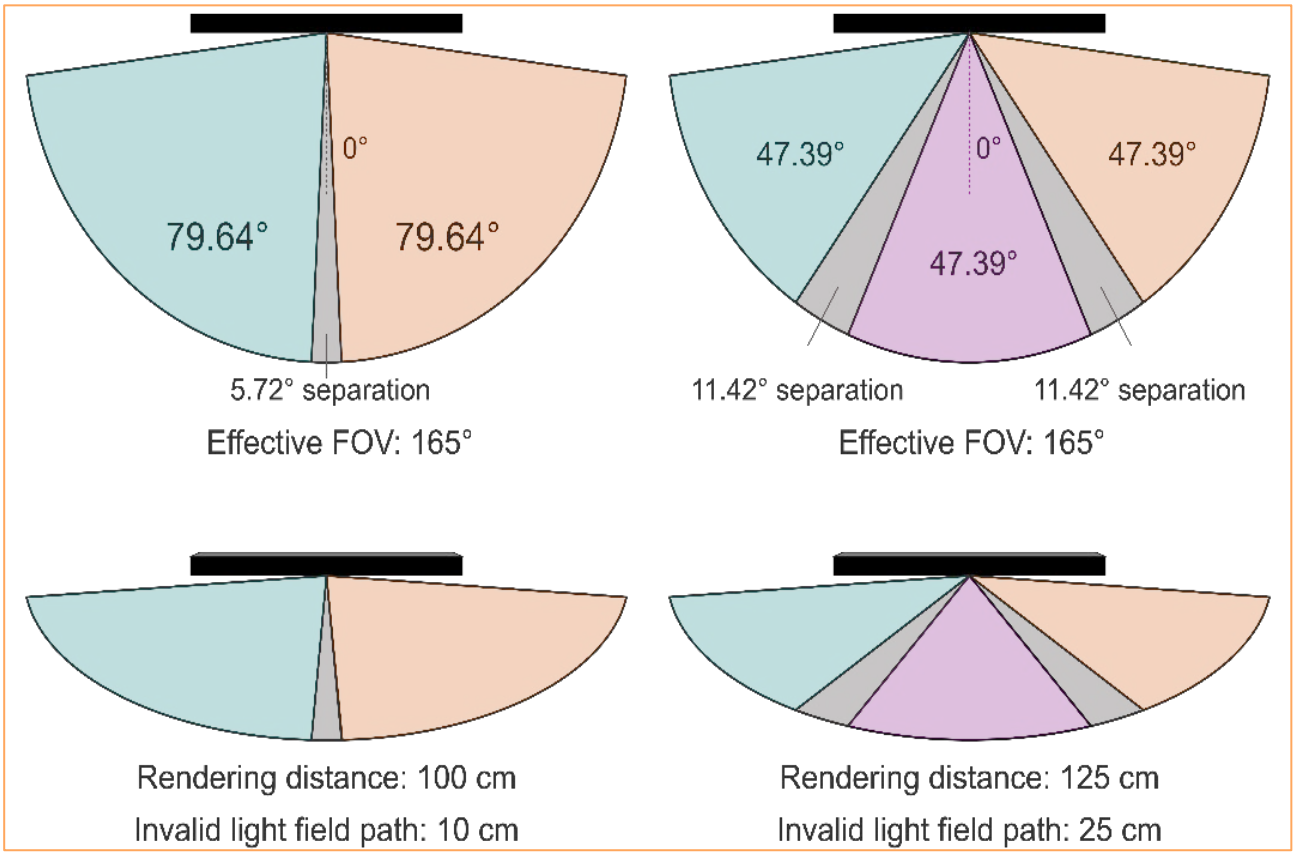


Figure 87. *Characteristics of two- (left) and three-domain (right) split-domain light field visualization.*

Single-participant studies

When only a single test participant is involved at a given time, the vast majority of the relevant ITU recommendations (e.g., ITU-T P.910) and the best practices of the IEEE 3333.1.4-2022 are applicable. However, regardless of observer motion, the default viewing angle should be the center of the domain – not to be confused with the general center view at zero degrees. Figure 2 illustrates this concept and the orientation of content via two-domain visualization. Our current research effort investigates the preference between the two depicted orientations.

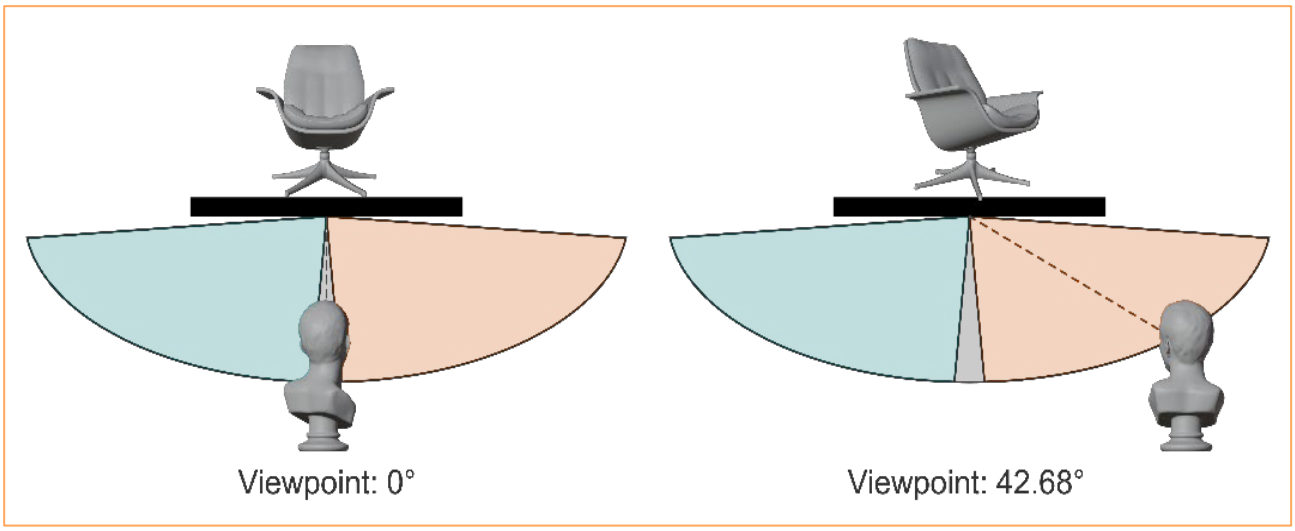


Figure 88. *Two-domain split-domain light field visualization with content orientation facing the center view (left) and the domain-aligned view (right).*

Both single-stimulus and double-stimulus methodologies are viable for single-participant studies. Beyond passive content consumption (e.g., observing a light field video), interactive use cases may also be addressed (e.g., gaming).

Simultaneous-participant studies

Having multiple test participants in a study at a given time poses notable challenges. Generally, participants should not influence each other in terms of assessment. While single-participant studies may employ verbal evaluation, it is to be avoided if at least two participants are present. Another option is to use the regular paper-based approach, yet that could necessitate changes in lighting conditions. Handheld devices may circumvent such an issue; however, these screens should be turned off during the observation of the visualized stimulus.

Regarding interaction, the influence between test participants is mitigated by the fact that the users' perspectives are isolated and thus everyone sees only their own. A solution for the quality assessment of interactive use cases is to use the display and a controller for the assessment task. In such a case, quality assessment is integrated into the interactive software. This approach is also applicable to passive observation, yet implementing such for a light field display – instead of using a dedicated viewer software – may unnecessarily complicate the work.

In contrast to single-participant studies, simultaneous-participant studies enable the investigation of inter-user effects. At the time of writing this paper, no such phenomenon has been addressed by subjective studies on light field visualization. Research efforts on inter-user effects are relevant for both competitive and collaborative use cases, as well as media consumption (e.g., watching different videos at the same time on the same display). While influences between participants should be avoided for most studies, those investigating inter-user effects are designed to accommodate and measure such phenomena.

While both single-stimulus and double-stimulus methodologies are viable for simultaneous-participant studies, certain test variables may be constrained, such as observer motion. During free-motion tests, the motion domain should be limited to avoid observer collision, since the separation zone may be rather narrow – as illustrated by Figure 1. After all, during movement, the test participants look at and focus on the screen of the light field display to observe the motion parallax effect.

Conclusion

In this paper, we addressed the assessment methodologies for single- and simultaneous-participant studies for split-domain light field visualization. Our ongoing single-participant subjective study on the HoloVizio 80WLT display investigates the preference between two approaches of content orientation via comparison category rating. As only a single individual participates in the test at a given time, verbal evaluation is used, although such a method is not recommended for simultaneous-participant studies. In future work, we aim to employ absolute category rating and degradation category rating as well for subjective data collection. The latter shall be of particular use for the measurement of just-noticeable differences. The vast majority of the research efforts shall exclude observer motion altogether, unless it is a test variable. The default viewing angle in the experiments is the center of the domain. For simultaneous-participant studies, our goal is to study both two- and three-domain configurations, with particular emphasis on inter-user effects.

Acknowledgments

The work in this paper was funded from the European Union's Horizon Europe 2021-2027 research and innovation program, under the HORIZON-CL4-2023-HUMAN-01-CNECT - HEAT Project, No: 101135637.

References

Kara, P.A. & Balogh, T. (2026). Valid-viewing-area-focused feasibility study on split-domain light field visualization. In *3rd Workshop on Interaction and Storytelling with(in) Immersive Media* (ISIM 2026), Athlone, Ireland, 9 June 2026, pp 1–7.

Javidi, K., Martini, M.G. & Kara, P.A. (2024). Investigation of the separation zone for split-domain light field visualization. In 2024 *28th International Conference on Information Technology (IT), Žabljak, Montenegro*, 21–24 February 2024, pp. 1–4. https://doi.org/10.1109/IT61232.2024.10475755

Kovács, P.T., Boev, A., Bregović, R. & Gotchev, A. (2014). Quality measurements of 3D light-field displays. In *Proceedings of the Eighth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2014), Chandler, Arizona, USA*, 30–31 January 2014, pp. 1–6.

Pan, X., Zheng, M., Yang, J. & Campbell, A.G. (2020). An adaptive low-cost approach to display different scenes to multi-users for the light field display. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology (VRST '20)*, Virtual Event, 1–4 November 2020, pp. 1–3. https://doi.org/10.1145/3385956.3422127

Seuntiëns, P.J.H., Meesters, L.M.J. & IJsselsteijn, W.A. (2005). Perceptual attributes of crosstalk in 3D images. *Displays*, 26(4–5), pp. 177–183. https://doi.org/10.1016/j.displa.2005.06.005

Direct comparison of 3D content on glasses-free light field displays and in extended reality: experimental design alternatives and preliminary observations

Tibor Balogh[1], Tamas K. Tottszer[2], and Peter A. Kara[2*]

[1]*Holografika, Baross str. 3, Budapest, 1192, Hungary*

[2]*Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Műegyetem rkp. 3, Budapest, 1111, Hungary*

Corresponding author: kara@hit.bme.hu

Keywords: Light field, extended reality, experimental methodology, user experience

Introduction

One of the greatest advantages of light field visualization is that it does not require any personal viewing apparatus. While solutions for near-eye light field visualization do exist (Han et al., 2022; Wang & Hua, 2021), projection-based systems such as the HoloVizio displays do not rely on the usage of special glasses or helmets to provide a full 3D visual experience (Balogh, 2006; Balogh et al., 2007). In contrast, virtual reality (VR), augmented reality (AR), and mixed reality (MR) are all fundamentally built on viewing devices.

There are numerous approaches to characterize the performance of implemented systems and solutions of 3D visualization. In the world of research, empirical studies involving human test participants – also known as subjective tests – provide valuable information regarding perception, preference, usability, and many other vital descriptors of user-related performance. Such studies also aid the creation and evaluation of novel objective quality metrics.

In most utilization contexts, the comparison of technologies and solutions is necessary to answer questions regarding relative performance (e.g., which device enables the highest level of user interaction accuracy and the shortest task completion time). Devices and systems may be compared in an indirect manner, meaning that they are assessed by using similar methodologies all whilst independent of one another – investigated in separate subjective studies. Once the tests of both studies are concluded, the collected data can be compared.

Conversely, a direct comparison between two vastly different technologies is quite far from being straightforward. In fact, there may be challenges that not only complicate the design and execution of such subjective studies, but may even undermine the validity of the output of the research effort. Due to the differences between their individual characteristics, the direct comparison between the performance of glasses-free and near-eye 3D technologies is inevitably plagued by these barriers. Additionally, the assessment methodologies of direct comparison are also necessary to collect information on user preference.

In this paper, the experimental design alternatives and considerations for the direct comparison of light field and extended reality (XR) visualization are addressed, and the preliminary observations of a pilot study are discussed. The paper separately highlights six areas that are the most crucial when comparing light field and XR within the same subjective study.

Apples, oranges, and everything in between

First and foremost, a scientifically valid comparison of two visualization technologies necessitates that the same contents are compared, using identical methodologies. Such methodologies extend to the test environment (e.g., the characteristics of the laboratory where the tests are conducted), variables (e.g., the instances of quality degradation that are applied to the source contents), protocol (e.g., the rating scale and temporal structure of subjective data collection), and visualized stimuli. In the case of a direct comparison, the test environment is shared, and the test variables and protocol do not typically pose challenges either. However, the content itself may be a significant obstacle, as most visual stimuli of light field imaging are not necessarily compatible with XR devices. For example, a frequent method of producing light field is based on image sequences that capture the same scene from different perspectives. One way to circumvent this issue and find common ground is to use point clouds. In our current pilot study, we use the display-specific converter of the HoloVizio system to match the input point clouds to the visualization properties of the display. Figure 1 exhibits such a point cloud visualized on the HoloVizio 640RC light field display – observed by multiple individuals simultaneously from different viewing angles.

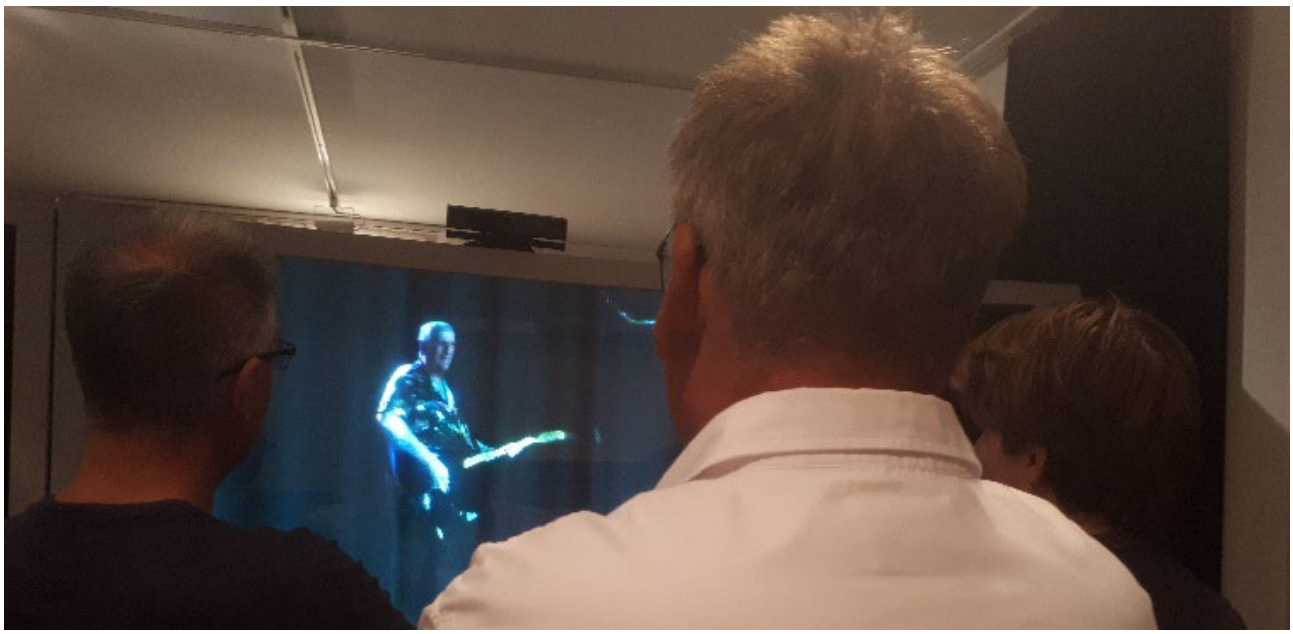


Figure 89. *Sparse point cloud video visualized on the HoloVizio 640RC.*

The ocean or a multitude of drops

Another important consideration of subjective studies is the descriptor of perceived quality. One approach to measure this is to rate the aggregated overall quality, with the task of incorporating everything related to user experience. The other – equally important – approach is to separately rate every relevant aspect of the experience. Rating the overall quality simplifies the protocol and shortens the test duration, whilst on the other hand, rating the individual aspects separately may highlight important tradeoffs between the investigated technologies. In the context of our work, the latter approach was applied not only to perceived quality (e.g., separately addressing perceived depth), but to other experience-related factors that stem from the differences of the two devices (e.g., user comfort) as well.

To infinity and beyond, if possible

Similarly to conventional 2D visualization, XR technologies support perception towards virtually infinite distances. However, as light field imaging describes a portion of 3D space, the visualized content must also be spatially constrained. Additionally, note that light field visualization is, by definition, the best in terms of quality within the plane of the screen, and loses angular density proportionally with content divergence along the two directions of the depth axis. In our work, this issue is solved by using point clouds that are bound within a well-defined spatial domain.

Conflict without resolution, and resolution without conflict

Potentially the greatest drawback of near-eye visualization technologies is the vergence–accommodation conflict (VAC) (Kim et al., 2012). It is due to the mismatch of eye rotation and focus, as the eyes need to focus on the same close surface regardless of the perceived changes in distance. However, due to having no necessity for such viewing apparatus, light field displays do not suffer the same issue (Huang & Hua, 2017). In our pilot, preliminary reports already indicate the potential perceptual fatigue of changing between the two technologies.

A change is as good as a rest, sometimes

In the case of the direct comparison of light field and XR visualization, a simple yet greatly impactful decision regarding the experimental design of a subjective study is whether to employ a sequential (i.e., all tests are on one display, followed by the other display) or alternating (i.e., changing between displays for each comparison) methodology. Even though the latter may introduce additional issues – as stated above – it is still necessary in order to carry out a double-stimulus comparison category rating method, which we too employed in our pilot study. Figure 2 depicts instances of our pilot study using the alternating methodology to assess point cloud videos reconstructed from sparse camera arrangements.

Time flies when you're having fun

The duration of the tests may not only influence the overall experiment itself, but also impact the investigated technologies differently. For XR, the previously mentioned VAC and other issues may pose limitations that are not relevant for light field visualization. However, alternating between the displays and rating multiple aspects do extend the total duration of the experiment. Therefore, both the fatigue and discomfort caused by XR should be considered, as well as the additional fatigue resulting from the alternating methodology. In our work, the total duration was shortened by reducing the number of source point clouds, and by narrowing the scope of the research effort through limited test variables and conditions.



***Figure 90**. Sparse point cloud video visualized on the HoloVizio 80WLT (left), on the Meta Quest 3 (middle), and as seen through the XR helmet (right).*

Conclusion

In this paper, we discussed the considerations regarding our ongoing pilot study on the direct comparison of light field and XR visualization. In order to avoid comparison of apples and oranges, we converted point cloud videos to the characteristics of the light field display used in the research effort. As the contents are spatially constrained, the potential problem with virtual infinity is circumvented. Quality assessment through comparison category rating is segmented to cover the individual aspects of quality and user experience. Due to the issues of VAC and the alternating between the devices, the source contents and test variables were limited to keep the subjective study as brief and focused as possible.

Acknowledgements

The work in this paper was funded from the European Union's Horizon Europe 2021-2027 research and innovation program, under the HORIZON-CL4-2023-HUMAN-01-CNECT - HEAT Project, No: 101135637.

References

Balogh, T. (2006). The HoloVizio system. *Stereoscopic Displays and Virtual Reality Systems XIII*, Proceedings of SPIE, Vol. 6055, pp. 279–290. https://doi.org/10.1117/12.650907

Balogh, T., Kovács, P.T. & Barsi, A. (2007). Holovizio 3D display system. In *2007 3DTV Conference, Kos, Greece*, 7–9 May 2007, pp. 1–4. https://doi.org/10.1109/3DTV.2007.4379386

Han, W., Han, J., Ju, Y.-G., Jang, J. & Park, J.-H. (2022). Super multi-view near-eye display with a lightguide combiner. *Optics Express*, 30(26), pp. 46383–46403. https://doi.org/10.1364/OE.477517

Huang, H. & Hua, H. (2017). Systematic characterization and optimization of 3D light field displays. *Optics Express,* 25(16), pp. 18508–18525. https://doi.org/10.1364/OE.25.018508

Kim, J., Kane, D. & Banks, M.S. (2012). Visual discomfort and the temporal properties of the vergence–accommodation conflict. In *Stereoscopic Displays and Applications XXIII, Proceedings of SPIE*, Vol. 8288, pp. 352–360. https://doi.org/10.1117/12.912223

Wang, X. & Hua, H. (2021). Depth-enhanced head-mounted light field displays based on integral imaging. *Optics Letters*, 46(5), pp. 985–988. https://doi.org/10.1364/OL.413676

An AI Data Assistant for Collaborative XR Inspection of Building Digital Twins

Panagiotis Iatrou[1], Moatasim Mahmoud[1], Rahool Dembani[1], and Stamatia Rizou[1]

[1]*Singular Logic, Greece*

Corresponding author: PIatrou@singularlogic.eu

Keywords: XR, Digital Twin, Data Assistant, AI Assistant



***Figure 91**. Overview of the AI data assistant (WARDY) in the collaborative XR digital twin environment.*

This work develops a collaborative extended reality (XR) digital twin application, supported by an artificial intelligence (AI) data assistant. To help overcome the technical complexity of XR-based digital twin systems, our conversational AI assistant supports natural language queries to real-time and historical sensor data, statistical analysis, interactive data visualization, and spatial navigation of a building digital twin.

Digital twin systems can be quite complex, especially when having many sensors and intertwined subsystems, often leading to high cognitive burden where monitoring and analysis can become overwhelming. While XR technologies can support more effective monitoring and inspection in digital twin environments (Tu et al., 2023; Grego et al., 2024), such systems remain highly complex to navigate and process (Sthapit & Olbina, 2026). Nowadays, LLM agents can process natural language queries, independently identify and use the right tools and generate coherent responses, without needing domain-specific technical expertise (Chen et al., 2025; Chirumamilla2025). As

advancements in AI and large language models (LLMs) continue to grow, their role in XR applications is getting more attention (Rahimi et al., 2025). Researchers have investigated and developed AI assistants in a variety of XR domains including education (Alipour et al., 2025; Mohameda et al., 2025), medicine (Wu et al., 2024; Chheang et al., 2024), career guidance (Tandaroudas et al., 2026), and creative arts (Chen et al., 2025). Yet, their role in XR-based digital twins and data analysis remain underexplored.
Motivated by recent progress in LLM-based data understanding and analysis, we decided to investigate how AI assistants can lower the technical complexity and cognitive load in collaborative XR-based digital twin systems. To achieve this, we develop and present WARDY, a conversational digital twin AI assistant for real-time data analysis and spatial navigation. WARDY is developed to assist XR users navigate digital twin environments and analyze sensor data more easily.

Implementation

Sensor Streams and Data Management: In order to supply the digital twin and the AI assistant with live and historical sensor measurements, the application uses a dedicated data management pipeline. The deployed sensors consist of Milesight AM103/VS133/VS350 devices that measure temperature, humidity, CO_2, and people count data. They publish to an MQTT broker to which a NiFi flow subscribes and maps the stream to NGSI-LD entities stored in Orion-LD. Finally, the agentic tools query Orion-LD for live and historical data.

XR System: We implement the XR interactions and collaboration capabilities using Unity's XR Interaction Toolkit and NGOs. The system supports virtual (VR) and mixed reality (MR) modes and gives the option for hand-based and controller-based interactions. The user is presented with a 3D building model with live sensor data readings taken from the data management pipeline, an AI interaction UI with voice and text input options, and a data visualisation UI for plotting historical data.

AI Assistant & Agentic Tool Calling: Based on Ollama and using the Qwen3 model, WARDY is implemented as a conversational AI assistant embedded within a Unity-based building digital twin application. It adopts an agentic tool-calling architecture in which the LLM is provided with a suite of callable functions and autonomously decides which tools to invoke based on the user's query. The system supports multi-tool execution within a single conversational turn, allowing the agent to issue several parallel tool calls and aggregate their results before generating a response. This capability is essential for handling complex queries that require synthesizing data from multiple sensors and analysis functions simultaneously. The tool suite is organized into four functional categories. **Real-time sensing tools** provide immediate access to current sensor readings, with support for individual queries, multi-sensor comparison, room-level aggregation, and threshold-based alerting. **Historical analysis tools** enable temporal reasoning over stored sensor data, supporting average, minimum, and maximum queries over both flexible durations and explicit time ranges. A **visualization tool** renders time-series data as a labelled line chart directly within the DT interface. Finally, **spatial navigation tools** allow the agent to control the DT's visual representation, including toggling floor visibility and switching between viewing modes.

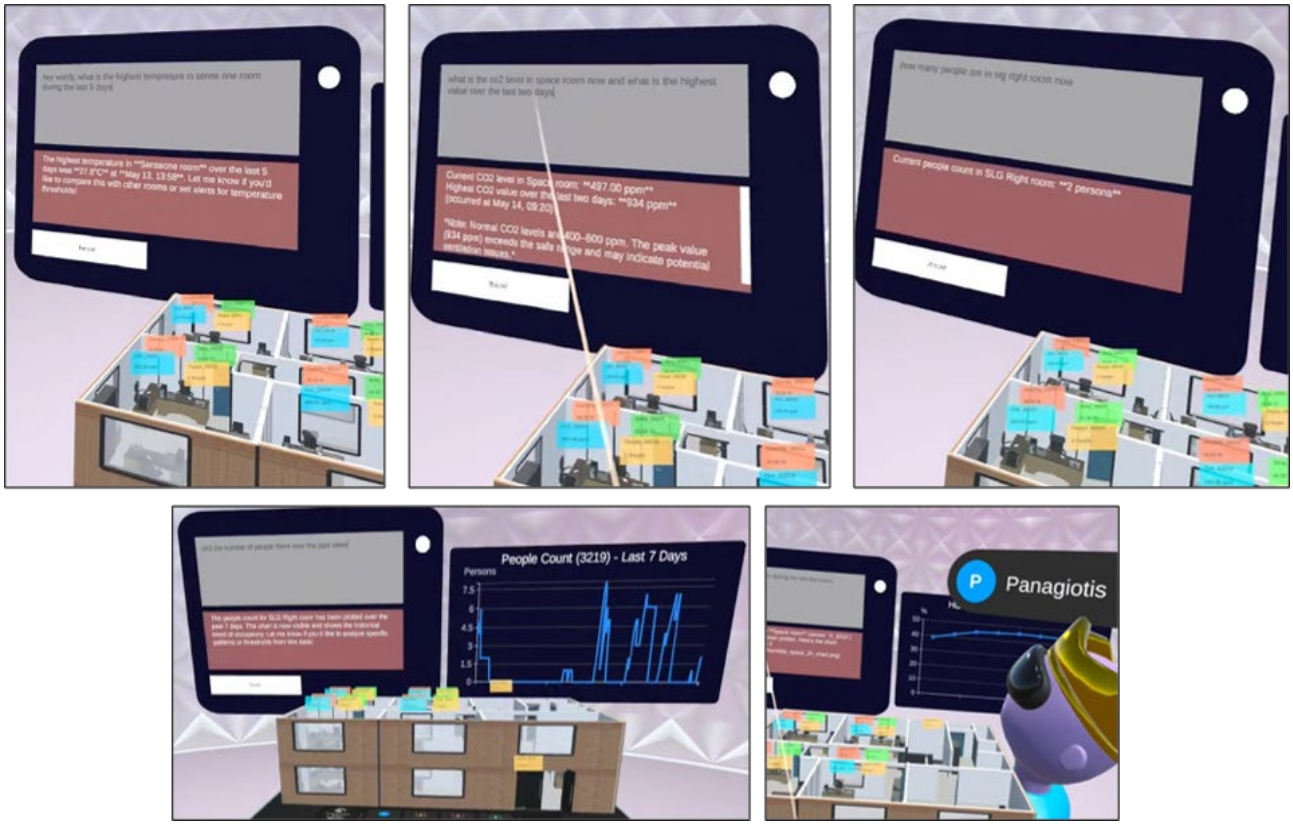


Figure 2. *Examples of WARDY's assistance and execution of tools in response to different prompts by XR users.*

System Demonstration

Figure 2 presents the implemented collaborative XR-based digital twin system, along with the AI assistant's responses to XR user prompts. In the first example, the user asks WARDY about the highest recorded temperature in a specific room over the last five days. WARDY correctly calls the relevant function to retrieve the maximum temperature in the specified room and time period and formulates the response accordingly. The second example demonstrates multi-tool execution where WARDY invokes two functions to report the current and highest recorded CO_2 levels in another room. The third example shows the use of a real-time sensing tool, to report the number of people at the moment. The last two examples illustrate the use of the visualization tool to plot historical data as requested by the user.

In future work, we intend to expand the tools library that WARDY uses for more effective assistance. We aim to provide tools with deeper data processing capabilities, allowing AI assistants not just to execute direct commands, but to truly "understand" trends in live and historical data to become as collaborators themselves, giving timely interventions and offering relevant data-driven suggestions.

Acknowledgment

This work was supported by the CLIMRES "Leadership for Climate Resilient Buildings" Horizon Innovation Actions project supported by the European Commission under grant agreement No 101147777.

References

Alipour, S., Coward, L., Burris, J., Karki, B., & Liao, T. (2025, October). Evaluating an AI Teaching Assistant in VR Physics Education: A Case Study of User Experience and Effectiveness. In *2025 IEEE International Symposium on Emerging Metaverse (ISEMV)* (pp. 62-68). IEEE. https://doi.org/10.1109/ISEMV67326.2025.00021

Chheang, V., Sharmin, S., Márquez-Hernández, R., Patel, M., Rajasekaran, D., Caulfield, G., ... & Barmaki, R. L. (2024, January). Towards anatomy education with generative AI-based virtual assistants in immersive virtual reality environments. In *2024 IEEE international conference on artificial intelligence and eXtended and virtual reality (AIxVR)* (pp. 21-30). IEEE. https://doi.org/10.1109/AIxVR59861.2024.00011

Chen, K., Wang, P., Yu, Y., Zhan, X., & Wang, H. (2025). Large language model-based data science agent: A survey. *arXiv preprint arXiv:2508.02744*. https://doi.org/10.48550/arXiv.2508.02744

Chen, Q., Mishra, R., Purnamasari, L. S., El-Zanfaly, D., & Kitani, K. (2025, April). Origami sensei: a mixed reality AI-assistant. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1-18). https://doi.org/10.1145/3706598.371409

Chirumamilla, K. R. (2025). Multi-Agent AI Assistant for Real-Time Business Insights. *Journal of Recent Trends in Computer Science and Engineering (JRTCSE)*, *13*(3), 17-36. https://doi.org/10.70589/JRTCSE.2025.13.3.3

Grego, G., Nenna, F., & Gamberini, L. (2024, September). XR-based digital twin for industry 5.0: a usability and user experience evaluation. In *International Conference on Extended Reality* (pp. 387-403). Cham: Springer Nature Switzerland.

Mohamed, A., Faisal, R., Al-Gindy, A., & Shaalan, K. (2025). Artificial intelligence and immersive technologies: Virtual assistants in AR/VR for special needs learners. *Computers*, *14*(8), 306. https://doi.org/10.3390/computers14080306

Rahimi, F., Sadeghi-Niaraki, A., & Choi, S. M. (2025). Generative AI meets virtual reality: A comprehensive survey on applications, challenges, and future direction. *IEEE Access*, *13*, 94893-94909. https://doi.org/10.1109/ACCESS.2025.3574779.

Sthapit, I., & Olbina, S. (2026). Digital Twin (DT) and Extended Reality (XR) in the Construction Industry: A Systematic Literature Review. *Buildings*, *16*(3), 517. https://doi.org/10.3390/buildings16030517

Tantaroudas, N. D., McCracken, A. J., Karachalios, I., Papatheou, E., & Pastrikakis, V. (2026). XR-CareerAssist: An Immersive AI-Powered Platform for Personalised Career Guidance Using Extended Reality and Multimodal Artificial Intelligence. *Open Research Europe*, *6*, 140.

Tu, X., Autiosalo, J., Ala-Laurinaho, R., Yang, C., Salminen, P., & Tammi, K. (2023). TwinXR: Method for using digital twin descriptions in industrial eXtended reality applications. *Frontiers in Virtual Reality*, *4*, 1019080. https://doi.org/10.3389/frvir.2023.1019080

An Interactive Virtual-Reality Platform for Cultural Heritage Exploration, Risk Assessment and Emergency Preparedness

Amir Azizi, Panayiotis Charalambous, Marina Tompouri, and Yiorgos Chrysanthou

CYENS Centre of Excellence, Nicosia, Cyprus

Corresponding author: a.azizi@cyens.org.cy

Keywords: Cultural Heritage; Virtual Reality; Risk Assessment; Immersive Visualization

Abstract

Cultural heritage sites face multiple natural and human-induced hazards, yet risk-assessment outputs are often communicated through static maps, tables, and reports with limited spatial context. This paper presents HeritageGuard, an integrated framework combining a reusable virtual-reality environment with a real-time cultural-heritage risk-assessment platform. The system provides two complementary experiences: a visitor mode for immersive exploration of heritage sites, public information, images, and audio narratives; and a heritage-site-manager mode for screening-level risk assessment, scenario exploration, and emergency-preparedness training.In manager mode, HeritageGuard evaluates hazard, exposure, vulnerability, and capacity indicators and returns structured assessment results to the interactive environment. These results are communicated through spatial risk overlays, ranked risk areas, compact score summaries, and three-dimensional site visualisation. Users can explore flood and earthquake scenarios in relation to the digital heritage model, compare evacuation conditions, and undertake guided or scored emergency-evacuation rehearsals. The rehearsal component supports route and assistance decisions and provides a post-exercise debrief including time, exposure, decision outcomes, safety score, and recommendations.The framework demonstrates how real-time risk computation, immersive visualisation, and scenario-based emergency preparedness can be combined to support early-stage risk communication, comparative analysis, and training. It is intended to complement, rather than replace, calibrated hydrodynamic, structural, fire-safety, and probabilistic risk models.

Introduction

Cultural heritage is increasingly recognised within disaster-risk-reduction agendas because heritage assets are vulnerable to hazardous events while also contributing to identity, continuity, and community resilience. The Sendai Framework for Disaster Risk Reduction 2015–2030 explicitly identifies cultural heritage protection as part of resilience building (UNDRR 2015). Heritage risk can be understood through the interrelated dimensions of hazard, exposure, vulnerability, and capacity (UNDRR 2017), encompassing physical assets as well as visitors, access conditions, emergency procedures, and management capacity.

Indicator-based and multi-criteria approaches have been proposed for cultural-heritage risk assessment. Ravankhah, Schmidt & Will (2021) developed an indicator framework for World Heritage sites in seismic zones, while Datola et al. (2024) introduced a multi-hazard risk-analysis

interface. However, the resulting information is commonly communicated through maps, dashboards, tables, and reports. Although valuable for analysis, these formats provide limited support for understanding how risk, hazards, and emergency decisions relate to the spatial structure of a heritage site.Immersive technologies offer an opportunity to address this limitation. Guo et al. (2024) reported improved situation awareness through X-reality for cultural-heritage risk management, while Li et al. (2024) proposed a digital-twin approach for natural-disaster monitoring. Building on this direction, this paper presents HeritageGuard, an interactive virtual-reality framework connected to a real-time cultural-heritage risk-assessment platform. It supports immersive historical exploration for visitors and screening-level risk visualisation, flood and earthquake scenario exploration, and emergency-evacuation rehearsal for heritage managers. The framework aims to support early-stage screening, spatial risk communication, comparative scenario analysis, and preparedness training; it is not intended to replace specialised engineering, hydrodynamic, structural, or probabilistic risk models.

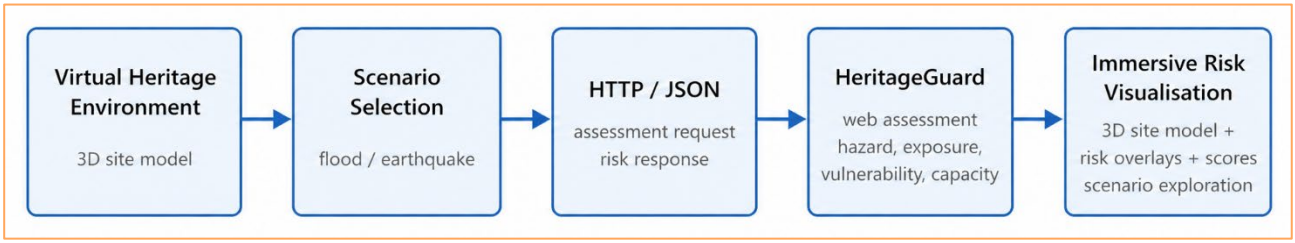


***Figure 1**. Architecture linking the VR heritage environment with HERITAGEGUARD through HTTP/JSON.*

System Architecture and Interaction

The framework combines an interactive VR environment with the HeritageGuard risk-assessment platform. The application offers two complementary user experiences. Visitor mode presents a selected heritage site through public information, imagery, and audio narration. Heritage Site Manager mode provides access to live screening-level assessment, hazard-scenario exploration, and emergency-preparedness activities.In the management workflow, the application sends the selected site and assessment context to HeritageGuard. The platform evaluates indicators related to hazard, exposure, vulnerability, preparedness, and resilience capacity. Structured results,including risk scores, qualitative ratings, data completeness, and recommended actions,are returned to the interactive environment and translated into colour-coded overlays, ranked risk locations, and compact score cards associated with the three-dimensional site model.Users can subsequently explore flood and earthquake scenarios in the same spatial environment. These views highlight risk patterns and locations requiring attention while allowing the site to be inspected from different viewpoints. The emergency-evacuation component supports guided training and scored practice: users assess route options, make assistance decisions for vulnerable groups, and receive a post-exercise debrief reporting time, exposure, decision outcomes, safety score, and recommendations. In the desktop version, the same workflow is represented through scenario configuration, risk-aware routing, crowd simulation, and a comparable debrief. This design extends conventional heritage-risk interfaces by embedding assessment outputs and preparedness decisions within an embodied spatial context (Datola et al. 2024; Guo et al. 2024).

a)	b)	c)

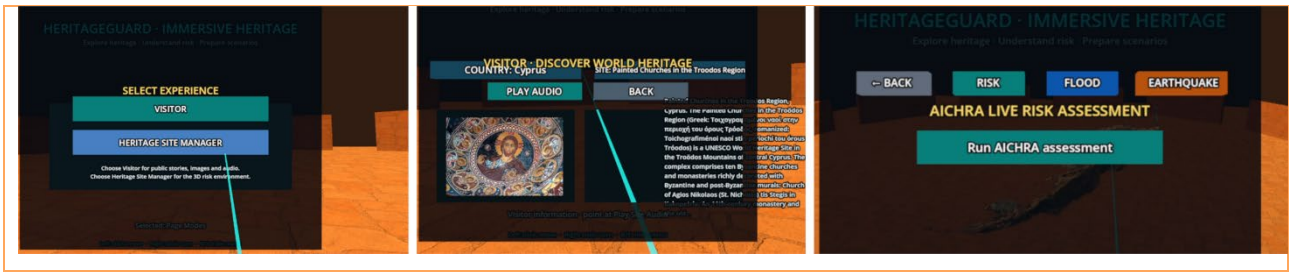


Figure 2. *VR interfaces: (a) mode selection; (b) Visitor Mode; (c) Heritage Manager Mode.*

Prototype Demonstration and Preliminary Findings

The prototype was demonstrated using the archaeological site of Choirokoitia, Cyprus, represented through a three-dimensional heritage model. In Heritage Site Manager mode, users request a live HeritageGuard assessment and inspect the returned screening-level indicators, qualitative ratings, data completeness, recommended actions, and scenario outputs within the interactive environment. The implementation demonstrates the complete request–response workflow without modifying the original assessment platform and supports risk assessment, flood, earthquake, and emergency-evacuation views.Risk values, hazardous areas, and ranked locations are presented in direct spatial relation to the heritage model rather than as isolated numerical outputs. Flood and earthquake overlays provide scenario-specific visualisation, while the emergency component supports guided training and scored practice. In VR, users select route and assistance decisions using controller-based interaction and receive a post-exercise debrief. The desktop version provides a complementary risk-aware evacuation simulation, enabling scenario configuration, routing, crowd representation, and comparable outcome reporting.

Prototype testing identified practical interface requirements, including alignment of risk layers with the three-dimensional model, readable VR labels, staged presentation of route decisions, and clear controller-based selection. These observations informed the use of simplified step-by-step VR interaction, compact score cards, colour-coded risk categories, and scenario-specific overlays. The findings are preliminary and demonstrate technical feasibility and interaction design rather than measured user-performance improvement; structured user evaluation is planned as future work. The framework can be adapted to other heritage sites by replacing the three-dimensional model and associated site metadata.

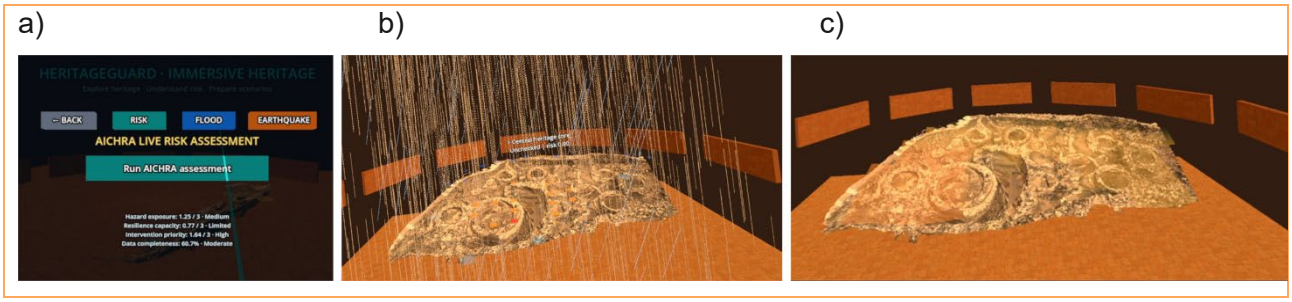


Figure 3. *Heritage Manager workflow: (a) screening-level assessment; (b) flood-risk visualisation; (c) 3D site model.*

Discussion and Conclusion

This paper demonstrates the feasibility of connecting real-time cultural-heritage risk assessment with an immersive virtual environment. The prototype combines visitor-oriented heritage exploration with a Heritage Site Manager workflow for screening-level assessment, spatial hazard visualisation, flood and earthquake scenario exploration, and emergency-evacuation rehearsal. Its main contribution is the integration of live risk computation, three-dimensional contextualisation, scenario-

based preparedness interaction, and post-exercise feedback within a reusable framework.The current implementation is a prototype and does not replace calibrated hydrodynamic, structural, fire-safety, or probabilistic risk models. Future work will extend the framework to additional heritage sites, improve scenario calibration and visual fidelity, and conduct structured user studies to evaluate usability, spatial understanding, decision-making, and preparedness-related outcomes.

References

Datola, G., Assumma, V., Appiotti, F., Bottero, M., Rinaldi, E., Campostrini, P. & Lombardi, P. (2024). The ResCult project: Implementation of the risk analysis interface for the Church of Santa Maria dei Miracoli and the Covent of San Nicola (Italy). *Journal of Cultural Heritage* 67, 164–182. https://doi.org/10.1016/j.culher.2024.02.013

Guo, Y., Liu, L., Huang, W., Shen, M., Yi, X., Zhang, J. & Lu, S. (2024). Extending X-reality technologies to digital twin in cultural heritage risk management: A comparative evaluation from the perspective of situation awareness. *Heritage Science* 12(1), 245.

Li, B., Dong, Y., Qiao, Y., Hou, M. & Wen, C. (2024). Natural disaster risk monitoring for immovable cultural relics based on digital twin. *Journal of Geodesy and Geoinformation Science* 7(1), 90–104. https://doi.org/10.11947/j.JGGS.2024.0107.

Ravankhah, M., Schmidt, M. & Will, T. (2021). An indicator-based risk assessment framework for World Heritage sites in seismic zones: The case of “Bam and its Cultural Landscape” in Iran. *International Journal of Disaster Risk Reduction* 63, 102405. https://doi.org/10.1016/j.ijdrr.2021.102405

UNDRR. (2015). Sendai Framework for Disaster Risk Reduction 2015–2030. *United Nations Office for Disaster Risk Reduction*, Geneva.

UNDRR. (2017). Terminology on Disaster Risk Reduction. *United Nations Office for Disaster Risk Reduction*, Geneva.

Impact of Virtual Reality Immersion on Upper-Limb Object Manipulation

Gabriela Calvo[1], Gloria Campos[1], Adrián Borrego[1], Luis Millán[2], Álvaro Page[3] and Roberto Llorens[1]

[1] *Neurorehabilitation and Brain Research Group, Human-Centered Technology Research Institute, Universitat Politècnica de València, Valencia, Spain*

[2] *Faculty of Physical Activity and Sports Sciences, Universitat de València, Valencia, Spain*

[3] *Institute of Mechanical and Biomechanical Engineering, Universitat Politècnica de València, Valencia, Spain*

Corresponding author: gcalmal@htech.upv.es

Keywords: Upper limb, virtual reality, motor performance, Milgram's continuum.

Introduction

Upper limb deficits are common following central nervous system or musculoskeletal injuries and can impair strength, coordination, spasticity and motor control, affecting patients' daily activities and quality of life (Sathian et al. 2011). Accurate assessment of upper limb function is essential to monitor recovery and schedule rehabilitation programs. However, conventional clinical tools, although quick and easy to administer, lack precision and can be biased by the evaluator's interpretation. Motion capture systems such as videophotogrammetry provide more objective kinematic information, but their clinical implementation can be limited by high cost, subject preparation and laboratory space constraints.

Current virtual reality (VR) display devices are a promising alternative in the clinic, as they function as a hand-tracking motion capture system, without the need for gloves or sensors. Before VR can be reliably used for clinical assessment, it is essential to determine whether virtual environments alter movement execution. Previous studies have shown that movements performed in VR can be comparable to those performed in physical environments, although differences in movement speed and precision have been reported (Everard et al. 2024; Knaut et al. 2009). Although these studies provide highly relevant information on how VR immersion affects movement, they focus exclusively on fully virtual environments. Consequently, it remains unclear whether VR-based assessments across intermediate levels of immersion can provide motor performance comparable to that obtained in the physical world.

Studying healthy individuals provides a controlled first step to isolate the effects of virtuality before extending VR-based assessment to clinical population. The hypothesis of the present study is that at least one of the evaluated virtual environments will produce motor performance comparable to that observed in the real environment during a reach-and-grasp task, supporting the use of VR headsets as a feasible tool for upper limb assessment. Therefore, the objective of this study was to compare motor performance across different levels of immersion defined by Milgram's continuum

and to determine which virtual environment most closely resembles the physical world using commonly reported kinematic variables. Milgram et al. described a spectrum of immersion ranging from a completely real environment to a fully virtual one, including intermediate states such as mixed reality (Milgram 1994).

Material and Methods

A total sample of 40 healthy adults (18 women and 22 men) participated in the study. Ethical approval was obtained from the Ethics Committee of the Universitat Politècnica de València (P16280225), and all participants signed an informed consent form prior to taking part in the study.

Based on Milgram's continuum, four different environments were defined and implemented, ranging from total reality to total virtuality (Figure 1).

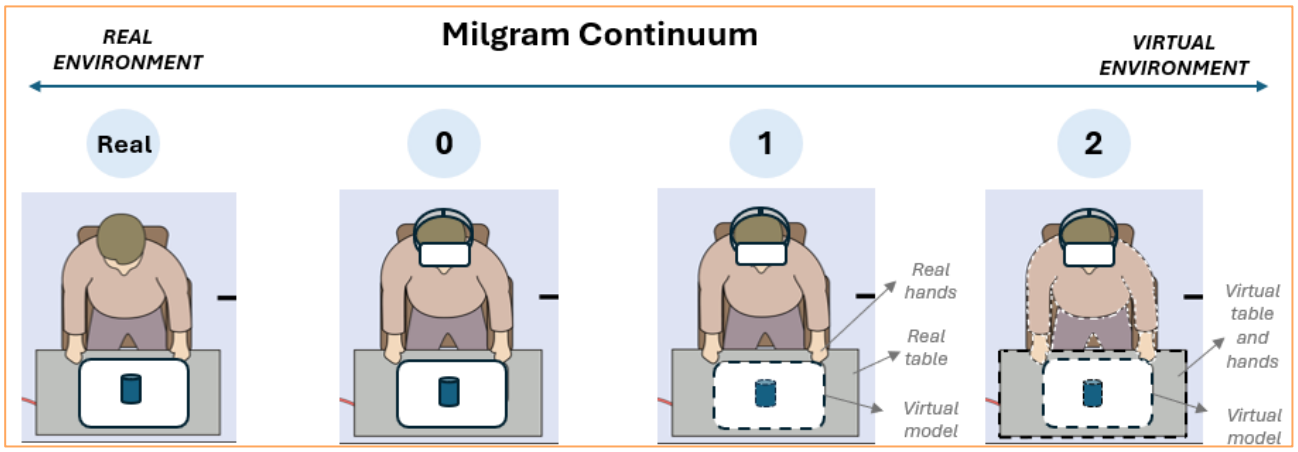


***Figure 92**. Representation of the four experimental environments along the Milgram Continuum: (Real) real environment, (0) passthrough environment, (1) mixed environment, and (2) virtual environment. A participant performing the task is shown in each condition.*

The task consisted of a reach-and-grasp activity. A cup (9 cm high, 7 cm diameter) was placed 30 cm from the edge of the table, and participants were required to grasp it and move it as far as possible. The VR environments were implemented using Unity (version 2022.3 10f1) for the Meta Quest 3 VR headset (Meta Platforms, Inc., California, USA). Data acquisition was conducted in a gold-standard three-dimensional photogrammetry motion capture laboratory (Optitrack, Natural Point, Corvallis, OR, USA). Six reflective markers were placed on anatomical landmarks: elbow, wrist and tip of the index finger of the dominant arm, at the sternum and both shoulders. Participants were seated on a chair without armrests or a backrest, with the elbow formed an angle of approximately 90º when the hand rested on the table edge. The setup was adjusted to allow participants to reach the target without trunk compensation.

The presentation order of the environments was randomized for each participant, ensuring that each environment appeared with equal frequency across all sequence positions. Motion trajectories were visually inspected, and missing data were interpolated using Motive software. Wrist-trajectory, which have been shown to provide a meaningful basis for assessing upper-limb function, were exported at 60 Hz (Lee et al. 2024). The exported data was smoothed using a local regression approach with cubic fitting, treating the movement trajectories as smooth functions affected by random errors.

Movement onset and offset were identified using a velocity threshold of 5 cm/s, combined with the detection of velocity peaks above the 25th percentile of each velocity profile. In the reaching task, two phases were analyzed separately: reaching, from the starting position to the cup, and grasp-and-transport, consisting of grasping the cup and moving it as far as possible. Movement performance was characterized by using spatiotemporal and kinematic variables: movement time, displacement, mean speed and maximum speed were calculated for each extracted movement or

task phase. Movement quality was assessed using the curvature index (CI) and the number of movement units (NMU).

Data normality was assessed using the Shapiro–Wilk test. As the data did not meet the normality assumption, the Friedman test was used to identify differences across the four environments. When significant differences were detected ($p < .05$), Wilcoxon signed-rank tests were performed for pairwise comparisons between the real environment and each VR condition. *P*-values were adjusted using the Bonferroni correction, with statistical significance set at adjusted $p < .05$.

Results

Pairwise comparisons against the physical world revealed that motor performance changes depended heavily on the task phase and the level of immersion. In the reaching phase we observe no differences between real and passthrough, while in the mixed and virtual a slight increase in movement time and displacement was observed (Figure 2).

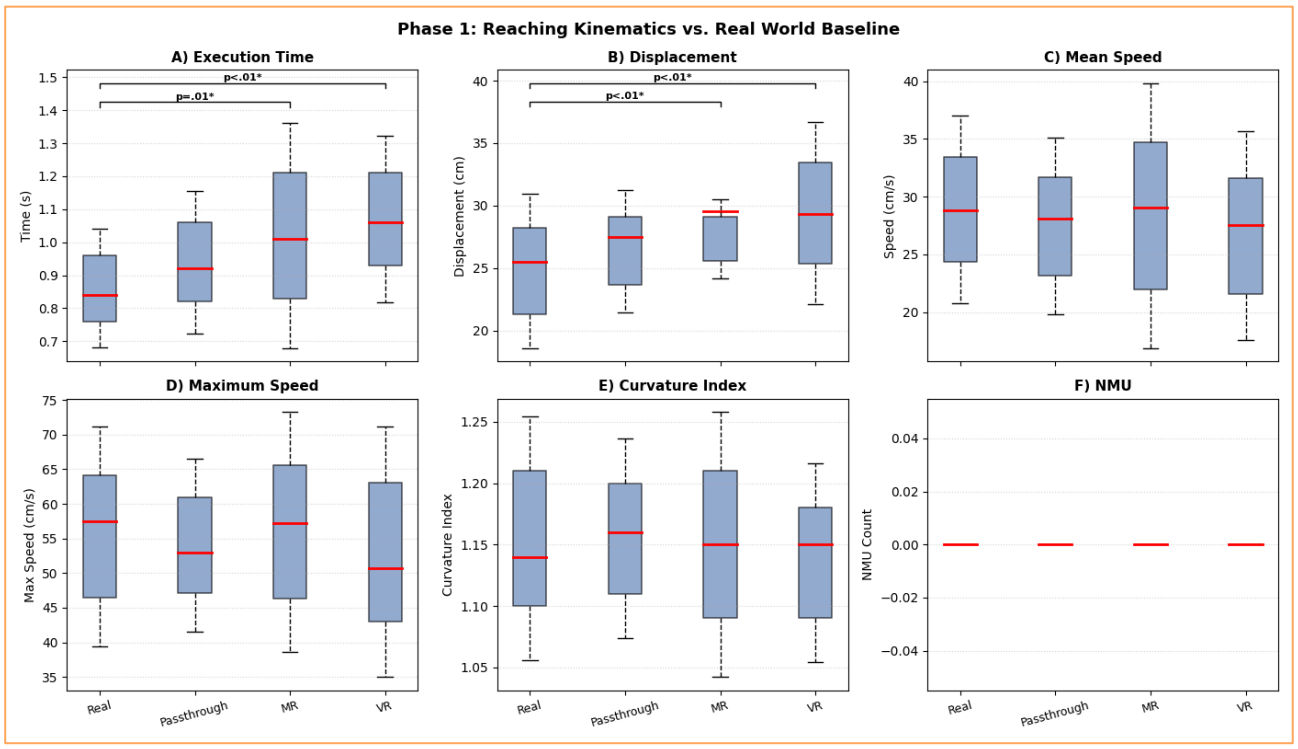


Figure 2. *Upper limb kinematic parameters during Reaching phase. Brackets highlight statistically significant pairwise comparisons against the physical world baseline (* $p < .05$, ** $p < .01$). Data represented as median and interquartile range (IQR). In panel F (NMU), boxplots appear as flat horizontal lines due to zero variance across all participants.*

In the grasping and transportation phase while passthrough exhibited no differences, mixed and virtual showed notable performance degradation characterized by reduced velocities, a sharp increase in trajectory curvature and loss of movement smoothness (Figure 3).

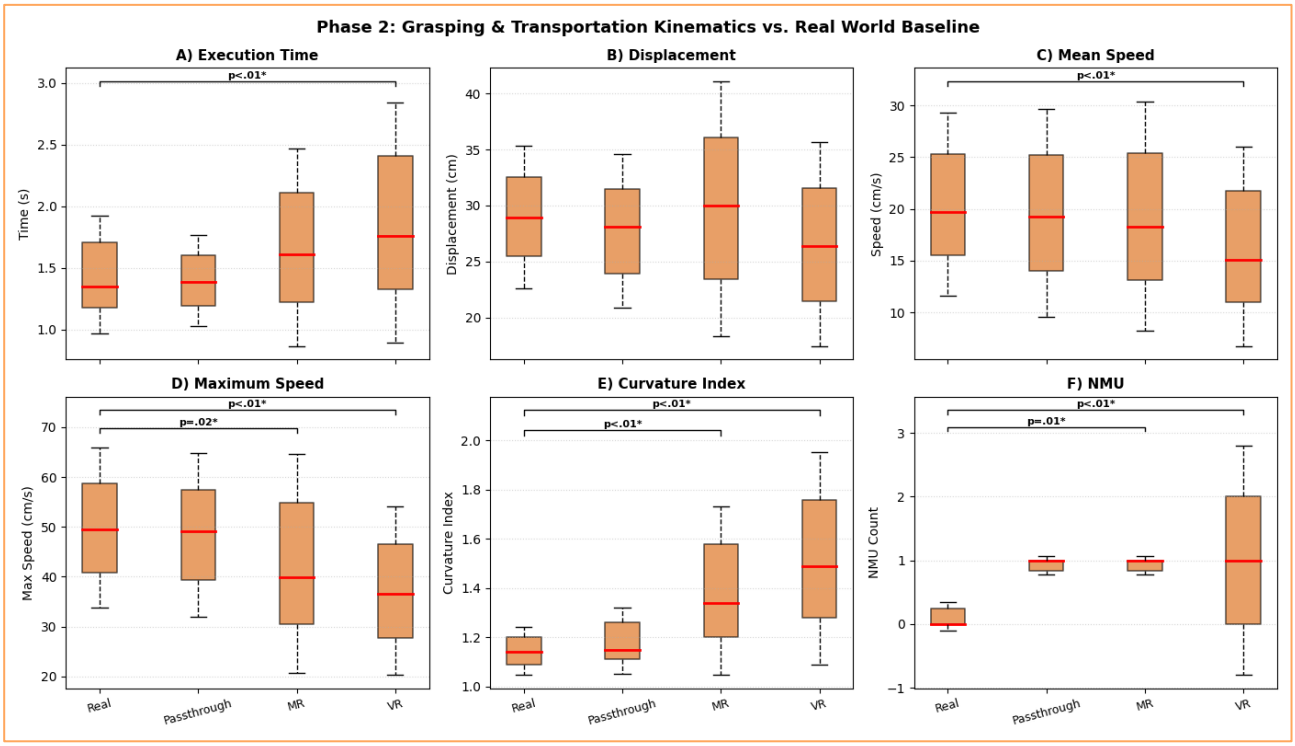


Figure 3. *Upper limb kinematic parameters during Grasping and Transportation phase. Brackets highlight statistically significant pairwise comparisons against the physical world baseline (* $p < .05$, ** $p < .01$). Data represented as median and interquartile range (IQR).*

Discussion

This study investigated whether VR environments can reproduce performance comparable to the physical world. The passthrough environment showed the highest agreement with the physical condition, whereas the introduction of virtual elements progressively altered motor performance, particularly during object manipulation.

During the reaching phase, movement time differed significantly between environments, while the remaining kinematic variables remained relatively consistent. This suggests that depth perception in the Meta Quest 3 may be sufficiently accurate to preserve most aspects of reaching performance. A recent meta-analysis concluded that in more modern head-mounted displays (HMDs), distance perception accuracy has improved compared to earlier devices (Kelly 2023). This is consistent with our results, where the only significant differences were observed in the time required to reach the glass, while other kinematic variables remained consistent across environments. However, the grasp-and-transport phase showed greater deterioration, particularly in mixed and fully virtual environments. The absence of physical and haptic feedback and the need to perform a pinch gesture to grasp virtual objects may require additional visuomotor adaptation (Jochmann et al. 2025).

The high agreement observed in the passthrough condition represents a key finding. By preserving real visual information and physical interaction while enabling headset-based motion tracking, passthrough may provide a suitable balance between ecological validity and objective kinematic assessment. Therefore, it appears to be the most promising condition for future HMD-based motor assessment.

Two main limitations should be considered. The lack of realistic tactile feedback may have affected grasping performance. Only healthy participants were included; therefore, future studies should determine whether these findings can be generalized to other populations.

References

Everard, G., Burton, Q., Van de Sype, V., Bibentyo, T. N., Auvinet, E., Edwards, M. G., ... & Lejeune, T. (2024). Extended reality to assess post-stroke manual dexterity: contrasts between the classic box and block test, immersive virtual reality with controllers, with hand-tracking, and mixed-reality tests. *Journal of neuroengineering and rehabilitation*, *21*(1), 36. https://doi.org/10.1186/s12984-024-01332-x.

Jochmann, E., Jochmann, T., Weber, M., Weigel, K., & Klingner, C. (2025). Impact of sensorimotor mismatch on virtual reality sickness and user experience: age-related differences in a randomized trial. *Journal of neuroengineering and rehabilitation*, *22*(1), 143. https://doi.org/10.1186/s12984-025-01677-x.

Kelly, J. W. (2022). Distance perception in virtual reality: A meta-analysis of the effect of head-mounted display characteristics. *IEEE transactions on visualization and computer graphics*, *29*(12), 4978-4989. https://doi.org/10.1109/TVCG.2022.3196606.

Knaut, L. A., Subramanian, S. K., McFadyen, B. J., Bourbonnais, D., & Levin, M. F. (2009). Kinematics of pointing movements made in a virtual versus a physical 3-dimensional environment in healthy and stroke subjects. *Archives of physical medicine and rehabilitation*, *90*(5), 793-802. https://doi.org/10.1016/j.apmr.2008.10.030.

Lee, S. I., Liu, Y., Vergara-Díaz, G., Pugliese, B. L., Black-Schaffer, R., Stoykov, M. E., & Bonato, P. (2024). Wearable-based kinematic analysis of upper-limb movements during daily activities could provide insights into stroke survivors' motor ability. *Neurorehabilitation and neural repair*, *38*(9), 659-669. https://doi.org/10.1177/15459683241270066.

Milgram, P. (1994). A Taxonomy of Mixed Reality Visual Displays. http://vered.rose.utoronto.ca/people/paul_dir/IEICE94/ieice.html.

Sathian, K., Buxbaum, L. J., Cohen, L. G., Krakauer, J. W., Lang, C. E., Corbetta, M., & Fitzpatrick, S. M. (2011). Neurological principles and rehabilitation of action disorders: common clinical deficits. *Neurorehabilitation and neural repair*, *25*(5_suppl), 21S-32S. https://doi.org/10.1177/1545968311410941.

A Virtual Reality System for Enhanced Indoor Fire Evacuation Training and Management

Moatasim Mahmoud[1], Athanasios Papanikolaou[1], Iason Sotiropoulos[1], and Stamatia Rizou[1]

[1]*Singular Logic*

Corresponding author: mmoatasim@singularlogic.eu

Keywords: VR training, fire evacuation, risk awareness, fire preparedness

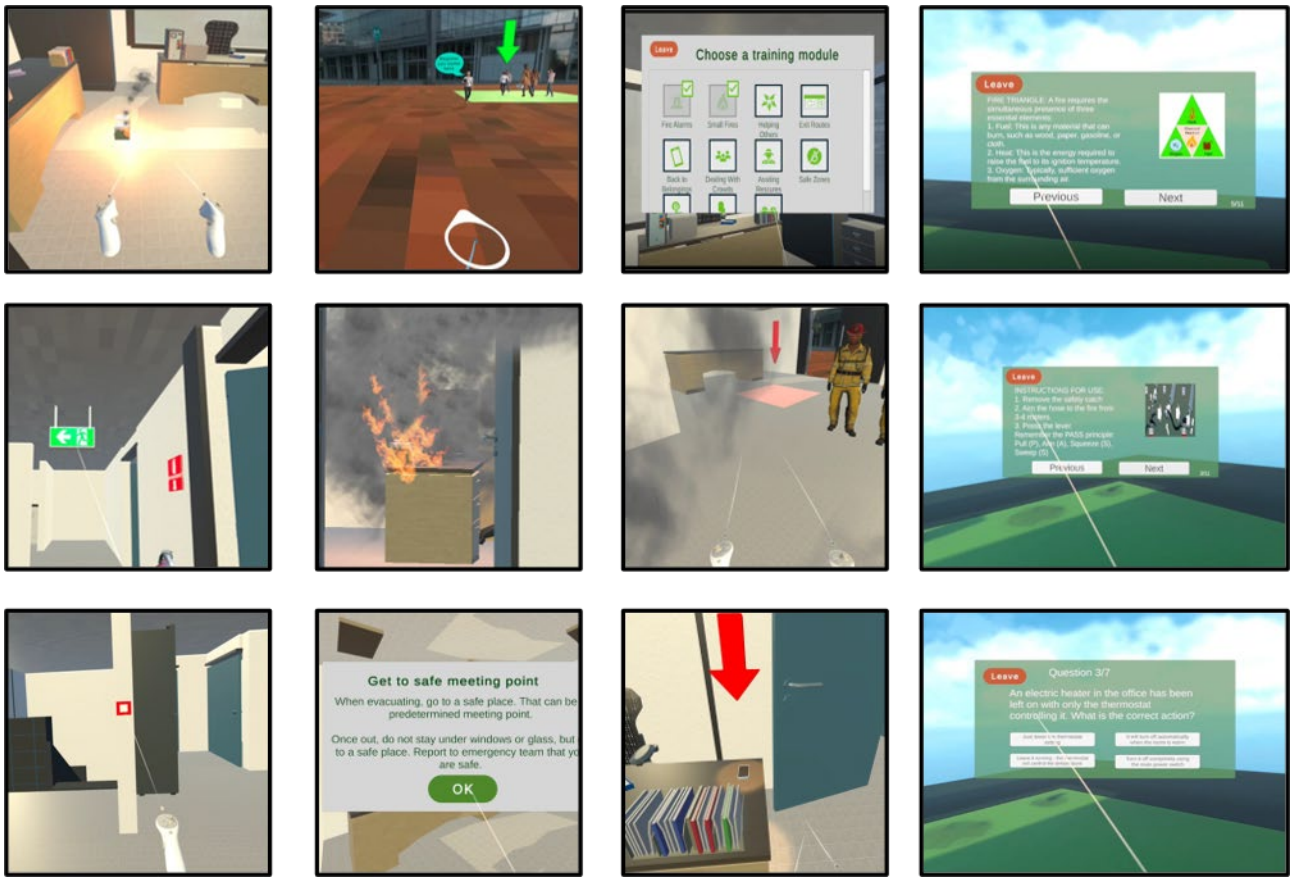


Figure 93. *Screenshots taken from the VR training sessions (user perspective)), showing examples on interactive procedural tasks, instructional module, and assessment module.*

This work demonstrates a virtual reality (VR) application designed for enhancing fire safety and evacuation training and management in indoor building environments. Through a complimentary approach, the system design and implementation provide trainees with interactive scenarios and guided procedural tasks, an instructional environment for knowledge acquisition, and an assessment module for testing learned material. To support building and safety managers, the system also includes a data collection layer, integrated with a centralized data management platform (DMP), supporting effective data-driven management and decision making. To evaluate our system and showcase its value, we conducted a subjective experiment and analyzed the training data collected from 27 users. The results prove the system's capabilities at improving learning outcomes, while offering crucial tools for safety management.

Benefits of VR-based training have been demonstrated across several domains, showcasing its role in acquiring knowledge and building new skills through interactive repeatable sessions (Thompson et al., 2026; Górski et al., 2025; Sharma et al., 2026). Immersive training has been of particular interest for risk preparedness as immersive systems can simulate high-stake real-life scenarios within safe, controlled, and affordable settings (Sparvoli et al., 2026; Kaggwa, et al., 2025). While multiple works have been dedicated to fire safety and evacuation preparedness (Mystakidis et al., 2022; Braun et al., 2022; Sun et al., 2025), most existing works are focused on specific procedural tasks (e.g., fire extinguishing) or limited aspects (e.g., evacuation routes) of the fire preparedness training. Moreover, most existing systems rely on self-reported efficacy and/or pre- and post-tests for assessing the training performance, neglecting vital insights that can be utilized from events and interactions within the VR training sessions. To address these limitations, we develop a VR fire safety and evacuation training application with complementary design that covers theoretical knowledge and procedural tasks over wide topics and principles of fire safety and evacuation. We followed a modular approach so that new training material can be easily added. Moreover, we implement a data collection layer and integrate it with a centralized DMP that records performance logs, quiz results, task completion times, and evacuation trajectories from training sessions. The system therefore is designed to improve VR training and management for fire safety and evacuation in indoor office building environments.

System Design and Implementation

The **Interactive Procedural Tasks** module comprises 15 hands-on scenarios covering key aspects of fire safety and evacuation. Training tasks include: Recognizing fire alarms and executing immediate response actions; Identifying and suppressing small, manageable fires; Assisting fallen or injured individuals; Navigating designated escape routes using exit signage; Managing crowd dynamics and prioritizing emergency personnel; Assembling at safe meeting points and coordinating with emergency response teams; Activating manual fire alarm call points; Applying smoke-inhalation mitigation techniques. Additionally, four tasks introduce routine preventive responsibilities, such as inspecting electrical appliances and maintaining clear evacuation paths. Some examples are presented in Figure 1. Each task requires trainees to complete specific actions based on scenario objectives. Each procedural module begins with an introductory overview of the scenario, objectives, and instructions delivered via multimodal (text and audio) cues to balance cognitive load in Virtual Reality (VR). The trainee then must perform the required action (e.g., reach a safety zone, or extinguish a small fire). Upon task completion, a feedback panel accompanied by voiceover reinforces correct behaviors and delivers a final summary point.

The **Instructional Environment Module** provides a knowledge base to complement practical training with essential theoretical concepts, delivered via a spatial UI using text, images, and voice narration. The content addresses fire causes, fire extinguisher classifications and proper usage, safe suppression techniques, and foundational evacuation principles, such as alarm recognition and route navigation. Information is organized into modular sections, each featuring a title panel followed by a sequence of spatial cards combining concise text with representative 2D illustrations. To enhance comprehension without inducing cognitive overload, voiceover narration reinforces key concepts rather than solely relying on texts.

The **Assessment Module** evaluates fire safety knowledge acquired during the instructional and procedural phases. The module utilizes a randomized multiple-choice format focusing primarily on declarative knowledge, such as selecting appropriate extinguisher types. When an answer is submitted, the system provides immediate feedback—confirming correct responses or offering concise explanations for incorrect choices to reinforce key learning points. Upon completing the quiz, a total score is generated to reflect overall theoretical understanding. The assessment interface employs a modular architecture that allows for straightforward expansion of the question pool.

Data Collection & Data platform integration: The solution implements a data collection layer for recording trainees' answers to in-app tests, task completion times, and evacuation trajectories. To support post-training analysis and long-term management of evacuation data, the VR application was integrated with a centralised Data Management Platform (DMP) based on CKAN, the world's leading open-source data portal, which stores and shares all training-related outputs posted as CSV via UnityWebRequest automatically when a session completes. Each session is tied to a dedicated dataset entry, letting facility managers, EMTs, and researchers access historical records, compare behaviour across sessions, and identify recurring bottlenecks, turning the application into a continuous evidence-based safety-monitoring system. The integration follows a metadata-driven approach aligned with the DCAT standard, with resources such as trajectory files and aggregated analytics stored as individual, richly described entries; role-based access control lets visibility be set to public or private. Figure 2 shows an example of the collected data on the platform. This collection, management, and analysis of this data is vital for exposing risk areas and unclear procedures, helping safety managers address these issues proactively.

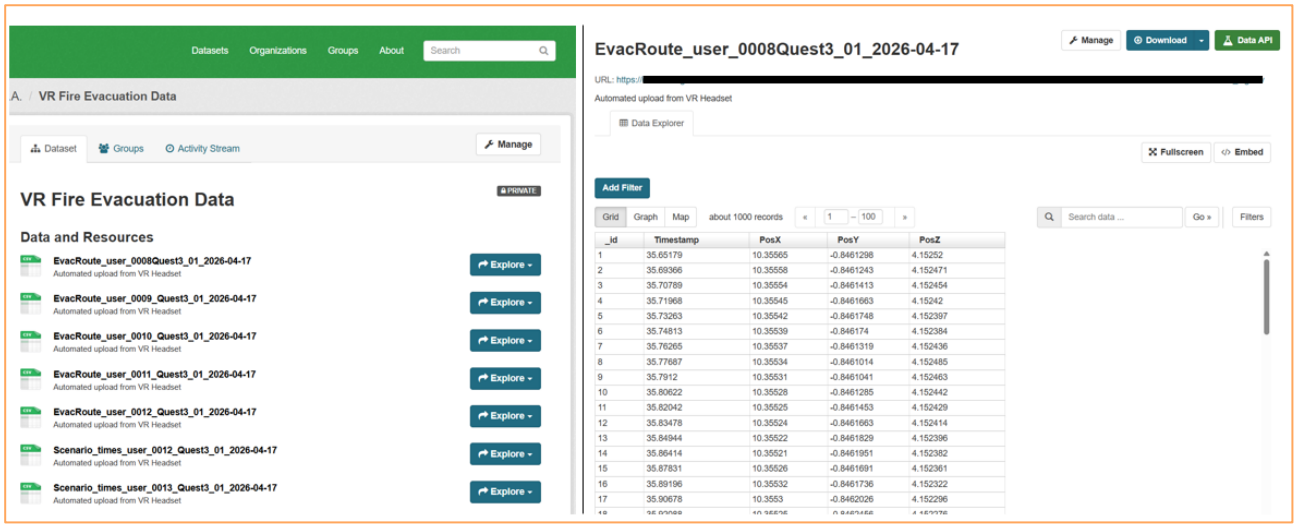


Figure 94. *The DMP with logs automatically sent from the VR headset at the end of training sessions.*

Subjective Assessment

Our subjective assessment with 27 participants demonstrates the system's capability at improving the training performance.

In addition to data collected from the application, we asked participants to fill in a post-training questionnaire and respond to a pre- and post-training test. Analysis of the this data combined, demonstrates substantial impact and high performance. Based on the pre- and post training test, the intervention produced a mean improvement of +17.8 pp across the 16 items (69.5%→87.3%). In terms of self-reported learning efficacy, based on 13 Likert items, the system scored 4.27/5 (SD = 0.39), indicating high confidence in perceived effectiveness. The system usability, following the System Usability Scale (SUS), averaged 80.3 (SD = 13.04, Median = 85). Moreover, analysis of scenario completion times and movement behavior offered important insights into inter-individual variability and spatial orientation, with indication on where the training should be improved further.

Acknowledgement

This work was supported by the CLIMRES “Leadership for Climate Resilient Buildings” Horizon Innovation Actions project supported by the European Commission under grant agreement No. 101147777.

References

Braun, P., Grafelmann, M., Gill, F., Stolz, H., Hinckeldeyn, J., & Lange, A. K. (2022). Virtual reality for immersive multi-user firefighter-training scenarios. *Virtual reality & intelligent hardware*, *4*(5), 406-417. https://doi.org/10.1016/j.vrih.2022.08.006

Górski, F., Gapsa, J., Grajewski, D., Zawadzki, P., Maik, M., Sobociński, P., ... & Walczak, K. (2025). The impact of virtual reality on technical training: A case study in live-line maintenance. *IEEE Access*, *13*, 80019-80032. https://doi.org/10.1109/ACCESS.2025.3566937

Kaggwa, M. M., Chaimowitz, G. A., Agboinghale, P., Marginean, D., & Olagunju, A. T. (2025). Virtual reality training programs in disaster preparedness: a systematic review. *Discover Education*, *4*(1), 325.

Mystakidis, S., Besharat, J., Papantzikos, G., Christopoulos, A., Stylios, C., Agorgianitis, S., & Tselentis, D. (2022). Design, development, and evaluation of a virtual reality serious game for school fire preparedness training. *Education Sciences*, *12*(4), 281. https://doi.org/10.3390/educsci12040281

Sharma, V., Park, S., Voinescu, A., & Jacobs, C. (2026). Medical students’ experiences with virtual reality simulation training: Qualitative study. *JMIR Medical Education*, *12*(1), e74301. https://doi.org/10.2196/74301

Sparvoli, G., Bernardini, G., Crescimbeni, L., Balducci, S., Bussaglia, A., Quagliarini, E., & Serroni, S. (2026, July). An Immersive Virtual Reality Serious Game for Preparing Citizens to Flood Emergencies in Urban Built Environments. In *2026 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv)* (pp. 305-310). IEEE. https://doi.org/10.1109/MetroLivEnv70468.2026.11659913

Sun, L., Lee, B. G., & Chung, W. Y. (2025). Enhancing fire safety education through immersive virtual reality training with serious gaming and haptic feedback. *International Journal of Human–Computer Interaction*, *41*(9), 5607-5622. https://doi.org/10.1080/10447318.2024.2364979

Thompson, K., Turkay, S., Brown, R., Winter, G., Burden, A., Burland, S., ... & Tay, A. (2026). The effectiveness of virtual reality (VR) for construction skills training. *Australasian Journal of Educational Technology*, *42*(1), 58-78. https://doi.org/10.14742/ajet.10693

A.R.C.O.: A Multimodal XR-LLM Cognitive Assistant for Technical Inspections in Virtual Worlds

Mathias Proboste-Martínez[1], Felipe Muñoz-La Rivera[2,3,*], Nancy Rodriguez[4] and Javier Mora-Serrano[5]

[1] *School of Civil Engineering, UPC - Barcelona Tech, Spain*

[2] *School of Civil Engineering, Pontificia Universidad Católica de Valparaíso, Chile*

[3] *Millennium Institute for Care Research (MICARE), Chile*

[4] *LIRMM, CNRS, University of Montpellier, France;* [5] *CIMNE, Barcelona, Spain.*

Corresponding author: felipe.munoz@pucv.cl

Keywords: Extended Reality (XR), Large Language Models (LLMs), Cognitive Digital Twins, Human-AI Teaming, Construction 5.0

Introduction and Operational Motivation

Technical auditing of built infrastructure is a high-stakes operational domain where undetected geometric or normative deficiencies compromise safety and economic performance. In civil engineering and social housing compliance, inspectors must evaluate physical geometries against dense regulatory frameworks. Despite digitalization in the AECO sector, field inspection remains fragmented: 3D geometries reside in Building Information Modeling (BIM) databases, while regulations remain isolated in static documents (Boje et al., 2020; Mora-Serrano et al., 2020). Consequently, inspectors act as manual data integration engines (Sepasgozar, 2021), incurring substantial cognitive load (Sweller, 1994) through procedural friction (Digital Muda; Dave et al., 2016).

While Extended Reality (XR) provides immersive visualization, conventional platforms operate as analytically passive displays (Delgado et al., 2020; Muñoz-La Rivera et al., 2021). Conversely, Large Language Models (LLMs) in engineering operate largely through two-dimensional desktop interfaces detached from spatial context (Wu et al., 2024; Zheng et al., 2023). Lacking physical grounding, generative models exhibit spatial blindness and yield technical hallucinations (Ji et al., 2023). This work presents A.R.C.O. (Augmented Reasoning Construction Operation), an XR-LLM framework that embeds an Intelligent Avatar into Virtual Worlds to assist technical auditing.

Architecture and Multimodal Grounding

A.R.C.O. adopts a decoupled client-server architecture isolating Unreal Engine 5 graphics rendering from asynchronous multi-agent reasoning on a Python FastAPI backend. Multimodal interaction relies on a native Unreal line-trace (raycast) emitted from the VR motion controller. Intersecting a 3D component collider extracts its unique BIM identifier (IfcGUID or bim_id) preserved from Revit Datasmith (Epic Games, 2024). Concurrently, speech is transcribed locally via the Azure

Speech plugin (Microsoft, 2024). Packaging spatial telemetry, room context, and transcribed text establishes a deterministic grounding mechanism that binds conversational intent directly to project databases.

At the backend, LangGraph state graphs orchestrate execution (LangChain, 2024). An entry supervisor routes queries across specialized expert nodes: chat_expert handles general dialogue using short-term memory, bim_expert queries IFC tables for dimensional and material properties, and norm_expert retrieves regulatory standards from FAISS using Google Generative AI embeddings (Johnson et al., 2019). Deterministic Priority Logic enforces safety: inspected IfcGUID attributes override probabilistic language generation, partition removals require verified non-load-bearing metadata with MEP warnings, and absent data triggers explicit uncertainty disclosures ('information unavailable') to prevent speculative inference.

Demonstration Scenario: Accessibility Auditing

The operational workflow was evaluated in a regulatory compliance audit based on the 'Awen' social housing project, designed under Chilean Universal Accessibility standards (DS No. 49; MINVU, 2016), depicted in Figure 1.



Figure 1. *Immersive inspection workflow in the Awen social housing scenario: (a) Level 1 voice induction with the Intelligent Avatar and smartwatch HUD; (b) Level 1 geometric calibration with the virtual measuring tape; (c) Level 2 entrance overview and ramp audit; (d) Level 2 in-situ measurement of door opening width; (e) Level 2 verification of corridor clearances and kitchen turning area; and (f) Level 2 restroom transfer evaluation via laser raycasting.*

In Level 1, the operator enters an onboarding environment to calibrate voice interaction with the Intelligent Avatar, verify speech capture on the smartwatch HUD (Figure 1a), and calibrate virtual measurement tools on reference geometries (Figure 1b). In Level 2, the user audits the housing unit, verifying exterior ramp slopes (Figure 1c), measuring door clear widths (Figure 1d), checking corridor clearances and the 1.50 m kitchen turning radius (Figure 1e), and evaluating restroom transfer zones via laser raycasting (Figure 1f).

Empirical Performance and Latency Benchmarks

System responsiveness was profiled across 585 live interactions, comparing local Edge execution against cloud-based generative inference (Llama 3.3 70B; Meta AI, 2024). Cloud deployment achieved an average response latency of 0.61 s (median 0.54 s, SD 0.28 s) during expert stress testing and 1.68 s (median 1.44 s, SD 0.81 s) during student evaluations, approaching conversational turn-taking thresholds. In contrast, local edge execution on workstation hardware suffered from computational contention between graphics rendering and on-device model execution, yielding a dispersed latency averaging 11.35 s (median 10.82 s, SD 3.42 s), which disrupts conversational fluency. Furthermore, query complexity did not cause proportional latency increases, confirming that responsiveness is governed primarily by context retrieval pipelines and network transport rather than linguistic complexity.

Discussion: Human-AI Teaming in Challenging Environments

Human-AI Agency: In safety-critical auditing, the avatar functions strictly as a consultative copilot, while formal compliance verdicts remain with the human engineer (European Commission, 2021). Articulating technical reasoning and forcing deterministic uncertainty disclosures mitigates Automation Bias (Parasuraman & Manzey, 2010), fostering calibrated trust under Construction 5.0 Human-in-the-Loop principles (Muñoz-La Rivera et al., 2021).

Attentional Ergonomics: Conventional XR interfaces relying on floating menus or handheld tablets induce visual divided attention. Shifting procedural and normative inquiries to the voice channel preserves the user's visual field (Schubert et al., 2001), maintaining focus on physical geometries while the wrist smartwatch provides unobtrusive status cues.

Industrial Deployment Tensions: Field deployment highlights three core tensions: (1) network sensitivity versus edge constraints, where cloud execution requires high bandwidth vulnerable in remote or shielded sites, while mobile headsets cannot yet run large models locally; (2) digital waste (Digital Muda), where parametric errors in source BIM data propagate directly into agent responses; and (3) conversational drift, where multi-turn dialogues require bounded memory buffers to prevent parameter cross-contamination across inspected elements.

Conclusion and Demonstration Scope

A.R.C.O. demonstrates the viability of coupling game engines with multi-agent generative architectures to create a spatially grounded cognitive assistant for technical inspections. Anchoring voice queries to physical BIM elements via collision raycasting and enforcing deterministic priority safeguards eliminates hallucinations while maintaining interactive response latencies. At EuroXR 2026, this contribution provides actionable lessons and live demonstration workflows for trustworthy Human-AI teaming in mission-critical virtual workspaces.

Acknowledgments

The first author acknowledges the National Agency for Research and Development (ANID) / Scholarship Program / DOCTORADO NACIONAL/2024 - 72240171. The authors acknowledge the support of the AWEN project, funded by the FONDEF IDEA R&D Program of ANID, Chile, Grant ID26I10835.

References

Boje, C., Guerriero, A., Kubicki, S. & Rezgui, Y. (2020). Towards a semantic construction digital twin: Directions for future research. *Automation in Construction*, 114, 103179.

Dave, B., Kubler, S., Främling, K. & Koskela, L. (2016). Opportunities for enhancing lean construction management using internet of things standards. *Automation in Construction*, 61, 86–97. https://doi.org/10.1016/j.autcon.2015.10.009

Delgado, J.M.D., Oyedele, L., Demian, P. & Beach, T. (2020). A research agenda for augmented and virtual reality in architecture, engineering and construction. *Advanced Engineering Informatics*, 45, 101122. https://doi.org/10.1016/j.aei.2020.101122

Epic Games. (2024). Datasmith overview and Unreal Engine 5 documentation. https://dev.epicgames.com/documentation/en-us/unreal-engine/datasmith-overview.

European Commission. (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. *Technical Report, DG Research and Innovation*.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Yan, Y., Ishii, E., Bang, Y., Dai, W. & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38. https://doi.org/10.1145/3571730

Johnson, J., Douze, M. & Jégou, H. (2019). Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, 7(3), 535–547. https://doi.org/10.1109/TBDATA.2019.2921572

LangChain. (2024). LangGraph: Multi-agent workflows as state graphs. https://github.com/langchain-ai/langgraph.

Meta AI. (2024). The Llama 3 herd of models. arXiv:2407.21783.

Microsoft. (2024). Azure AI Speech Services: Speech-to-text. https://learn.microsoft.com/azure/ai-services/speech-service/.

MINVU. (2016). Decreto Supremo Nº 49: Fondo Solidario de Elección de Vivienda - Accesibilidad universal. Ministerio de Vivienda y Urbanismo, Chile.

Mora-Serrano, J., Muñoz-La Rivera, F. & Valero, I. (2020). Design science research in engineering education: Methodological guidelines. Developments in the Built Environment, 3, 100016.

Muñoz-La Rivera, F., Mora-Serrano, J. & Oñate, E. (2021). Conceptual framework for BIM-VR integration in Construction 5.0. Archives of Computational Methods in Engineering, 28, 4501–4521.

Parasuraman, R. & Manzey, D.H. (2010). Complacency and bias in human use of automation. Human Factors, 52(3), 381–410. https://doi.org/10.1177/0018720810376055

Schubert, T., Friedmann, F. & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. Presence: Teleoperators and Virtual Environments, 10(3), 266–281.

https://doi.org/10.1162/105474601300343603

Sepasgozar, S.M. (2021). Differentiating digital twin from digital shadow: Methodological guidelines. Buildings, 11(8), 359. https://doi.org/10.3390/buildings11040151

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Wu, J., Zhang, C. & Zhang, J. (2024). LLM-based multi-agent collaboration for automated construction schedule generation. *Advanced Engineering Informatics*, 62, 102654.

Zheng, Z., Lu, X. & Chen, K. (2023). ChatGPT for construction safety: Evaluating the potential of large language models. *Safety Science*, 168, 106325.

LoomaXR: An Engine-Agnostic Hub for Collaborative XR Scene Creation

Mercè Álvarez de la Campa Crespo, Chrysostomos Chadjiminas, and Panayiotis Charalambous

CYENS – Centre of Excellence, Nicosia, Cyprus

Corresponding author: m.alvarez@cyens.org.cy

Keywords: Collaborative XR; interoperability; shared virtual environments; heterogeneous clients; semantic asset retrieval; 3D content generation

Abstract

LoomaXR is a collaborative platform that lets applications built with different technologies work together on the same live XR scene. A central authoritative hub holds the scene and keeps every connected application in sync, so multiple people can work on it at the same time and see each other's changes as they happen. A client contract defines how an application joins, exchanges scene information, and responds to updates. For example, a web client and an Unreal Engine client that implement the same contract can join the same scene while choosing how to render it and how people interact with it. Services built on the hub help participants find assets by describing what they need in text or speech, or by providing an image. When nothing suitable is available, they can generate a new object. Participants share what the scene contains and how it is organised, but not a fixed view of it. Each client can present the same scene differently, from high-quality graphics to audio or haptic feedback.

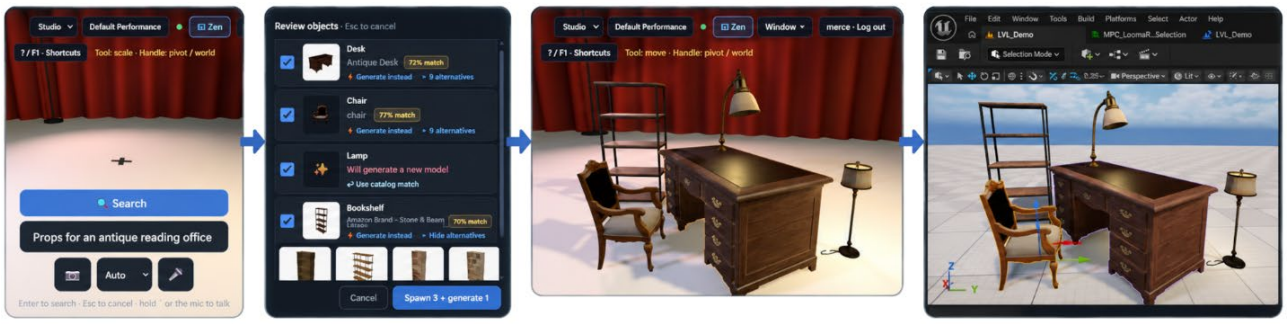


***Figure 95**. LoomaXR workflow. Users describe the scene they want, review suggested or newly generated objects and collaboratively arrange them in a shared scene. The same scene can be viewed and edited through different clients, illustrated here by the web and Unreal Engine clients.*

Introduction and contribution

Creating an XR scene can involve people with different roles and preferred tools. One participant may arrange the scene in a browser, while another views or develops it in a game engine. These applications can exchange assets, but they do not normally provide a shared live workspace. LoomaXR provides a central hub that keeps the scene in sync while allowing each application to control its visual presentation and interactions.

Its main contribution is an engine-agnostic approach to collaboration, supported by optional search and generation services that are not tied to a single 3D cat software tool or game engine. LoomaXR is being developed within the HAMLET project, where collaborative stage design for dance and theatre is its first application. The underlying approach is intended for a wider range of XR projects.

Related work and positioning

Shared editing of 3D scenes is not new, but existing systems set different thresholds for participation. Distributed virtual environment architectures established replicated scene graphs and update propagation decades ago (Frécon and Stenius 1998), generally assuming one homogeneous application. Unreal Multi-User Editing synchronises edits across peers but expects every participant to run the same engine, in practice the same version and project (Epic Games 2024). OpenUSD and hosted implementations of it offer a rich, engine neutral interchange and live layering model (Pixar Animation Studios 2023), at the cost of a substantial runtime in every client and, in practice, a vendor stack. LoomaXR takes a middle-ground approach. Its shared scene format is simple enough to work across web browsers, game engines and headset applications. Guests can also join a scene from a phone by opening a shared link without creating an account.

Collaborative scene creation and client contract

In LoomaXR the hub holds the scene, and every object in it has a shared identity, so every connected application knows which object has changed. When a participant adds, moves, groups, renames, or removes an object, the change appears for everyone else working in that scene.

The client contract deliberately distinguishes core scene exchange from optional capabilities. A client implements only what its role needs. Rendering conventions are outside the contract: two clients may differ in materials, lighting or level of detail while agreeing on object identity, hierarchy and transform. For example, the Three.js and Unreal Engine clients can render their own visual representations of the same authoritative scene state, while preserving the same objects, hierarchy and transforms (see Figure 3). The mandatory part is therefore a statement of what makes the experience shared, which object is which, how objects stand in relation to one another, and the order in which changes apply; everything beyond it, including whether the scene is drawn at all, belongs to the client.

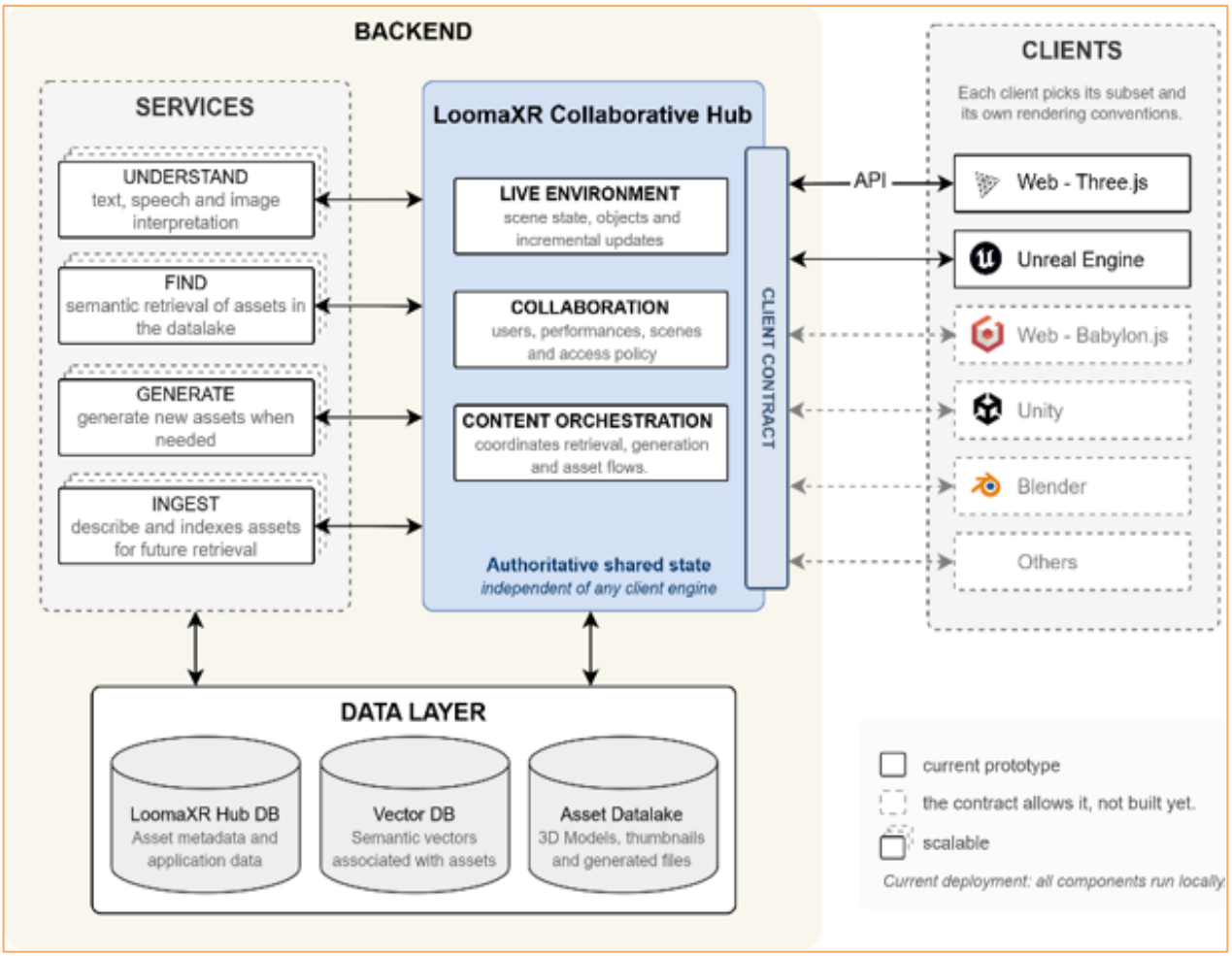


***Figure 96**: LoomaXR architecture: one authoritative hub between modular services and heterogeneous clients.*

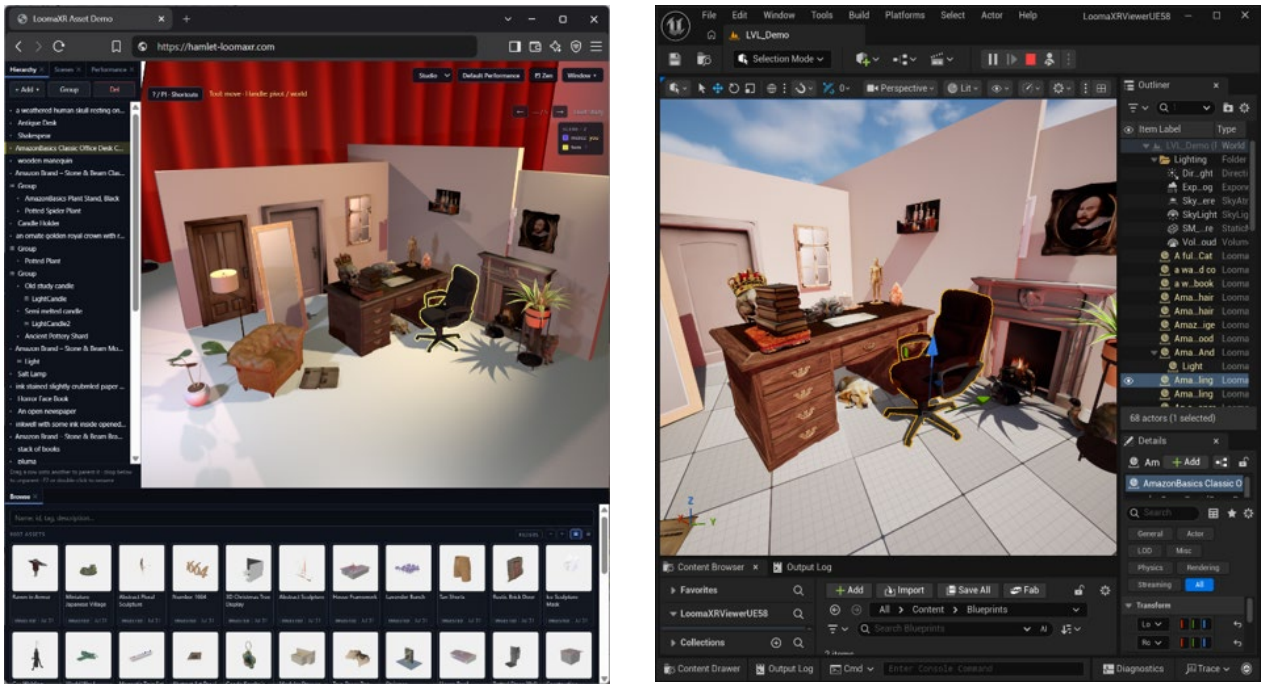


***Figure 97.** Two participants rendering the same scene, one in Unreal Engine, and the other using a web browser.*

Working on the same scene does not mean that every participant must experience it in the same way. A performer may need cues, a lighting designer may need controls, and a director may need annotations and a running order. An audience-facing client may present only the final scene. LoomaXR keeps the important scene information shared while allowing each client to present it in a form suited to its users. This flexibility could also support clients that communicate the scene through speech, spatial audio, or haptic feedback instead of graphics. We have not yet developed such a client, but the shared scene already contains information about object names, positions, relationships, and descriptions. Exploring these different experiences is part of our future work.

Prototype and content services

The web client provides the most complete environment for creating and organising scenes. Participants can find assets using text, speech, or images, including requests containing several objects. For example, a request “props for an antique office,” provides a set of suggested objects, as shown in Figure 5. LoomaXR searches a catalogue of about 9,000 assets from openly licensed 3D collections (Collins, et al. 2022), (Downs, et al. 2022). When a suitable asset is not available, participants can generate a new one through the connected generation service, as shown in Figure 4 (Xiang, et al. 2025). Retrieved or generated objects can then be added and arranged directly in the shared scene. The Unreal Engine client connects to the same shared scene and renders it using its own visual pipeline (see Figure 3). Participants can also access the generation service from Unreal Engine to create and add new objects directly to the scene.

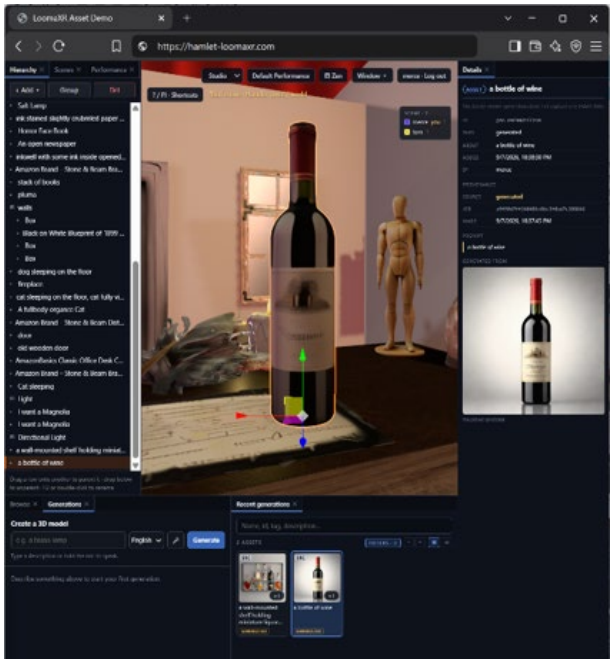


Figure 98. *Generated asset using the prompt “a bottle of wine”.*

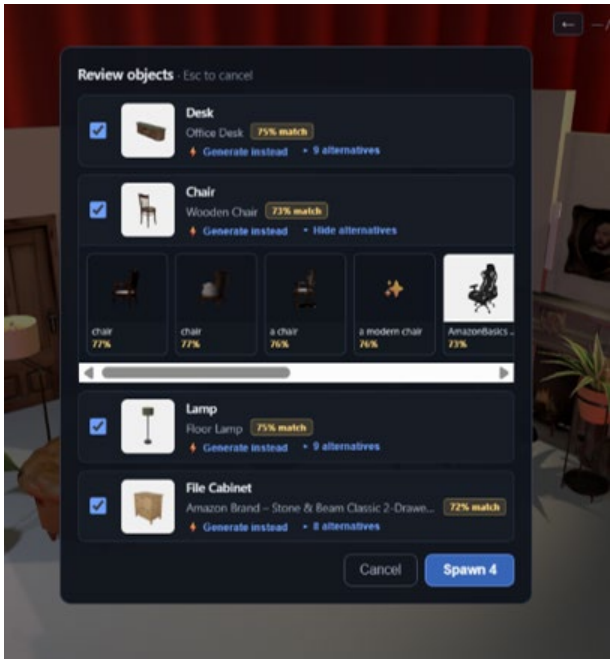


Figure 99. *A user searched for “props for an antique office”. The request suggests multiple objects which the user can select.*

Limitations, future work and conclusion

Content generation currently processes one object at a time because the available setup uses a single GPU. Additional requests wait in a queue until the current generation is complete. Future work will support several generation requests at once, extend LoomaXR to more applications such as Unity and Blender, and evaluate the system with more participants under realistic production conditions.

LoomaXR allows applications built with different technologies to collaborate through one live scene. The hub keeps the shared scene consistent, while each client retains control over how it is presented and experienced. Search and generation services extend this environment and support a more flexible creative process.

Acknowledgements

This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101178362 (HAMLET-HORIZON-CL2-2024-HERITAGE-01-03). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the granting authority. Some of the 3D models used for the demo are from the Amazon Berkeley Objects dataset (© Amazon.com, Inc.) and Google Scanned Objects (© Google LLC), both used under CC BY 4.0.

References

Collins, J., Shubham G., Kenan D., Achleshwar L., Xu, L., Gundogdu, E., Zhang, X., … Malik, J. (2022). ABO: Dataset and Benchmarks for Real-World 3D Object Understanding. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 21126–21136. https://doi.org/10.1109/CVPR52688.2022.02045

Downs, L., Francis, A., Koenig, N., Kinman, B., Hickman, R., Reymann, K., ... & Vanhoucke, V. (2022, May). Google scanned objects: A high-quality dataset of 3d scanned household items. In *2022 international conference on robotics and automation (ICRA)* (pp. 2553-2560). https://doi.org/10.1109/ICRA46639.2022.9811809

Epic Games. (2024). Multi-User Editing in Unreal Engine. *Multi-User Editing in Unreal Engine.*

Frécon, E., & Stenius, M. (1998). DIVE: A scaleable network architecture for distributed virtual environments. *Distributed Systems Engineering*, *5*(3), 91-100. https://doi.org/10.1088/0967-1846/5/3/002

Khronos Group. (2021). glTF 2.0 Specification, version 2.0.1. *glTF 2.0 Specification, version 2.0.1.*

Pixar Animation Studios. (2023). Universal Scene Description (OpenUSD). *Universal Scene Description (OpenUSD).*

Stefik, M., Bobrow, D. G., Foster, G., Lanning, S., & Tatar, D. (1987). WYSIWIS revised: Early experiences with multiuser interfaces. *ACM Transactions on Information Systems (TOIS)*, *5*(2), 147-167. https://doi.org/10.1145/27636.28056

Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., ... & Yang, J. (2025, June). Structured 3d latents for scalable and versatile 3d generation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 21469-21480). IEEE. https://doi.org/10.1109/CVPR52734.2025.02000

.

XR Skills Between Campus and Workplace in Ten Countries

Javier Sánchez García on behalf of the Metaverse Academy Consortium

Universitat Jaume I Castelló de la Plana Spain

Corresponding author: jsanchez@uji.es

Keywords: Extended reality; XR skills; higher education; industry needs; mixed methods; AR and VR

Introduction

Extended reality (XR) is now used beyond entertainment in training, healthcare, manufacturing, retail and cultural heritage. Yet exposure in higher education remains uneven and is often limited to demonstrations rather than supervised practice. This extended abstract presents the needs analysis that informed the Metaverse Academy, a three-year Erasmus+ Alliance for Innovation. The study examines the gap between students' practical experience of XR and the skills employers report needing when they introduce XR into real workflows.

Method

The study used a sequential mixed-method design. A harmonised online survey was completed by 2,039 higher-education students from 13 partner organisations in ten countries between May and October 2024. The 19-item questionnaire covered XR awareness and use, the place of XR in degree programmes, perceived competences, barriers and attitudes towards online learning. Partners translated and reviewed the instrument before fieldwork; responses were anonymous. In September and October 2024, the consortium also conducted 200 semi-structured interviews with staff from 160 companies. Interviewees included founders, general managers and project leads from XR specialists and organisations in manufacturing, education, healthcare, retail, cybersecurity and heritage. Student data were compared across national samples and interview material was coded thematically. The sample is partner-recruited rather than probabilistic, so the results should be read as a cross-country profile rather than population estimates.

Findings

The student results show a clear awareness-to-practice gap. Across the full sample, 29% had never heard of XR, 31% had heard of it but never used it, and 18% had only seen a demonstration. Only 5% reported regular or experienced use. National differences were substantial: 44% of respondents in South Africa and 38% in Spain had never heard of XR, whereas 54% of the Swedish sample had used it at least a few times. Nevertheless, systematic use in higher education was rare. Forty-three per cent said XR was not used at all in their programme and a further 27% said it was used rarely. Only 11% reported frequent or very active use. Students generally believed that XR could benefit their learning and their country, but they identified equipment cost, limited training, technical difficulties and time as practical barriers.

Table 1. *Student experience with XR technologies (% of national sample)*

Country	Never heard	Heard, never used	Demo only	Used a few times	Regular or experienced use
Türkiye (449)	26	31	23	15	5
Romania (310)	16	31	23	27	3
Spain (470)	38	24	18	17	3
South Africa (426)	44	36	12	5	3
Bulgaria (103)	11	36	17	22	15
Germany (31)	19	29	23	23	7
Greece (103)	14	47	18	16	7
Sweden (40)	13	13	8	54	13
Slovakia (45)	31	33	13	22	0
Israel (62)	27	34	15	10	15

Note: Figures are rounded. Regular or experienced use combines the two highest-use response categories.

The industry interviews point in the same direction. Employers described active or emerging XR use in practical settings, including VR laboratories for training, AR overlays for maintenance, virtual walkthroughs, retail visualisation and heritage interpretation. Their skills requirements were not confined to a single software package. Firms sought people able to develop in engines such as Unity or Unreal, create and animate 3D assets, design spatial interactions, and integrate an XR prototype into a classroom, service or production process. They also stressed client communication, teamwork and project experience. Cost, specialist recruitment, legacy systems, infrastructure, privacy and resistance to change were recurring obstacles, especially for smaller firms and public institutions.

Implications for XR education

Taken together, the two datasets suggest that the challenge is not only to raise awareness of XR. Students need repeated, guided opportunities to use the technology, while employers need graduates who can connect technical development with a real organisational problem. The findings informed the Academy's subsequent 19-course offer. The practical implications were to introduce core concepts through first supervised use, place hardware in learners' hands early, combine engines, 3D production and spatial UX with ethics and data protection, and use project briefs linked to external organisations. Flexible and hybrid formats are also important where lab access, connectivity or equipment availability is limited. The results support a shift from occasional demonstrations towards structured practice, industry-linked projects and routes that work across different national infrastructures.

Conclusion

This ten-country study finds a persistent mismatch: students generally see value in XR but have little opportunity to practise it in their degree programmes, while companies report a need for technically capable people who can apply XR in context. The results do not establish prevalence for Europe as a whole, and the self-reported measures need future replication with observed use. They do, however, provide a grounded basis for curriculum and capacity-building decisions. For the EuroXR community, a useful measure of progress is not awareness alone, but the amount of supervised use learners receive and their ability to complete an XR brief that matters outside the lab.

Acknowledgements

This work was co-funded by the European Union through Erasmus+ Alliance for Innovation, Grant Agreement 101140232. The authors thank the partner organisations, students and industry interviewees who contributed to the study. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or EACEA.

References

Al-Adwan, A. S., Li, N., Al-Adwan, A., Abbasi, G. A., Albelbisi, N. & Habibi, A. (2023). Extending the technology acceptance model to predict university students' intentions to use metaverse-based learning platforms. *Education and Information Technologies* 28(11), 15381–15413.

Creswell, J. W. & Plano Clark, V. L. (2018). Designing and Conducting Mixed Methods Research. 3rd ed. *Thousand Oaks*, CA: Sage.

Johnson, R. B. & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher* 33(7), 14–26. https://doi.org/10.3102/0013189X033007014

López-Belmonte, J., Pozo-Sánchez, S., Lampropoulos, G. & Moreno-Guerrero, A. J. (2022). Design and validation of a questionnaire for the evaluation of educational experiences in the metaverse in Spanish students. *Heliyon* 8(11), e11461.

Metaverse Academy. (2024). D2.3 Market Research Report: Need Analysis of the Industrial Stakeholders and D2.4 Skills Analysis Report and Practical Guidelines: Need Analysis of Students. *Castelló de la Plana*: Universitat Jaume I.

Feel the Hue: Exploring Colour-Vibrotactile Correspondences in VR for Affective Regulation

Noelia González-Sánchez[1], Arturo S. García[1,2], Antonio Fernández-Caballero[1,2,3], and José P. Molina[1,2]

[1]*Instituto de Investigación en Informática de Albacete, Albacete, Spain*

[2]*Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Albacete, Spain*

[3]*CIBERSAM (Biomedical Research Networking Centre in Mental Health), Madrid, Spain*

Corresponding author: noelia.gonzalez@uclm.es

Keywords: Crossmodal correspondences, affective regulation, virtual reality, vibrotactile feedback.

Abstract.

Crossmodal correspondences between colour and haptic stimulation have been documented in general contexts, but there is limited evidence on how they operate on the glabrous skin of the hand during affective touch stimulation. This work explores, through a preliminary study in virtual reality, how congruence between hue and vibrotactile intensity influences affective response. Eight participants rated the valence and arousal of vibrotactile stimuli delivered through haptic gloves when touching virtual cubes of six hues, manipulating intensity (20% and 100%) and duration (0.5, 2 and 5 s) under conditions congruent and incongruent with the hue-heat hypothesis. Vibration intensity was the dominant factor in both dimensions, distributing the stimuli into two clusters of the affective space, whereas congruence produced an additional effect on valence but not on arousal. Low-intensity stimuli combined with cool hues were most strongly associated with states of calm, making them candidates for the design of haptic interfaces aimed at affective regulation.

Introduction

Crossmodal correspondences are defined as the tendency to associate sensory features across two or more modalities (Spence, 2011) and may exist even when the stimuli are imagined. Among them, the Hue-Heat Hypothesis holds that colours are implicitly and consistently associated with specific thermal sensations (Plass et al., 2014), so that the brain uses visual temperature cues to modulate the response to direct physical stimulation (Ho et al., 2014). Slobodenyuk et al., 2015 have documented mappings between haptic stimulation and colour attributes, supported by three mechanisms: structural, semantic and statistical (Delazio et al., 2017). There is, however, limited evidence on how these principles operate on the glabrous skin of the hand during affective touch stimulation. Despite lacking C-tactile (CT) afferent fibres, glabrous skin can display a hedonic response to stimuli (Gonzalez-Sanchez et al., 2026), so that a visual stimulus congruent with the vibration could act as a psychophysiological regulator. Which specific combinations of vibration and colour shape affective response remains an open question.

The present study explores how crossmodal correspondence and the hue-heat hypothesis influence the affective response to vibrotactile touch, using haptic interfaces with actuator arrays in a controlled virtual reality environment. When vibrotactile stimulation is accompanied by congruent visual information, the central nervous system integrates both modalities coherently, enhancing hedonic processing and potentially relieving anxiety.

Experiment

The sample comprised 8 participants (5 men, 3 women; mean age 33), all right-handed, healthy and neurotypical, with no history of neurological disorders. A within-subject design was used, exposing all participants to the full set of experimental conditions.

The immersive experience was developed in Unity and displayed through a Meta Quest 2 headset. Vibrotactile stimulation was delivered using bHaptics TactGlove DK2 gloves (bHaptics Inc., 2026), which provide feedback through six linear resonant actuators (LRA) per glove (five on the fingertips and one on the wrist at a nominal resonance frequency of approximately 170 Hz. Stimuli were modelled as abstract geometric shapes (cubes) to prevent bias from prior cognitive associations (Slobodenyuk et al., 2015). Colour was encoded according to the Munsell Color System to ensure replicability, defining six hues divided equally by perceived chromatic temperature (Table 1). Backgrounds and ambient surfaces were set to neutral grey (B0B0B0) to prevent simultaneous contrast or chromatic induction effects.

***Table 11**. Chosen colours and their representation in the Munsell and hexadecimal systems*

Hue	Perceived temperature	Value in the Munsell Colour System	Hexadecimal colour value
Red	Warm	7.5R 5/22	F31010
Orange	Warm	10R 6/16	FC6015
Yellow	Warm	7.5Y 9/12	FDE80C
Light blue	Cool	10BG 9/8	50FAFC
Blue	Cool	5PB 6/14	1997F8
Purple	Cool	10PB 4/26	7B1AF1

Experimental design

The experiment manipulated three main factors in a within-subject factorial design:

1. Hue: 6 levels (red, orange, yellow, light blue, blue, purple).
2. Stimulus intensity: 2 levels, corresponding to 100% and 20% of the maximum output power of the bHaptics TactGlove DK2 vibrotactile actuators.
3. Stimulus duration: 3 levels of active stimulation (5, 2 and 0.5 seconds).

These variables were distributed across 6 independent blocks structured according to two sensory congruence hypotheses. In the congruent hypothesis (3 blocks), warm hues were paired with 100% intensity and cool hues with 20%; in the incongruent hypothesis (3 blocks) this assignment was reversed. Each block consisted of six vibrotactile trials, one per hue, presented in a fixed order, so that participants experienced all combinations of intensity and duration defined for each congruence condition, with 18 stimuli per hypothesis. The primary dependent variables were participants' subjective affective responses, assessed along two dimensions: hedonic (pleasant–unpleasant) and arousal (calming–activating). As a secondary control variable, the presence of residual tingling was recorded after stimulation.

Experimental procedure

The protocol was structured to avoid biases from fatigue or physiological adaptation. Participants first completed a demographic questionnaire (age, sex, handedness) and the state subscale of the STAI (Spielberger, 2010) as an anxiety baseline and were then fitted with the headset and the glove on their dominant hand. In each block, they touched the virtual cube and received the corresponding stimulation, in a fixed order (red, orange, yellow, light blue, blue, purple), reporting valence and arousal verbally (5-point Likert) after each one, with a 20-second rest between colours. Open questions on thermal sensations and colour–vibration associations closed each block. The congruent condition was administered first, followed by a 1-minute rest and a check for numbness, and then the incongruent condition in the same sequence. Two final questions addressed perceived realism and associations with natural elements.

Results

Global mean valence was M = 3.42 and mean arousal M = 2.79. Intensity was the dominant factor in both dimensions: stimuli at 20% obtained higher valence (M = 4.04; SD = 0.95) and lower arousal (M = 1.90; SD = 0.97) than those at 100% (M = 2.80, SD = 1.39; M = 3.68, SD = 1.18). Congruence showed an additional effect on valence (congruent M = 3.65; incongruent M = 3.19), but barely on arousal (M = 2.67; M = 2.91). Duration produced no appreciable differences in either dimension (range of means below 0.3 points). As Figure 1 shows, stimuli were distributed into two clusters defined by intensity: those at 20% in the high-valence, low-arousal region, and those at 100% in the low-valence, high-arousal region.

Discussion

The results show that vibration intensity was the factor that most determined differences in valence and arousal, although congruence between colour and vibration produced an additional effect in the expected direction. At equal physical intensity, the congruent condition obtained higher valence scores than the incongruent one, a pattern consistent with structural and semantic correspondence processed by the central nervous system (Slobodenyuk et al., 2015). This effect was reinforced by participants' comments: in incongruent blocks they described a decrease in perceived realism, noting that the vibration felt inverted with respect to what they expected for each colour. Given that hedonic assignment on glabrous skin relies on cortical evaluation (Gonzalez-Sanchez et al., 2026), an incongruent stimulus would demand greater cognitive effort, reducing positive hedonic evaluation (Laura et al., 2024). In the final questions, participants associated colours with specific natural elements (red and orange with fire or lava; light blue with water or ice) and reported subjective sensations of temperature change. These results point in the same direction as (Ziat et al., 2016), which explains that, under strict spatial and temporal synchrony between physical action and visual feedback, the brain integrates the colour signal to bias somatosensory interpretation. Colour would therefore not be a passive aesthetic attribute, but a signal the nervous system uses to modulate tactile afferents. When valence and arousal are integrated jointly, however, stimuli organised into two clusters defined primarily by intensity level Figure 1. It was therefore physical intensity that determined the region of the affective space towards which each stimulus shifted, following Russell's circumplex model (Russell, 1980), while congruence acted as a modulator within each level: congruence affected valence but barely arousal, suggesting that crossmodal correspondence operates on perceived pleasantness rather than on arousal level, which is governed largely by physical intensity. The only exception to this pattern was blue in incongruent condition, positioned at the boundary between both regions (M = 3.50 in both dimensions). Low-intensity stimuli combined with a congruent (cool) colour were those most located in the calm region.

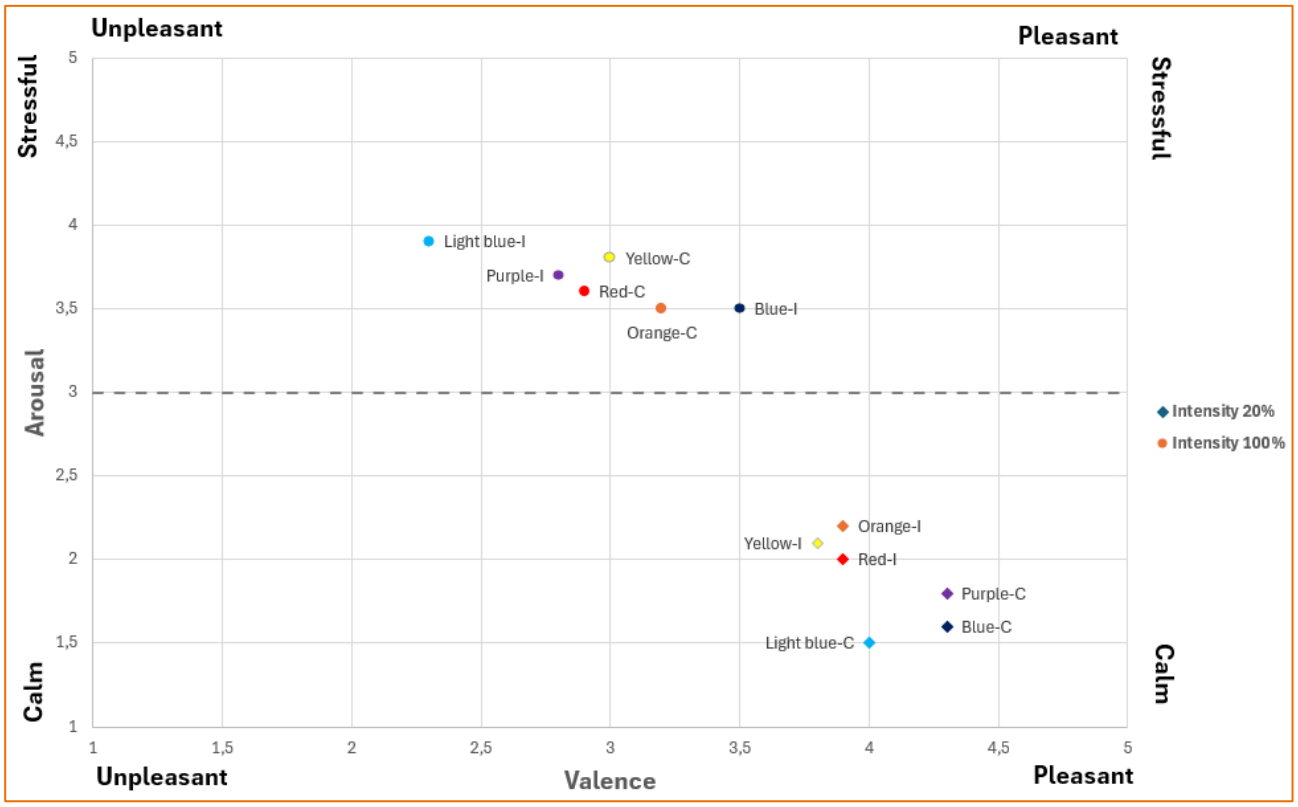


***Figure 100**. Distribution of stimuli in the valence-arousal space. Each point represents a colour in a congruence condition (C = congruent, I = incongruent). The dotted line marks the midpoint of the arousal scale.*

Conclusions

This study evaluated the effect of congruence between colour and vibration intensity on perceived valence and arousal, with the aim of identifying which combinations might be effective in inducing states of calm. Vibration intensity was the factor that most affected both dimensions across the six colours; nonetheless, the correspondence between colour and vibration temperature was explicitly recognised by participants when experiencing the stimuli. When valence and arousal were integrated into a joint affective space, low-intensity stimuli combined with cool (congruent) colours were most strongly associated with states of calm; within this group blue and light blue stood out, followed by purple, while warm colours showed higher arousal levels even at low intensity. These findings identify the combination of low intensity with cool colours as the most promising, among those evaluated, for the design of vibrotactile stimuli oriented towards affective regulation, although the choice of duration appears to allow some flexibility. This work should be understood as a preliminary, exploratory study. The sample was small (n = 8), which prevents generalisation of the findings, and the analysis was descriptive in nature, so it cannot establish whether the observed differences are statistically significant. Nevertheless, the findings provide initial evidence on how vibrotactile intensity and its congruence with colour modulate affective response, a principle transferable to the design of virtual reality environments and haptic devices that favour multisensory feedback.

Acknowledgments

This work was partially supported by grant PID2023-149753OB-C21 funded by MICIU/AEI/10.13039/501100011033, grant 2025-GRIN-38441 funded by Universidad de Castilla-La Mancha, CYTED through Red 225RT0169, and CIBERSAM-ISCIII, all co-funded by "ERDF A way to make Europe

References

bHaptics Inc. (2026). *Manual TactGlove DK2*.

Delazio, A., Israr, A., & Klatzky, R. L. (2017). Cross-modal correspondence between vibrations and colors. *2017 IEEE World Haptics Conference (WHC)*, 219–224. https://doi.org/10.1109/WHC.2017.7989904

Gonzalez-Sanchez, N., García S., A., Fernández-Caballero, A., & Molina P., J. (2026). Affective touch: Estudio preliminar de estímulos vibrotáctiles en interfaces hápticas. *Congress Interacción*. https://doi.org/10.65234/interaccion.153

Ho, H.-N., Van Doorn, G. H., Kawabe, T., Watanabe, J., & Spence, C. (2014). Colour-Temperature Correspondences: When Reactions to Thermal Stimuli Are Influenced by Colour. *PLoS ONE*, *9*(3), e91854. https://doi.org/10.1371/journal.pone.0091854

Laura, B., Claudio, Z. C., Massimiliano, Z., & Riccardo, P. (2024). Investigating the validity of the hue-heat effect on thermal sensitivity. *Scientific Reports*, *14*(1), 21413. https://doi.org/10.1038/s41598-024-71784-7

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, *29*, 128–140. https://doi.org/10.1016/j.learninstruc.2013.02.006

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Slobodenyuk, N., Jraissati, Y., Kanso, A., Ghanem, L., & Elhajj, I. (2015). Cross-modal associations between color and haptics. *Attention, Perception, & Psychophysics*, *77*, 1379–1395.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spielberger, C. D. (2010). State-Trait Anxiety Inventory. In *The Corsini Encyclopedia of Psychology* (p. 1). Wiley. https://doi.org/10.1002/9780470479216.corpsy0943

Ziat, M., Balcer, C. A., Shirtz, A., & Rolison, T. (2016). *A Century Later, the Hue-Heat Hypothesis: Does Color Truly Affect Temperature Perception?* (pp. 273–280). https://doi.org/10.1007/978-3-319-42321-0_25

Supporting Low Vision in VR: Accessibility Challenges and Design Principles

Brigida Bonino[1], Franca Giannini[1], Katia Lupinetti[1], Silvia Dini[2], and Chiara Terrile[2]

[1]*Istituto di Matematica Applicata e Tecnologie Informatiche – CNR*

[2] *Fondazione David Chiossone*

Corresponding author: brigida.bonino@cnr.it

Keywords: Virtual reality; Low vision; Accessibility; Virtual museums; Design principles

Introduction

Virtual Reality (VR) is increasingly used to enhance traditional approaches in education, healthcare, training, and cultural heritage. However, its potential for accessibility and inclusion remains underexplored, particularly across diverse disabilities. This work focuses on people with low vision, a significant reduction in visual function that cannot be fully corrected through conventional lenses or surgery, but allows the use of residual sight. Unlike solutions for blind users, accessible approaches for this group should not replace vision solely with audio or tactile feedbacks, but enhance remaining visual capabilities and support autonomous interaction. VR can contribute to this goal by reducing physical barriers and enabling visual simplification, contrast adjustment, and multimodal content access. However, most immersive applications are designed for sighted users and rely heavily on visual interfaces and spatial awareness (Zhao et al., 2019; Creed et al., 2024). Research on VR and visual impairment mainly addresses simulation, rehabilitation, and training rather than the use of VR as an assistive technology. Consequently, relatively few studies have investigated how immersive experiences can be made accessible to people with low vision (Dudley et al., 2023). This gap also extends to cultural activities, such as museum visits, where barriers include fixed viewing distances, reflective display cases, inaccessible labels, and visual crowding (Salameh et al., 2026). Accessible virtual museums must therefore avoid reproducing existing barriers or introducing new ones. Accordingly, this work identifies challenges encountered in physical and virtual museums and, more broadly, in environments requiring users to perceive, distinguish, and explore objects. Through an iterative expert-driven co-design process involving VR researchers, orthoptists, and assistive technology specialists, these challenges are translated into design principles aimed at supporting residual vision, autonomous exploration, and accessible interaction.

Identification of the Challenges

The accessibility challenges for low vision in immersive environments are identified by combining findings from previous studies on virtual museum and VR accessibility (Zhao et al., 2019), relevant guidance from the WCAG 2.2 (https://www.w3.org/TR/WCAG22) and the XAUR (https://www.w3.org/TR/xaur), and the clinical and technical knowledge of the experts from the Fondazione David Chiossone in Genoa, who are involved in the co-design process. Challenges are organized into two complementary categories: museum-related challenges (MRC), concerning the

barriers experienced in physical museums; and technology-related challenges (TRC), concerning the potential accessibility issues arising from interaction with immersive VR systems. Together, the challenges listed in the following provide the basis for the definition of VR accessibility requirements and design principles.

- **Navigation and spatial orientation (MRC):** unfamiliar and crowded exhibition spaces can make it difficult for visitors with low vision to identify paths, locate artifacts, and move independently, often increasing their reliance on accompanying persons or museum staff.
- **Limited artifact visibility and exploration (MRC):** fixed viewing distances, reflective display cases, glare, shadows, and inadequate lighting may prevent visitors from perceiving relevant visual details. Moreover, valuable or fragile objects generally cannot be approached closely, touched, or observed from alternative viewpoints, limiting active and detailed exploration.
- **Visual crowding and inaccessible information (MRC):** the simultaneous presence of artifacts, visitors, decorative elements, supports, and information panels can make relevant content difficult to isolate. Labels may also be too small, insufficiently contrasted, overly detailed, or positioned at inaccessible distances.
- **Limited support for the residual vision (MRC):** museum accessibility solutions often rely mainly on audio descriptions and, when available, tactile reproductions, providing alternatives to vision rather than supporting visitors who wish to use their remaining visual capabilities.
- **Visual complexity and inaccessible elements (TRC):** realistic textures, decorative elements, competing light sources, insufficient contrast, small icons, selection cues, and subtle visual feedback may increase visual crowding, making objects, interfaces, and information difficult to distinguish and locate. Users may also struggle to estimate distances and confirm whether a target has been correctly selected.
- **Complex interaction, disorientation, and cybersickness (TRC):** controller buttons, menus, and command sequences may increase cognitive load and require users to learn unfamiliar interaction mechanisms. Artificial locomotion, sudden viewpoint changes, and unclear spatial references may also cause disorientation, discomfort, or cybersickness**.**
- **Limited and unimodal object exploration (TRC):** virtual museums may reproduce the restrictions of physical exhibitions if users cannot approach, enlarge, rotate, or manipulate artifacts. Moreover, information provided through a single sensory channel may not accommodate different perceptual and functional capabilities.
- **Lack of personalization (TRC):** fixed settings for contrast, colors, scale, lighting, information presentation, and interaction modalities may not address the heterogeneity of low vision conditions and individual preferences.

From Challenges to Requirements and VR Design Principles

The identified challenges are translated into accessibility and interaction requirements for people with low vision, also considering relevant WCAG 2.2 guidelines (i.e., perceivable, adaptable, distinguishable, navigable, and input modalities) and the XAUR user needs. The related requirements are consolidated into the five design principles described below.

Table 1 summarizes the relationships among challenges, requirements, and design principles.

Table 1. *Relationships among challenges, requirements, and VR design principles for low vision.*

Challenge	Requirement	Design Principle
Insufficient contrast, poorly defined contours, and difficulty in distinguishing artifacts and interface elements.	Make relevant environmental and interactive elements distinguishable across residual visual capabilities	High-Contrast and Distinguishable Coloring and Rendering
Visual crowding, environmental complexity, competing light sources, and excessive information.	Reduce perceptual and cognitive load while preserving the information necessary to understand and explore.	Environmental Simplification and Visual Hierarchy
Poorly visible pointers, ambiguous selection feedback, and difficulty locating interaction targets.	Provide clearly perceivable interaction cues and immediate, multimodal confirmation of user actions.	Salient Interaction Cues and Multimodal Feedback
Fixed viewing distances, restricted artifact exploration, and complex interaction techniques.	Enable users to manipulate and inspect artifacts through accessible and intuitive interaction modalities.	Accessible Objects Exploration and Natural Interaction
Heterogeneity of visual conditions, preferences, and interaction capabilities.	Make visual rendering and interaction modalities tailored to user needs.	Personalization and Adaptability

i. **High-Contrast and Distinguishable Coloring and Rendering.** Virtual objects, interfaces, and interaction cues should be clearly distinguishable from the background through strong

ii. luminance and chromatic contrast, well-defined edges, and sufficiently large visual elements. Since low vision includes heterogeneous perceptual conditions, users should also be able to customize colors, brightness, contrast, outlines, and visualization modes according to their residual visual capabilities.

iii. **Environmental Simplification and Visual Hierarchy.** Highly detailed environments may increase visual crowding and make relevant information difficult to isolate. VR museums should therefore reduce decorative elements, complex textures, reflections, competing light sources, and unnecessary objects. Scenes should contain a limited number of clearly recognizable points of interest, using visual hierarchy to distinguish artifacts and interactive elements from secondary environmental content.

iv. **Salient Interaction Cues and Multimodal Feedback.** Pointers, selection rays, and target indicators should be sufficiently large and strongly contrasted with the environment. User actions should generate immediate and unambiguous feedback by combining visual highlighting with audio messages or vibration. This multimodal approach supports target localization and selection while reducing dependence on a single sensory channel.

v. **Accessible Object Exploration and Natural Interaction.** Unlike physical museums, VR can allow users to approach, enlarge, rotate, and inspect artifacts from different viewpoints. These actions should be supported through natural interaction modalities, such as direct hand manipulation and voice commands, without requiring complex sequences or controller combinations. Audio descriptions and alternative visualizations should complement object exploration whenever details remain difficult to perceive.

vi. **Personalization and Adaptability.** No single configuration can accommodate the heterogeneity of low vision conditions. VR museum experiences should therefore allow users to adapt visual parameters, information presentation, and interaction modalities to their needs and preferences. Personalization should be considered a core accessibility principle rather than an additional feature, enabling users to exploit their residual vision more effectively.



***Figure 1**. Virtual museum environments for normal vision and adapted for low vision.*

An example of the proposed design principles application is shown in Figure 1. A virtual museum environment originally developed for users with normal vision (Bonino et al., 2025) and a version adapted for low vision are compared. The new environment reduces visual complexity and clutter by simplifying wall and floor textures and removing non-essential elements, such as supports and information panels. Stronger contrast distinguishes points of interest and interaction elements, such as the user's virtual hands, while highlighted contours and high contrast color changes facilitate object identification and selection. Multimodal interaction is supported through natural bare-hand gestures for object manipulation and speech-recognition-based voice commands for exploration and navigation. A virtual pointer enables objects selection and triggers audio feedback for accessing information.

Conclusions

This work presents design principles for developing immersive VR environments accessible to people with low vision. Drawing on challenges encountered in physical museums and immersive systems, these principles translate perceptual and functional needs into practical recommendations for environmental simplification and distinguishability, interaction feedback, object exploration, and personalization. Although developed for museums, they also inform general virtual environments involving object perception, understanding, and interaction. Future work will evaluate the proposed principles with end users with different low vision conditions and refining them according to their experiences and preferences. Further developments will also investigate more adaptive and personalized configurations and assess the applicability of the principles across different immersive scenarios.

References

Bonino, B., Giannini, F., & Lupinetti, K. (2025). Inclusive and Engaging Virtual Museum Experience for Children. In *Proc. CSEDU 2025*, 1, 907–914.

Creed, C., Al-Kalbani, M., Theil, A., Sarcar, S., & Williams, I. (2024). Inclusive AR/VR: Accessibility barriers for immersive technologies. *Universal Access in the Information Society*, 23, 59–73.

Dudley, J., Yin, L., Garaj, V., Kristensson, P.O., (2023). Inclusive immersion: a review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse. *Virtual Reality* 27, 2989–3020.

Salameh, R., Lanir, J., & Szpiro, S.F. (2026). Remote Assistance with Augmented Reality: A Flexible Accessibility Solution for Low Vision Museum Visitors. *International Journal of Human–Computer Interaction* 42, 2988–3011. https://doi.org/10.1080/10447318.2025.2531268

Zhao, Y., Cutrell, E., Holz, C., Morris, M. R., Ofek, E., & Wilson, A. D. (2019). SeeingVR: A set of tools to make virtual reality more accessible to people with low vision. In *Proc. CHI 2019*, 1–14. https://doi.org/10.1145/3290605.3300341

Material Degradation and Structural Behavior of Cultural Heritage Burial Stone Column through XR

Ioannis Filippos Kyriakidis [1,2], Emmanouil K. Tzimtzimis [1], Thomas Panagiotopoulos [1,2], Kyriacos Themistokleous [3], Konstantinos Tsongas [2], and Dimitrios Tzetzis [1],

[1]Digital Manufacturing and Materials Characterization Laboratory, School of Science and Technology, International Hellenic University, 57001 Thessaloniki, Greece;

[2]Advanced Materials and Manufacturing Technologies Laboratory (AMMT) Lab, International Hellenic University, Thessaloniki, 57001

[3]ERATOSTHENES Centre of Excellence, Franklin Roosevelt 82, 3012 Limassol, Cyprus

Corresponding author: d.tzetzis@ihu.edu.gr

Keywords: Cultural Heritage, 3D Scanning, Representative Volume Element, Porosity, Finite Element Analysis, Extended Reality

Abstract

Cultural heritage monuments provide an important physical connection between contemporary society and the architectural, technological, and cultural achievements of the past (Turchetti et al., 2026). Nevertheless, prolonged exposure to environmental conditions, weathering processes, material ageing, and accumulated structural damage can progressively alter the mechanical properties of historic structures. The geometrical complexity and material heterogeneity of such monuments also make their structural assessment particularly challenging. Conventional investigation methods require direct contact with the monument, while destructive material characterization is generally restricted because of the cultural and historical significance of the structure. Recent developments in three-dimensional scanning, computational mechanics, micromechanical modelling, and Extended Reality (XR) offer new opportunities for the non-destructive documentation, numerical assessment, and interactive visualization of cultural heritage assets. This work presents an integrated digital methodology for investigating the compressive structural response of a historic stone column located at the Tombs of the Kings archaeological site in Paphos, Cyprus (Figure 1).

The proposed approach combines reality-based geometrical documentation with multiscale material modelling and finite element analysis. The geometry of the column was initially captured using three-dimensional scanning techniques, enabling its external morphology and existing geometrical irregularities to be digitally documented (Alfio et al., 2022; Gonizzi Barsanti et al., 2022). The acquired scanning data were processed and converted into a triangulated STL model representing the geometry of the original column. Unlike an idealized computer-aided design model, the resulting digital reconstruction retains the characteristic surface features and geometrical variations of the actual monument. The STL geometry was subsequently processed and imported into a finite element environment, where it provided the geometrical basis for the numerical

compression analysis. In this way, the methodology establishes a direct connection between the physical cultural heritage asset and its computational representation.

To account for the non-uniform condition of the historic material, the digital column was divided into five distinct zones. Each zone was assigned a prescribed porosity level, allowing spatial variations in the mechanical condition of the stone to be incorporated into the numerical model. This zonal representation was adopted because deterioration and pore development may not occur uniformly throughout a historic structural element. Different parts of a column can experience different levels of weathering depending on their position, environmental exposure, moisture accumulation, loading history, and proximity to the ground. Dividing the structure into separate material regions therefore enables the influence of spatially varying porosity to be investigated without assuming that the entire column possesses identical mechanical properties. The five zones form an intermediate level between the detailed pore-scale description of the material and the full-scale structural model of the column.

The mechanical properties associated with the different porosity levels were determined using a Representative Volume Element (RVE) micromechanical approach (Kalo et al., 2017; Miled et al., 2011). For each porosity condition, an RVE was employed to represent the internal material structure at the microscale. The micromechanical models were used to estimate the corresponding homogenized effective properties, describing the equivalent mechanical behaviour of the porous stone as a continuous material. These effective properties were subsequently transferred to the macroscale finite element model and assigned to the respective column zones. Consequently, individual pores did not need to be explicitly reproduced throughout the full geometry of the column, which would have resulted in a highly complex model and considerable computational requirements. The adopted multiscale approach instead provides an efficient connection between microscale porosity and macroscale structural behaviour, while retaining the influence of material heterogeneity within the numerical assessment.

Following the assignment of the porosity-dependent effective properties, a static structural compression simulation was conducted in finite element analysis software. Appropriate loading and boundary conditions were applied to reproduce the axial compression of the column, and the resulting distributions of stress, strain, and displacement were evaluated. The numerical analysis was used to investigate how the five material zones contribute to the overall mechanical response of the structure and to identify regions potentially exhibiting increased deformation or stress concentration. Particular attention was given to the interaction between the reality-based geometry and the spatial variation of the material properties. The methodology therefore extends beyond a conventional homogeneous analysis, in which a single set of mechanical properties would be applied to the entire structural component. By introducing zone-specific properties obtained through the RVE models, the simulation provides a more representative description of the potential influence of porosity on the compressive response of the historic column.

The numerical results were subsequently transferred into an XR environment to support their interactive exploration and communication (Guo et al., 2024). Through XR visualization, users can observe the reconstructed column in three-dimensional space, inspect the five material zones, and examine the calculated structural response directly on the digital geometry. Stress, strain, or displacement distributions can be superimposed on the virtual representation, allowing areas of interest to be identified more intuitively than through conventional two-dimensional contour images alone. The XR environment can therefore act as an interface between advanced engineering analysis and cultural heritage interpretation. It can assist researchers and conservation specialists in examining numerical results while also making the assessment more accessible to students, museum visitors, and other non-specialist audiences. The integration of multiscale modelling, finite element analysis, and immersive visualization establishes a comprehensive workflow connecting physical documentation, material characterization, structural evaluation, and interactive communication.

The presented poster demonstrates the potential of combining three-dimensional scanning, RVE-based homogenization, static structural analysis, and XR technologies for the digital

assessment of cultural heritage structures. The developed workflow provides a non-destructive framework for representing geometrical complexity and spatial material heterogeneity within a common computational model.

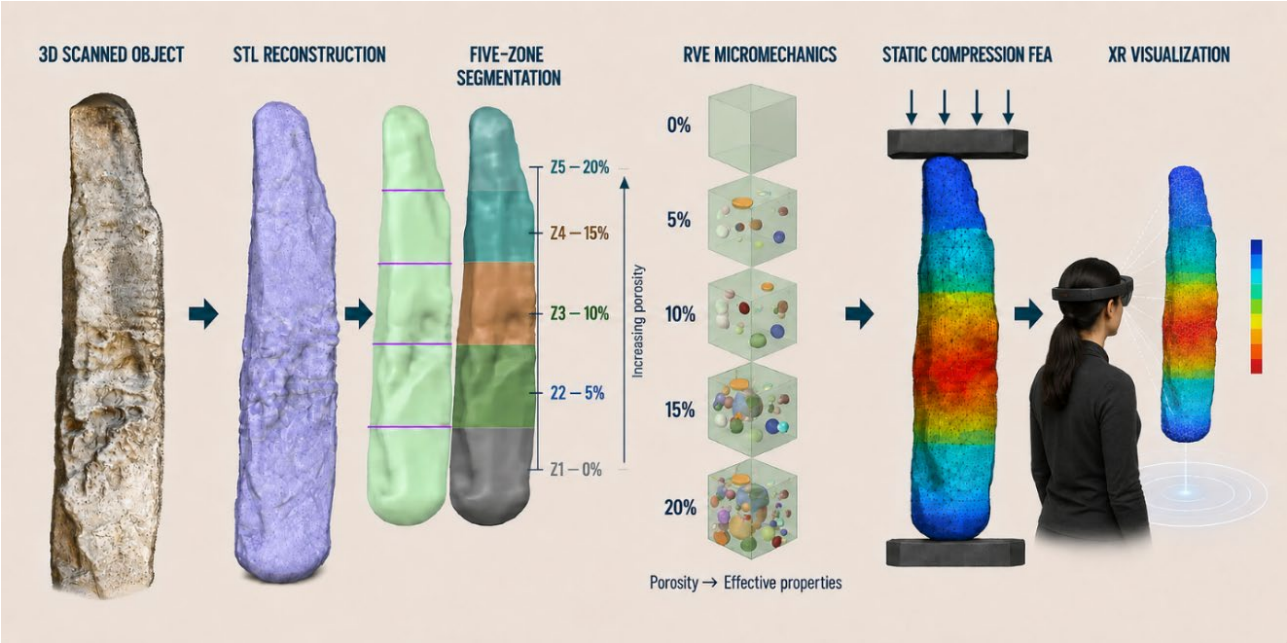


***Figure 1**. Conceptual overview of the proposed multiscale methodology for the structural assessment and immersive visualization of the historic column with the aid of 3D-Scanning, RVE Modelling, Finite Element Analysis, and Extended Reality Visualization.*

This research is part of the EXCALIBUR project that is funded by the European Commission under the Horizon Europe Framework Program and the grant agreement N◦ 101233712. Views and opinions expressed are however, those of the authors only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Alfio, V. S., Costantino, D., Pepe, M., & Restuccia Garofalo, A. (2022). A geomatics approach in scan-to-FEM process applied to cultural heritage structures: The case study of the “Colossus of Barletta”. *Remote Sensing*, 14(3), 664. https://doi.org/10.3390/rs14030664

Gonizzi Barsanti, S., Guagliano, M., & Rossi, A. (2022). 3D reality-based survey and retopology for structural analysis of cultural heritage. *Sensors*, 22(24), 9593. https://doi.org/10.3390/s22249593

Guo, Y., Liu, L., Huang, W., Shen, M., Yi, X., Zhang, J., & Lu, S. (2024). Extending X-reality technologies to digital twin in cultural heritage risk management: A comparative evaluation from the perspective of situation awareness. *Heritage Science*, 12, 245. https://doi.org/10.1186/s40494-024-01363-6

Kalo, K., Grgic, D., Auvray, C., Giraud, A., Drach, B., & Sevostianov, I. (2017). Effective elastic moduli of a heterogeneous oolitic rock containing 3-D irregularly shaped pores. International *Journal of Rock Mechanics and Mining Sciences*, 98, 20–32. https://doi.org/10.1016/j.ijrmms.2017.07.009

Miled, K., Sab, K., & Le Roy, R. (2011). Effective elastic properties of porous materials: Homogenization schemes versus experimental data. *Mechanics Research Communications*, 38(2), 131–135. https://doi.org/10.1016/j.mechrescom.2011.01.009

Turchetti, F., Oreni, D., Votsis, R., Kyriakides, N., Cuca, B., & Agapiou, A. (2026). The analysis of a column of the Tomb 7 colonnade at the Tombs of the Kings archaeological site: A comparative evaluation of scan-to-FEM methodologies. *Heritage*, 9(3), 100. https://doi.org/10.3390/heritage9030100

.

Voice-Driven Generation of 3D Assets and Immersive Environments in Mixed Reality

Josep Carreres-Fluixa, Jose Luis Soler-Dominguez, Samuel Navas-Medrano, and Patricia Pons

HCI research group, Instituto Tecnológico de Informática, Valencia, Spain

Corresponding author: jcarreres@iti.es

Keywords: Mixed Reality; Generative AI; 3D Asset Generation; Voice Interaction; Immersive Authoring.

Introduction

Traditional XR asset creation workflows typically occur at edit time, relying heavily on flat-screen modeling tools or complex programmatic pipelines, resulting in a steep learning curve for general XR creators. Text-to-3D spatial generative AI pipelines offer an alternative (Poole et al., 2022; Zhu et al., 2026), enabling users to materialize geometric assets through simple natural language prompts. However, most generative workflows remain separated from the immersive experience itself, requiring users to formulate and refine generated content outside the XR experience. Bringing real-time generative capabilities directly into lightweight, standalone Head-Mounted Displays (HMDs) introduces both strict computational bottlenecks—due to mobile GPU limits—and critical interaction design challenges. Providing intuitive authoring mechanisms is essential, particularly for users with cognitive or mental health conditions, where cognitive overload, sensory saturation, and fine-motor interface fatigue severely hinder engagement.

To overcome these barriers, we present an interactive tool, integrated in a Unity Mixed Reality app for Meta Quest 3, that leverages a distributed, decoupled server architecture to offload high-latency spatial AI inference. Grounded in iterative co-design sessions involving XR researchers, clinical specialists, and neurodivergent end users, our interaction design substitutes nested UI panels and virtual keyboards with an embodied tabletop workspace and tangible-like Toggle-to-Talk buttons. By articulating simple natural language prompts users can personalize mindful relaxation spaces in real time dynamically generating custom 3D assets and transforming the surrounding environment through 360-degree panoramic skyboxes viewed directly through virtual windows. This approach preserves immersion and eliminates the need for intermediate 2D evaluation steps.

Related Work

Recent work in immersive authoring environments has explored diverse modalities for human-AI interaction in XR. A growing body of research investigates the potential of generative AI to craft deeply contextual, emotionally resonant environments; for instance, (Wen et al., 2026) demonstrated the therapeutic value of generating personalized emotional sanctuaries in fully immersive VR However, existing frameworks often suffer from severe bottlenecks, imposing cognitive and semantic barriers that constrain the XR experience.

On the one hand, conversational systems like CoGen3D (Jiang et al., 2026) fragment spatial flow by forcing users to evaluate intermediate 2D images on flat displays prior to 3D instantiation.

While staged confirmation helps filter ambiguity, navigating flat 2D evaluation panels or virtual dialogue interfaces increases cognitive load and attentional shifts, which can be particularly overwhelming for vulnerable users. Similarly, in-situ authoring environments like ImaginatAR (Lee et al., 2025) and RealityCrafter (Kim et al., 2025) support situated scene editing, but still rely on multi-step menus or 2D input modalities that disrupt presence. On the other hand, modular approaches such as PartInteractor (Jiang & Li, 2026) restrict creative freedom by relying on rigid, predefined semantic hierarchies for part-based editing, which additionally introduces computational overhead in volumetric analysis.

In contrast to existing systems that rely on floating 2D UI panels or complex multi-stage dialogues, our work eliminates intermediate flat evaluations through direct natural language voice-to-3D-object or Skybox generation. This enables the interactive materialization and manipulation of any asset within an immersive environment, guided solely by user intent.

Distributed XR System Architecture

Our system architecture, as seen in Figure 1, relies on a clean decoupling between the XR client and a distributed backend of specialized inference workers. This separation keeps computationally intensive generation transparent to the user and preserves the responsiveness of the MR experience. Rather than introducing monolithic cloud-native architectures, our design utilizes a lightweight event- and message-driven loop, managed through a FastAPI gateway, a Redis message broker, and the workers responsible for executing the inferences.

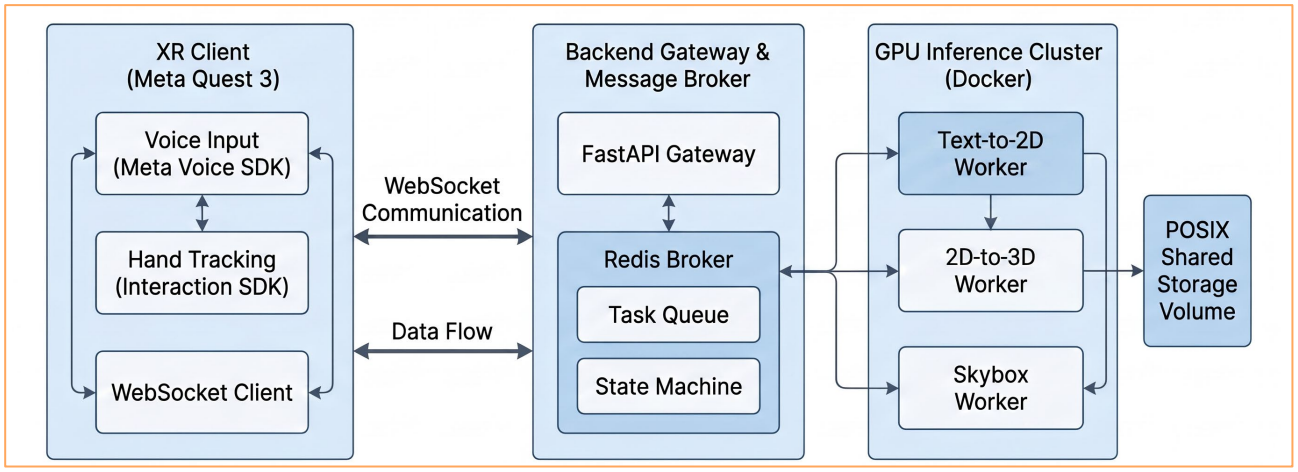


***Figure 101**. Distributed System Architecture Diagram*

Distributed Deployment Model and Orchestration Engine

The core backend is structured as a decentralized network composed of independent processing nodes. A central server acts as the primary API gateway, exposing WebSocket endpoints. To prevent resource contention and maximize GPU utilization, all heavy model inference (text-to-image, image-to-3D, and skybox generation) is completely decoupled from the API gateway. When an authoring request is received, the orchestration engine generates a unique job identifier, provisions tracking entries in Redis, and appends the payload to the designated queue. Remote GPU nodes consume tasks via non-blocking message-driven loops, executing inferences asynchronously without stalling client-side rendering.

The gateway communicates with dedicated remote GPU servers via a high-performance Redis message broker, enabling horizontal scalability as scene complexity or user load increases. To ensure reproducibility and isolated execution environments, each worker component is containerized with direct GPU access. Furthermore, worker services share a high-performance POSIX-compliant storage volume; once an asset is synthesized, only lightweight file path metadata is transmitted across network protocols, virtually eliminating file serialization overhead and network latency.

Worker Framework Extensibility

A fundamental architectural requirement of this framework is the ease of integrating new and emerging AI models without changing the communication and interaction layers. To ensure this, we established a worker abstraction that handles communication, state management, and common execution tasks. To validate this extensible design, we implemented and deployed specialized pipelines across our GPU workers: FLUX.1 via Z-Image (Black Forest Labs, 2024) executes the initial Text-to-2D stage, which feeds into 2D-to-3D mesh synthesis handled interchangeably by Hunyuan3D 2.0 (Zhao et al., 2025) or TRELLIS 2 (Xiang et al., 2025) to export clean GLB assets. Simultaneously, the Skybox worker integrates HunyuanWorld 1.0 (Tencent Hunyuan, 2025) to generate 360-degree equirectangular panoramic environments aligned with physical room boundaries. To change or deploy a newly released state-of-the-art AI model (e.g., a new model from Hugging Face) can therefore be integrated by developers in a transparent way from the MR application perspective. This structural isolation shields developers from the complexities of networking, cluster computing, and concurrent programming, enabling truly plug-and-play scaling.

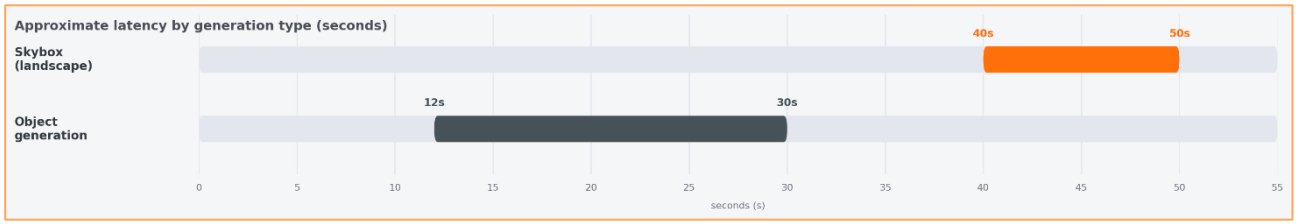


***Figure 102**. Average creation time*

Benchmarking the distributed inference pipeline on a NVIDIA A100 GPU (80 GB VRAM) demonstrated an end-to-end turnaround suitable for interactive authoring. For 3D assets, combining FLUX.1 concept generation (~5.0 s) with TRELLIS 2 (~5 s inference, ~7 s GLB export) achieved a total generation turnaround of ~17 s, substantially faster than Hunyuan3D 2.0 which required ~28 s. Meanwhile, panoramic skybox synthesis via HunyuanWorld completed in approximately ~50 s. This latency reduction keeps the Quest 3 client fully responsive and prevents spatial disorientation during runtime generation (Figure 2).

Mixed Reality Use Case: GenAI for Creating Personalized Mindful Environments

We have tested the proposed architecture in a MR scenario that allows users to create personalized relaxing scenarios This use case is particularly relevant in mental health contexts, where GenAI can support the personalization of therapeutic experiences for vulnerable users who may have specific interaction difficulties, cognitive or physical limitations, and highly individual needs, making the creation of personalized relaxation spaces especially valuable. We conducted several codesign sessions with XR experts, mental health professionals and users to identify the design requirements of the system. Direct physical interaction with 3D objects was prioritized over nested augmented menus. The following sections provide an overview of the final scene based on the codesign sessions, iterative implementation and user feedback.

Spatial Alignment and Voice-Driven Generation

Prior to launching the application, the client must perform an initial room scan using Meta Quest 3's native Space Setup tool. Leveraging the device's depth sensors and spatial meshing capabilities, physical surfaces are registered, specifically detecting tables and windows within the environment. In this manner, the largest table is selected as the central workspace, and the physical windows are replaced with virtual windows so that the generated skybox can be viewed through them.

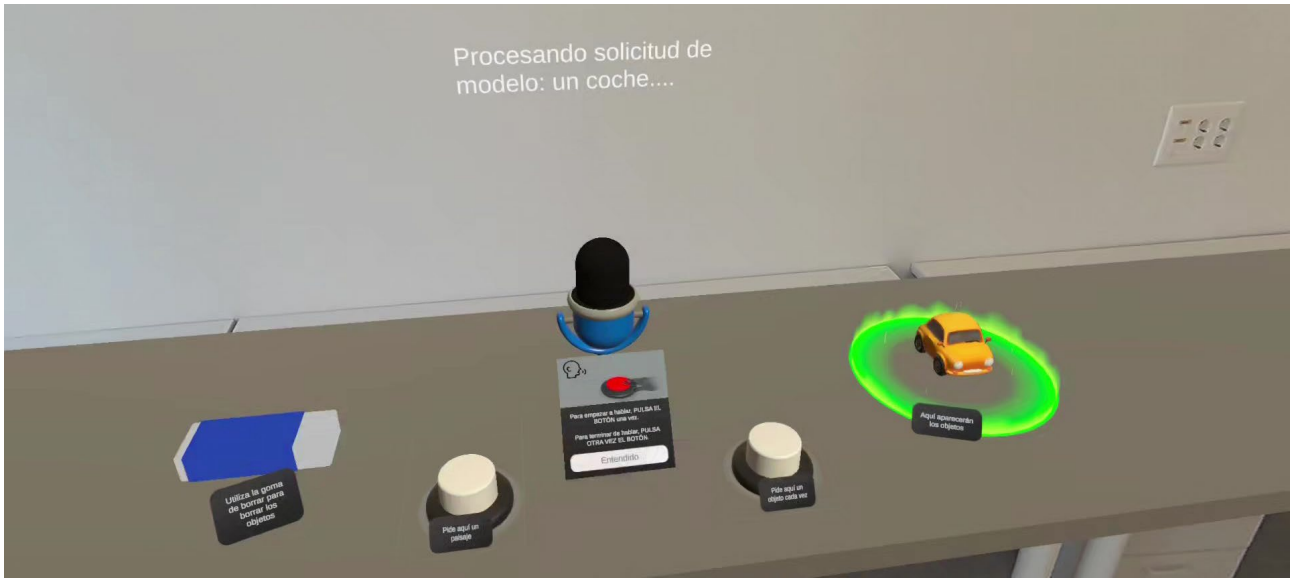


***Figure 103**. Physical workspace configured in Mixed Reality on the user's real table*

To anchor the experience in physical affordances and minimize spatial disorientation for neurodivergent users, the physical tabletop serves as the primary embodied workspace (Fig. 3). The tabletop consists of three distinct spatial elements, arranged for intuitive access. On the left is a virtual eraser designed to remove generated 3D objects from the scene. On the right side of the tabletop lies the generation zone: a glowing holographic ring that acts as the portal where generated objects materialize. Generation controls are positioned in the center via two virtual Toggle-to-Talk buttons: the left button coordinates skybox generation, while the right button manages 3D object creation. Pressing either button opens the microphone to capture Spanish speech; a second press ends the capture and triggers the entire generation process using the resulting transcription.

For the acquisition of user prompts, a speech recognition module based on the Meta Voice SDK (Wit.ai) was integrated. Once the text transcription is obtained, the resulting string is routed by an interaction manager that dynamically directs payloads to either the 3D asset or skybox pipeline based on the active button context.

Object Manipulation and Immersive Environment

Regarding 3D assets, once the GLB data is processed and downloaded into the scene, the system immediately attaches the necessary components using the Meta Interaction SDK, allowing the user to manipulate the virtual object with their physical hands to perform grabbing, scaling, and spatial placement actions anywhere in the environment. In parallel, when a skybox is processed, the panoramic texture is projected onto a custom equirectangular skybox shader to ensure proper lighting and visual coherence. This environment is then visible through the virtual windows identified during the preliminary spatial scan, replacing what users can see through the physical window with the digital 2D content that has been generated by the user (Fig. 4). Hence, voice is used not simply to replace text entry, but to preserve the immersive character of content generation. Users can

formulate requests while keeping their visual attention on the scene and their hands available for direct interaction.



Figure 4. *3D Object Manipulation*

Conclusion

In this work, we presented an interactive Mixed Reality framework for 2D and 3D asset generation at runtime, that overcomes the computational and memory constraints of lightweight standalone HMDs. Through a decoupled client-server architecture, our system offloads high-latency spatial AI inference from the real-time rendering loop. Benchmarking on an NVIDIA A100 verified interactive turnaround times (~17 s for complete 3D assets), enabling real-time authoring of mindful relaxation spaces without degrading client-side rendering. By the integration of a direct spatial interface featuring Toggle-to-Talk, natural language speech recognition, and physical hand-tracking manipulation (Fig. 4), the interface eliminates the cognitive barriers of 2D flat-screen workflows. This multimodal pipeline bridges the gap between spatial generative AI and accessible, situated interaction for neurodivergent individuals enabling the seamless generation of 3D objects and panoramic environments directly within the user's physical space.

Future work will continue along this research direction, focusing on improving the visual and geometric quality of the generated assets while further reducing generation latency to support more fluid and responsive immersive authoring experiences. We also plan to explore alternative representations for large-scale 3D content and environments, including point clouds and Gaussian Splatting, which may offer more efficient and scalable approaches for generating and rendering complex spatial scenes in Mixed Reality. Future iterations will also investigate edge computing and near-the-edge deployment strategies, bringing selected generative AI workloads closer to the XR device to reduce communication latency, improve responsiveness, and enable more flexible distributed inference architectures.

Acknowledgments

We would like to thank the mental health professionals and participants from Fundación SASM and Fundación ACAVALL for their valuable insights and feedback. This work has been funded by the IVACE/FEDER funds through the project CUIDAR (IMDEEA/2025/36).

References

Jiang, J. & Li, B. (2026). PartInteractor: Intent-Driven Part-Aware 3D Authoring for Continuous Co-Creation in XR. In: *Proceedings of the 39th Annual ACM Symposium on User Interface Software and Technology* (UIST '26). ACM, 13 p. https://doi.org/10.48550/arXiv.2608.01335

Jiang, W., He, W., Tan, Z., Kuang, Z., Yu, D., Hasegawa, S., ... & Sarsenbayeva, Z. (2026). CoGen3D: An Agentic Human-AI Co-Design Pipeline for 3D Asset Generation for Virtual Reality. *arXiv preprint arXiv:2607.03731*. https://doi.org/10.48550/arXiv.2607.03731

Kim, S., Kim, D., Son, T. & Woo, W. (2025). RealityCrafter: User-guided Editable 3D Scene Generation from a Single Image in Mixed Reality. In: *Adjunct Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (UIST '25 Adjunct). ACM, pp. 1–3. https://doi.org/10.1145/3746058.3758405

Lee, J., Aleotti, F., Mazala, D., Garcia-Hernando, G., Vicente, S., Johnston, O. J., ... & Van Brummelen, J. (2025, September). Imaginatear: Ai-assisted in-situ authoring in augmented reality. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-21).https://doi.org/10.1145/3746059.37476

Black Forest Labs. (2024). FLUX.1: Flow Matching for High-Fidelity Text-to-Image Generation. *Technical Report, Black Forest Labs*. Available at: https://github.com/black-forest-labs/flux

Tencent Hunyuan. (2025). HunyuanWorld 1.0: Generating Immersive, Explorable, and Interactive 3D Worlds from Words or Pixels. *arXiv preprint arXiv*:2507.21809. https://doi.org/10.48550/arXiv.2507.21809

Zhao, Z., Lai, Z., Lin, Q., Zhao, Y., Liu, H., Yang, S., ... & Guo, C. (2025). Hunyuan3d 2.0: Scaling diffusion models for high resolution textured 3d assets generation. *arXiv preprint arXiv:2501.12202*. https://doi.org/10.48550/arXiv.2501.12202

Xiang, J. et al. (2025). TRELLIS.2: Native and Compact Structured Latents for 3D Generation. *arXiv preprint* arXiv:2512.14692.

Zhu, M., Chen, J. & Li, B. (2026). When Generative AI Meets Extended Reality: Enabling Scalable and Natural Interactions. *IEEE Internet Computing*, 30(2), pp. 45–52. https://doi.org/10.1109/MIC.2025.3619462

Poole, B., Jain, A., Barron, J.T. & Mildenhall, B. (2022). DreamFusion: Text-to-3D using 2D Diffusion. In: *Advances in Neural Information Processing Systems* (NeurIPS 2022), 35, pp. 14556–14569. https://doi.org/10.48550/arXiv.2209.14988

Wen, R., Gupta, K., Hoermann, S., Billinghurst, M., Nassani, A. & Piumsomboon, T. (2026). From Prompt to Presence: Co-Creating Personalised Emotional Sanctuaries in VR with Generative AI. In: *Proceedings of the 31st International Conference on Intelligent User Interfaces* (IUI '26). ACM, pp. 955–971. https://doi.org/10.1145/3742413.3789067

XR NeedStruct: From Free-Text to Structured XR User Requirements for Industrial XR Prototyping

Fahd Maaroufi, Florence Danglade, and Jean-Remy Chardonnet

Arts et Métiers Institute of Technology, LISPEN, 71100 Chalon-Sur-Saône, France

Corresponding author: fahd.maaroufi@ensam.eu

Keywords: Extended reality, industrial XR, natural language processing, requirement formalization, XR prototyping.

Introduction and Research Problem

Extended Reality (XR), encompassing Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR), is increasingly used in industrial design, production, and maintenance (Rauschnabel et al., 2022). However, developing effective XR applications requires a clear formalization of end-user needs, which are often expressed through short, informal, and incomplete natural-language descriptions (Karre et al., 2024). Current requirements engineering practices for XR remain largely based on unstructured language, while more formal approaches are applied inconsistently. Existing industrial XR frameworks still rely mainly on manually defined, use-case-specific requirements (Gerhard & Siewert, 2024), and recent LLM-based XR assistants are generally limited to predefined functionalities (Tomkou et al., 2025). Consequently, a gap remains between informal user expressions and structured requirements that can directly support XR application development. To address this limitation, we propose a structured approach that transforms free-text industrial needs into explicit and exploitable XR user requirements.

Proposed Tool and Methodology

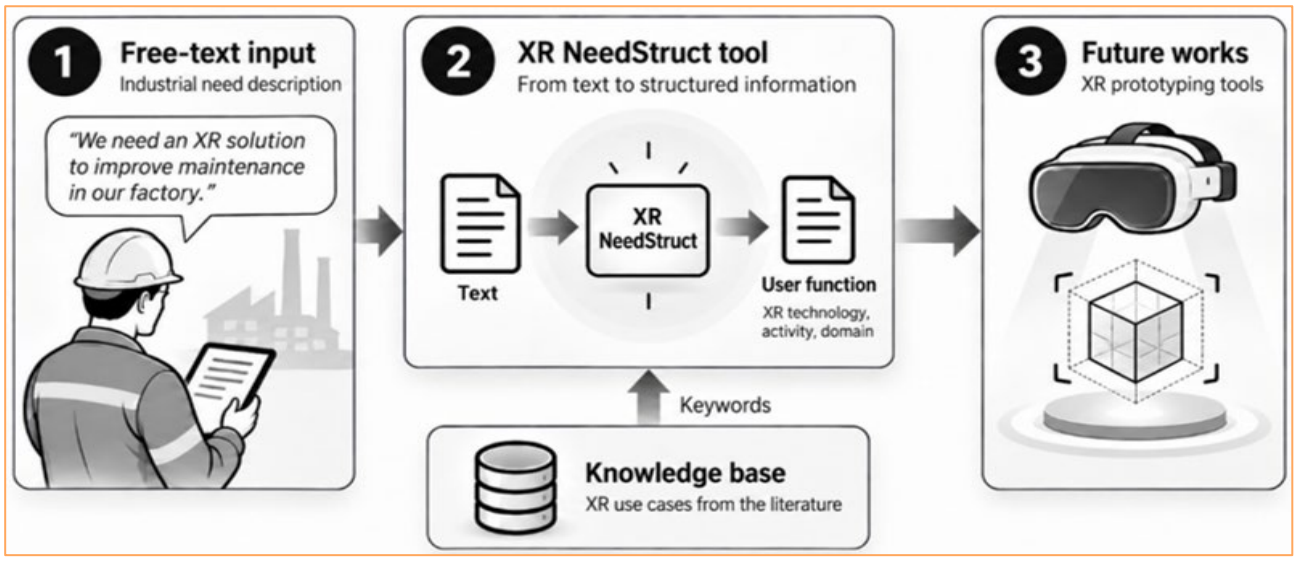


Figure 1. *Overall Workflow of the Proposed XR NeedStruct Approach*

Based on the user's free-text input, XR NeedStruct generates a structured representation of the user function (Figure 1). This representation includes the recommended XR technology, the relevant application domain (design, production, or maintenance), the targeted activity (e.g., design review, training, simulation, or assistance), the XR scene objects, and the real-world environment. The identified elements are then used to formalize the user functions according to a predefined sentence structure. In this structure, the generic fields are automatically replaced by the terms extracted by the tool:

"The **[XR technology]** system shall enable the user to **[target activity]** + **[application domain / environment]** of **[XR scene object]**."

This formulation converts the initial free-text description into a concise and structured user function that can be directly interpreted during the design of the XR application. For example, "The AR system shall enable the user to assist during pump assembly operations" or "The VR system shall enable the user to visualize the virtual prototype of a product during the design process."

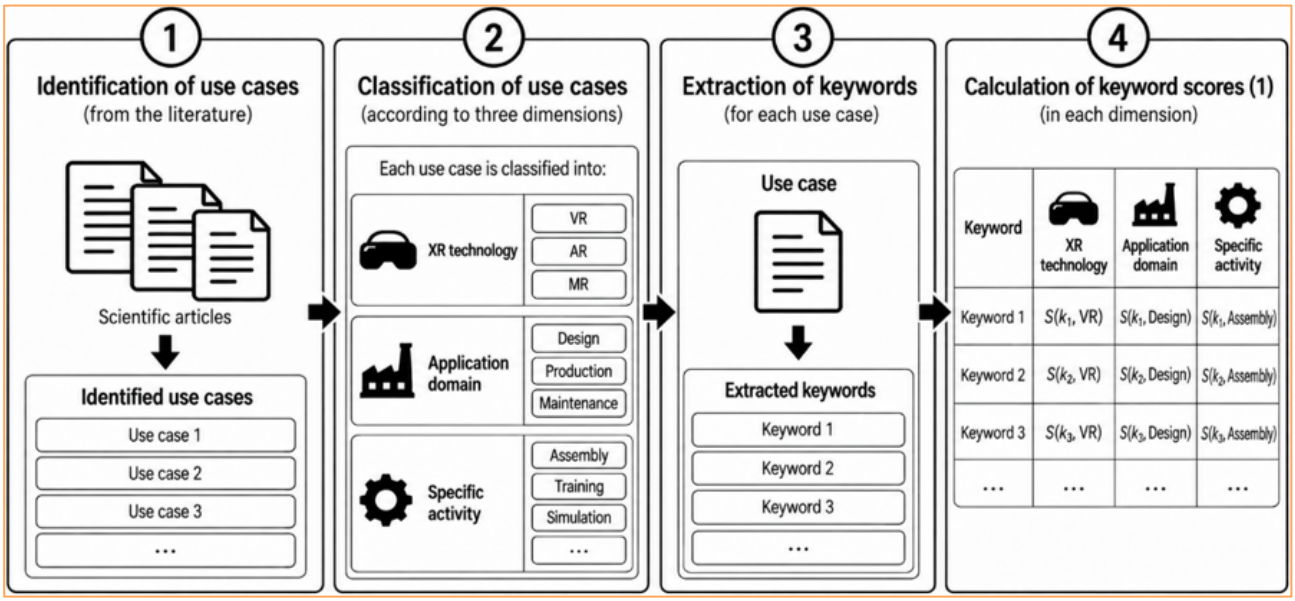


Figure 2. *Literature-Based Construction of Keyword Association Scores*

For the first three elements, the proposed method decomposes the free-text sentence expressing an industrial need into a set of keywords. Each keyword is assigned predefined association scores for XR technology, application domain, and activity. These scores are derived from a literature-based analysis of industrial XR use cases, in which each use case was classified according to these three dimensions and the corresponding keywords were extracted (Figure 2). For each keyword, a score was calculated independently for the three dimensions. The score of a keyword k for a given category c is defined as:

$$S(k,c) = \frac{N(k,c)}{\sum_{j=1}^{m} N(k,c_j)}$$

where N(k,c) represents the number of occurrences of keyword k associated with category c, and the denominator represents the total number of occurrences of the same keyword across all categories of the considered dimension. This calculation is performed separately for XR technology, application domain, and activity, thereby providing an independent association score for each dimension. For example, if a given keyword appears 10 times in studies classified as VR, 6 times in AR, and 4 times in MR, its score for VR is:

$$S(k,VR) = \frac{10}{10+6+4} = 0.50$$

meaning that 50% of the occurrences of this keyword are associated with VR within the technology dimension. The same calculation is independently applied to the application-domain and activity dimensions. When a user provides a free-text description, the tool identifies the

corresponding keywords and aggregates their literature-derived association scores for each candidate category. Within each dimension, the category with the highest aggregated score is selected, resulting in the recommended XR technology, application domain, and activity.

The identification of XR scene objects is based on a rule-based matching procedure supported by a predefined industrial object database. This database was constructed from industrial XR use cases reported in the literature. For each use case, the objects explicitly involved in the described task were extracted, including machines, equipment, components, small parts, tools, and elements of the industrial environment. Terms referring to the same object were grouped through synonyms and lexical variants. The extracted objects were then organized into functional categories inspired by manufacturing-domain ontology principles, particularly the distinction between manufacturing entities and resources proposed in MASON [Lemaignan et al., 2006]. The resulting hierarchy was adapted to XR scene representation and comprises the industrial scene, main object, sub-component, small component, tool, and virtual XR element. During analysis,the user's free-text description is first preprocessed through lowercasing, accent and irrelevant-character removal, spacing normalization, and spelling correction. The processed text is then matched against object names and associated synonyms in the predefined industrial object database. Detected objects are assigned to functional categories, including main objects (e.g., machines, robots, pumps, conveyors), sub-components (e.g., bearings, cables), small components (e.g., screws, bolts, nuts), and tools or measuring devices. Since the database cannot cover all possible industrial objects, unknown objects are flagged for user approval and, once confirmed, automatically added to the database for subsequent analyses. Finally, the role of each detected object within the XR scenario is determined from its functional category and the technical actions expressed in the user description, distinguishing central scenario entities, associated components, and tools involved in the required operations.

The involvement of the real environment is assessed through a separate rule-based analysis of linguistic indicators contained in the user input. The tool searches for expressions referring to physical objects, industrial spaces, superposition, scanning, recognition, localization, alignment, and spatial anchoring. These indicators are used to determine whether the scenario involves interaction with the physical environment and to characterize the relationship between real and virtual elements, such as virtual-content superposition in AR or spatial alignment in MR.

Experimental Validation

Validation was conducted using 16 free-text statements extracted from scientific articles and doctoral dissertations reporting needs and expectations explicitly expressed by industrial practitioners regarding XR technologies. Two XR domain experts independently translated each statement into a structured user function comprising five elements: XR technology, application domain, target activity, XR scene objects, and real-world environment. The use of 16 statements therefore resulted in 80 element-level evaluations (16 statements × 5 elements). For each statement, the expert-derived structured representation was compared, element by element, with the corresponding output generated by the proposed tool. Table 1 reports the resulting percentage agreement between the expert evaluations and the tool outputs for each of the five elements.

Validation against expert assessments showed strong agreement across the extracted user-function components, with correspondence ranging from 79.7% to 96.9%. The highest agreement was obtained for application domain identification (96.9%), while 3D object extraction showed the lowest correspondence (79.7%).

Table 12. *Correspondence between XR NeedStruct outputs and expert assessments across the extracted user-function elements*

User-function elements	Correspondence with experts
XR Technology	87.5%
Application Domain	96.9%
Target Activities	90.4%
Real-World Environment	90.6%
3D Objects	79.7%

Discussion and Future Work

This work contributes to industrial XR research by providing a structured method for converting early-stage industrial needs into formalized user functions. These user functions are intended to serve as direct inputs to the design guide developed in subsequent work. Each function specifies the recommended XR technology, application domain, target activity, XR scene objects, and real-world environment. Based on this structured representation, the guide will identify the engineering features relevant to the expressed need and use them to derive the corresponding functional and technical design specifications. These specifications will subsequently support the configuration and development of XR prototypes adapted to the targeted industrial use case. Accordingly, the proposed tool establishes a systematic link between informal industrial requirements and downstream XR design decisions.

By better aligning XR solutions with industrial needs, this approach can also support their adoption in industrial contexts. Overall, this work constitutes a first step toward the development of a prototyping tool for industrial XR applications.

References

Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H. & Alt, F. (2022). What Is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior* 133, Article 107289. https://doi.org/10.1016/j.chb.2022.107289

Karre, S. A., Reddy, Y. R. & Mittal, R. (2024). RE Methods for Virtual Reality Software Product Development: A Mapping Study. *ACM Transactions on Software Engineering and Methodology* 33(4), Article 88, 1–31. https://doi.org/10.1145/3649595

Gerhard, D. & Siewert, J. L. (2024). Industrial Augmented Reality: A Framework for Defining Requirements. *International Journal of Advances in Production Research* 1(1), 11–23.

Tomkou, D., Fatouros, G., Andreou, A., Makridis, G., Liarokapis, F., Dardanis, D., Kiourtis, A., Soldatos, J. & Kyriazis, D. (2025). Bridging Industrial Expertise and XR with LLM-Powered Conversational Agents. 2025 21st International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT). https://doi.org/10.1109/DCOSS-IoT65416.2025.00158

Lemaignan, S., Siadat, A., Dantan, J.-Y. & Semenenko, A.(2006). MASON: A Proposal for an Ontology of Manufacturing Domain. IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications (DIS'06), Prague, 195–200. https://doi.org/10.1109/DIS.2006.48

HoloHaptic Calls: Powering real-time, bidirectional haptic-augmented holographic communications using EU-based commercial-ready XR technology

Álvaro Egea[1], Marc Martos[2,3], Gijs Den Butter[4], Sergi Fernández[2,3], and Mario Montagud[1,3]

[1] *Universitat de València*

[2] *i2CAT Foundation*

[3] *Volum Tech*

[4] *SenseGlove*

corresponding author: mario.montagud@i2cat.net

Keywords: eXtended Reality (XR), Holographic Communications, Holoportation, Haptic Feedback, Haptic Gloves, Multimodal Interaction, Volumetric Video

Motivation - Scientific & Societal Relevance

Multiuser holoportation — enabling real-time, volumetric capture of users — offers a transformative leap for digital communications. Recent studies have demonstrated its advantages over avatar-based Social Virtual Reality (VR) and traditional 2D videoconferencing in terms of immersion, co-presence, and interaction quality (Montagud et al., 2022). Consequently, its potential is being actively explored across domains, including virtual meetings (Fernández et al., 2022), entertainment (Montagud et al., 2025), education (Egea et al., 2026), and virtual visits (Hjeijw t al., 2026), among others.

Beyond realistic audiovisual perception, tactile stimuli and interactions are a fundamental aspect for human-to-human relationships to convey emotional expressions, strengthen social bonds, contribute to naturalness, and build trust through non-verbal cues. While recent technological advances have resulted in promising haptic interfaces for interacting with digital elements in both Virtual and eXtended Reality (VR, XR) settings, state-of-the-art technology-mediated communication platforms still lack multiuser, bidirectional haptic and tactile capabilities—beyond preliminary studies restricted to avatar-based interactions in controlled setups (Kimmel et al., 2026).

To address this gap, this paper presents a pioneering proof-of-concept that seamlessly integrates cutting-edge XR holoportation technology from Volum Tech with force and tactile feedback gloves from SenseGlove (Figure 1), thus augmenting volumetric video communication paradigms with bi-directional, realistic and precise social and collaborative touch capabilities.

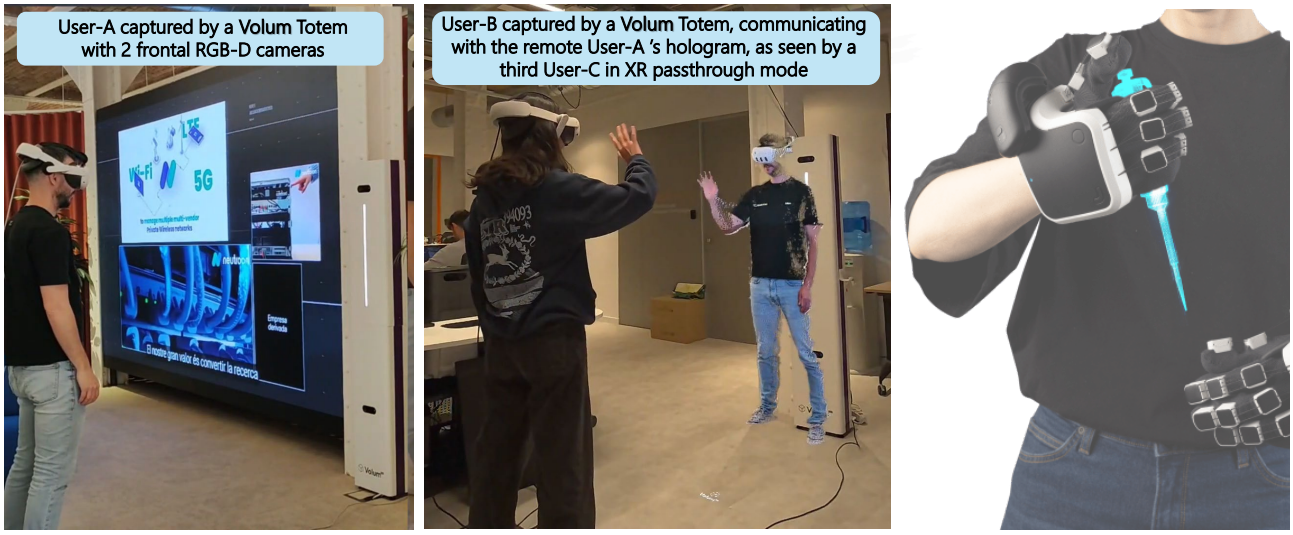


Figure 104. *(Left, Middle) XR holoportation technology by Volum Tech; (Right) Nova haptic gloves by SenseGlove*

Technological Architecture & Deployment Needs

The proposed proof-of-concept enables real-time, bidirectional tactile-augmented holographic communication between remotely captured users, offering a new paradigm for realistic and trustworthy digital interactions. It supports both full VR (where all users are holoported into a shared virtual environment) and hybrid XR (where volumetric content is projected directly into a physical space) setups. Therefore, the showcased demo will support two deployment variants, which will fit both the available space deployment and remote support resources:

- Deployment A — Local Totem-to-Totem: It features two locally deployed Volum Totem units (Figure 1). Each Totem integrates two RGB-D sensors for full human-size frontal capture, a dedicated mini-PC for spatial media processing, a Meta Quest 3 headset for XR/VR rendering, and a pair of SenseGlove Nova haptic gloves. It requires two adjacent / near-by 2x1m^2 physical areas.
- Deployment B — Local Totem to Remote / Pre-Recorded User: It eliminates the second local hardware station. The local participant interacts either in real-time with a remote user captured via a multi-sensor 360° rig at the i2CAT lab in Barcelona (Figure 2), or with a pre-recorded volumetric user performing audiovisual speech and interactive gesture narratives.

Audience Participation & Interaction

Conference attendees will physically and firsthand experience real-time holoportation augmented with haptic-enabled non-verbal interactions, being able to conduct diverse shared and interactive activities:

1. Performing natural handshakes with remote 3D holograms (Figures 2 and 3).
2. Feeling tactile feedback when touching volumetric digital elements, including the remote holograms.
3. Collaboratively manipulating shared virtual 3D objects in real time.

The following demo videos jointly support the capabilities of the proposed demo:

- Totem-to-Totem call scenario: https://youtu.be/xT-kBMecfqY
- Hybrid XR with a remote full volumetric user: https://youtu.be/0qH1wMZvIXA
- Handshake in a shared VR scenario: https://youtu.be/owL0bCR813o



***Figure 2**. XR holographic call through Volum technology.*

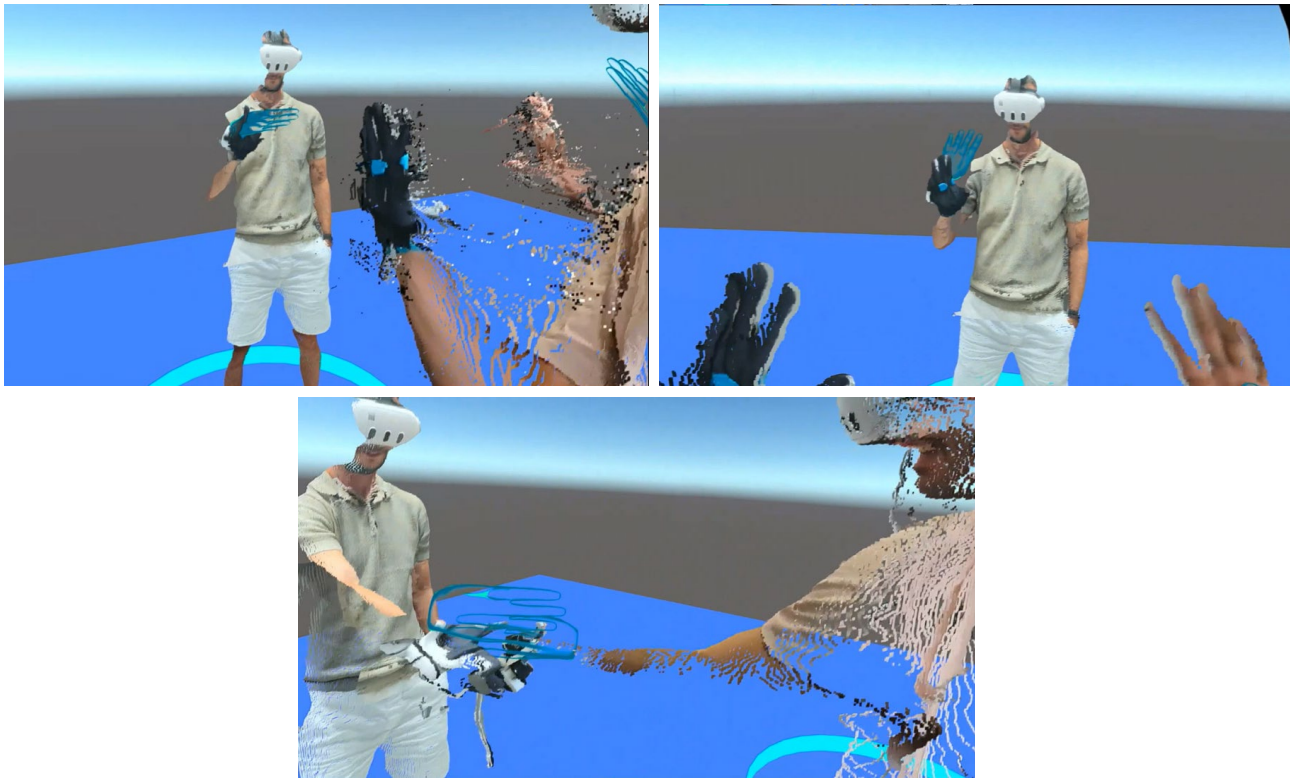


***Figure 3.** Haptic-augmented VR holographic calls through Volum and SenseGlove technologies.*

Acknowledgments

The work has funded by European Union's Horizon Europe program, under agreement nº 101135025 (PRESENCE project) and agreement n◦ 101298582 (ORIGINS project), and by Agencia Estatal de Investigación (AEI), in the framework of Proyectos Generación de Conocimiento 2022, under agreement PID2022-140749OBI0 (EVOLVE project). The work of Álvaro Egea has been funded by a pre-doctoral fellowship from the Conselleria de Educación, Universidades y Empleo of the Generalitat Valenciana (CIACIF/2024/338). The work of Mario Montagud has been funded by MCIN/AEI/10.13039/501100011033 under Grant RYC2020-030679-1 and by "the European Social Fund (ESF) Investing in Your Future".

References

Egea, Á, Gálvez, M., Rodrigo, A., Gómez-Barquero, D., Fernández, S., & Montagud, M., 2026. Enabling Interactive XR Holographic Telepresence in Education with Affordable and Easy-to-Use Systems. In *Proceedings of IEEE Telepresence* 2026. https://doi.org/10.1109/DIS.2006.48.

Fernández, S. Montagud, M., Cernigliaro, G., & Rincón-Rivera, D. (2022). Multiparty holomeetings: Toward a new era of low-cost volumetric holographic meetings in virtual reality. *IEEE ACCESS*, 10, 81856-81876.

Hjeij, M., Montagud, M., Rincón-Rivera, D., & Fernández, S. (2026). Multimodal and Social Virtual Reality (VR): exploring and validating promising enablers for next-generation interactive and group-based virtual visits. *Applied Sciences*, 16(8), 4002. https://doi.org/10.3390/app16084002

Kimmel, S., Goodarzi, S., Kaiser, J. N., Landwehr, E., & Heuten, W. (2026). Feel the Connection: Assessing the Influence of Haptic Social Touch on Social Presence in Remote VR Interactions. In *Proceedings of the Augmented Humans International Conference* 2026 (pp. 442-454). https://doi.org/10.1145/3795011.3795024

Montagud, M., Li, J., Cernigliaro, G., El Ali, A., Fernández, S., & Cesar, P. (2022). Towards socialVR: evaluating a novel technology for watching videos together. *Virtual Reality*, 26(4), 1593-1613.

Montagud, M., Martos, M., Egea, Á., Fernández, S. (2025). Social VR with holographic comms: Enablers for new engaging experiences within the TV / video consumption landscape. *IEEE Transactions on Broadcasting*, 71(3), 793-807.

A Traffic Interception Framework for Auditing Biometric Data in VR Applications

Abril Pérez, Marcos Morán, David Rodríguez, and Rodrigo García

Information Processing and Telecommunications Center (IPTC), Universidad Politécnica de Madrid, ETSI Telecomunicación, Avda. Complutense 30, Madrid, Spain

Corresponding author: abril.perez@alumnos.upm.es

Keywords: Virtual reality; privacy; man-in-the-middle; network traffic; biometric data

Introduction

Virtual Reality (VR) applications collect a large amount of varied data, including the user's environment characteristics, and biometric and physiological data: body movement patterns and biomechanics, hand tracking, eye tracking, facial expressions and maps of the user's physical space. Such data can uniquely identify an individual (Miller et al. 2020) and may reveal sensitive attributes, bringing them close to the special categories of data that the GDPR protects with enhanced safeguards (European Parliament and Council of the European Union 2016). Such data can uniquely identify an individual (Miller et al. 2020; Nair et al. 2023). However, there is often a gap between what these applications' privacy policies state and the data processing they carry out. As a preliminary study, this work establishes a non-root experimental infrastructure to intercept and capture encrypted network traffic from commercial VR applications developed with Unity, for the subsequent auditing of the captured data.

Methodology

Unlike standard Android environments, Meta Quest headsets restrict bootloader unlocking through proprietary cryptographic signatures. Privilege escalation relies entirely on firmware-specific vulnerabilities (e.g., CVE-2021-1931; NIST 2021; darknight1050 2021), which are fully patched in the latest firmware versions of these devices. Additionally, unauthorized modifications to the bootloader partition result in an unrecoverable software state (hard brick) due to Meta's hardware-level secure boot implementation (QFPROM). This prevents Emergency Download (EDL) recovery by withholding public cryptographic keys, so standard reflashing requires expensive physical chip reprogramming. Consequently, root-based traffic interception is not viable, and alternative non-root approaches are required.

To capture network traffic, a man-in-the-middle (MITM) setup (Figure 1) was deployed in a controlled laboratory environment, taking the OVRseen project (Trimananda et al. 2022) as a starting point. Initially, a Meta Quest 1 with Android 10 was used because AntMonitor, the tool used to intercept traffic in the original repository, generates a unique cryptographic CA whose user-level installation is blocked by the Quest security system on Android 11 and later.

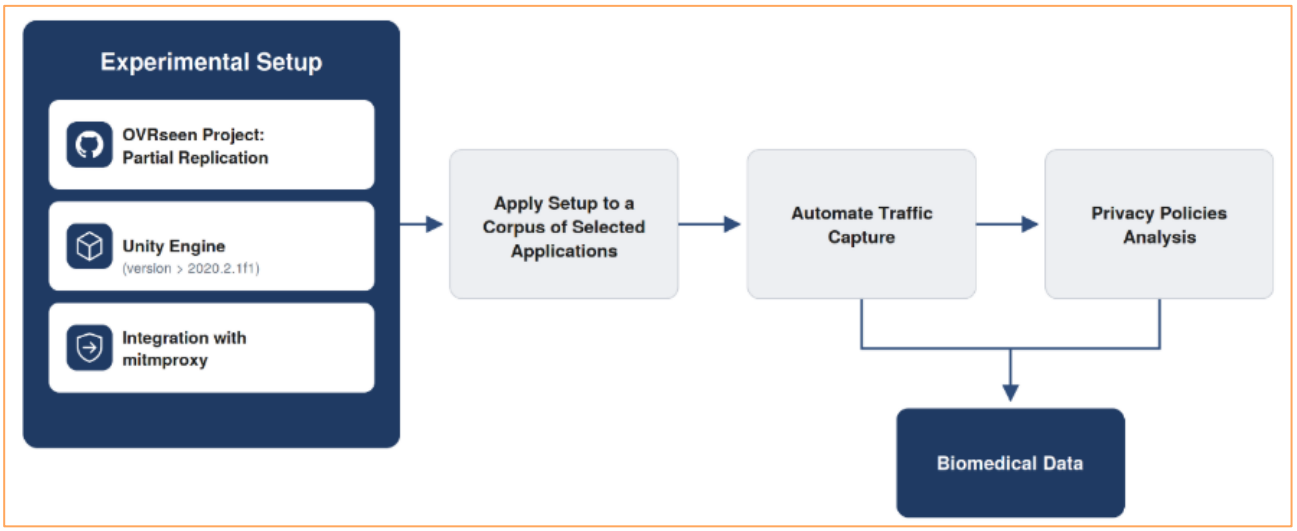


***Figure 1**. Methodological workflow for the technical audit of VR applications. The framework outlines the experimental setup for traffic interception, its deployment across a selected application corpus and the subsequent dual analysis of privacy policy compliance and biometric data extraction (future work).*

Preliminary Results

The experimental infrastructure is now fully operational: the initial phase has established a functional non-root laboratory setup for network traffic interception on Meta Quest devices. The current setup bypasses SSL pinning by decompiling target APKs, injecting the Frida-gadget library, adjusting JSON initialization parameters to evade SELinux blocks, and repackaging the applications for user-certificate trust. The tool mitmproxy was also integrated into the pipeline to replace AntMonitor; it routes all network traffic from the device through itself as an intermediary proxy, decrypting HTTPS streams with a custom root CA installed on the client.

The OVRseen framework (Trimananda et al. 2022) has been partially replicated and validated in an Ubuntu virtual machine to automate the capture of encrypted telemetry. Encrypted traffic flows from commercial VR applications have already been intercepted (Figures 2 and 3, Listing 1), confirming the technical viability of the pipeline and enabling the systematic audit that follows.

Path	Method	Status	Size	Time
https://perf-events.cloud.unity3d.com/api/v2/projects/1fc4d10…	POST	202	4.2kb	1s
https://config.uca.cloud.unity3d.com/	POST	200	762b	1s
https://cdp.cloud.unity3d.com/v1/events	POST	200	552b	934ms
https://cdp.cloud.unity3d.com/v1/events	POST	200	12.1kb	1s
https://perf-events.cloud.unity3d.com/api/v2/projects/1fc4d10…	POST	202	4.3kb	502ms
https://cdp.cloud.unity3d.com/v1/events	POST	200	530b	160ms
https://cdp.cloud.unity3d.com/v1/events	POST	200	576b	239ms
https://cdp.cloud.unity3d.com/v1/events	POST	200	531b	190ms
https://cdp.cloud.unity3d.com/v1/events	POST	200	530b	238ms
https://graph.oculus.com/graphql	POST	200	1.6kb	532ms

Figure 2. *Network requests captured via mitmproxy during a Playground VR session on the Meta Quest 1, showing repeated POST traffic to Unity Cloud Diagnostics (cdp.cloud.unity3d.com), Unity performance-events (perf-events.cloud.unity3d.com), and Oculus GraphQL (graph.oculus.com) endpoints.*

Request Response Details

```
POST https://cdp.cloud.unity3d.com/v1/events HTTP/1.1
Host                          cdp.cloud.unity3d.com
User-Agent                    UnityPlayer/2019.4.8f1 (UnityWebRequest/1.0, libcurl/7.52.0-DEV)
Accept-Encoding               deflate, gzip
Accept                        */*
Content-Type                  application/json
X-Event-Count                 5
X-Highest-Priority-Events     0
X-Data-Block-Id               [redacted]
X-Expired-Session-Dropped     0
X-Data-Retry-Count            1
X-Continuous-Request          2
X-Request-Ts                  1701793534514
X-Unity-Version               2019.4.8f1
Content-Length                12399
```

Figure 3. *HTTP request headers for a representative POST to cdp.cloud.unity3d.com/v1/events, showing Unity-specific tracking headers (X-Data-Block-Id, X-Unity-Version, X-Request-Ts). Identifiers blurred.*

```
{"common": {
    "appid": "1fc4d104-818e-435e-92a2-05077564c235",
    "userid": "<REDACTED>", "sessionid": "<REDACTED>", "platform": "AndroidPlayer",
    "platformid": 11, "sdk_ver": "u2019.4.8f1", "sdk_rev": "60781d942082", "session_count": 1,
    "localprojectid": "<REDACTED>", "build_guid": "<REDACTED>", "deviceid": "<REDACTED>",
    "identity_token": "<REDACTED>"
  }
}{
  "type": "analytics.appInstall.v1",
  "msg": { "unity_ver": "2019.4.8f1", "app_ver": "1.3.3", "ts": 1781793533329, "t_since_start": 63725889 }
}{
  "type": "analytics.deviceInfo.v1",
  "msg": {
    "ts": 1781793533333, "t_since_start": 63729336,
    "os_ver": "Android OS 10 / API-29 (QQ3A.200805.001/49845030443200410)",
    "app_ver": "1.3.3", "rooted_jailbroken": false, "debug_build": false,
    "vr_device_name": "Oculus", "vr_device_model": "Oculus Quest",
    "vr_devices": ["Oculus", "OpenVR"],
    "model": "Oculus/Quest/monterey", "device_name": "Meta Quest",
    "cpu": "ARM64 FP ASIMD AES", "cpu_count": 3, "cpu_freq": 1900,
    "ram": 3808, "vram": 1024, "screen": "2880 x 1600", "dpi": 235,
    "lang": "en", "sensors": 150,
    "gfx_name": "Adreno (TM) 540", "gfx_vendor": "Qualcomm",
    "gfx_ver": "OpenGL ES 3.2", "gfx_shader": 50, "max_texture_size": 16384
  }
}{
  "type": "analytics.xrDeviceInfo.v1",
  "msg": {
    "ts": 1781793533337, "t_since_start": 63733049,
    "xr_device_name": "Oculus", "xr_device_model": "Oculus Quest",
    "vr_refresh_rate": 72.0,
    "vr_play_area_dimension": { "x": 7.147627353668213, "y": 10.0, "z": 4.763949394226074 },
    "vr_tracked_area_dimension": { "x": 7.147627353668213, "y": 10.0, "z": 4.763949394226074 },
    "vr_play_area_geometry": [
      { "x": 3.8547046184539797, "y": 0.0, "z": 3.4848313331604006 },
      "...",
      { "x": 0.835798442363739, "y": 0.0, "z": -2.9939632415771486 }
    ],
    "vr_tracked_area_geometry": [
      { "x": 0.3732522428035736, "y": 0.01015329360961914, "z": 6.9700608253479 },
      "...",
      { "x": -1.2201303243637086, "y": 0.010135054588317871, "z": 5.434751033782959 }
    ]
  }
}
```

***Listing 1**. Excerpt of intercepted telemetry payload corresponding to the POST reproduced in Figure 3. Identifiers redacted; coordinate array elided between endpoints.*

As shown in Listing 1, the intercepted telemetry transmits explicit spatial metrics that can be classified as sensitive personal data. Specifically, the payload extracts the vr_play_area_dimension and the exact 3D coordinates of the vr_tracked_area_geometry. These data points map the precise size, shape, and physical boundaries of the user's room. Because a physical room's layout is highly stable and mathematically unique to the individual's home, these metrics enable persistent spatial fingerprinting. By logging these geometric signatures, third parties can uniquely identify and track users across different applications, accounts, or even different VR headsets.

Future Work

Immediate future work involves automated capture of a selected application corpus, followed by static and dynamic analysis to determine which categories of data, particularly biometric and physiological, are collected and transmitted, for what purpose, and to which recipients (including SDKs and third parties). A privacy policy analysis will then compare these observed data flows against each application's stated policies and GDPR requirements.

Long-term future work will explore anonymization techniques to safeguard user privacy in VR environments, and will extend this auditing framework to newer hardware, such as the Meta Quest 3 and 3S.

References

Darknight 1050. (2021). quest-bootloader-unlocker. GitHub. https://github.com/darknight1050/quest-bootloader-unlocker

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). *Official Journal of the European Union* L 119, 1–88.

Miller, M. R., Herrera, F., Jun, H., Landay, J. A. & Bailenson, J. N. (2020). Personal identifiability of user tracking data during observation of 360-degree VR video. *Scientific Reports* 10, 17404.

Nair, V., Guo, W., Mattern, J., Wang, R., O'Brien, J. F., Rosenberg, L. & Song, D. (2023). Unique identification of 50,000+ virtual reality users from head and hand motion data. In: *Proc. 32nd USENIX Security Symposium*, Anaheim, CA, USA, pp. 895–910.

NIST. (2021). CVE-2021-1931 Detail. National Vulnerability Database. https://nvd.nist.gov/vuln/detail/CVE-2021-1931

Trimananda, R., Le, H., Cui, H., Tran Ho, J., Shuba, A. & Markopoulou, A. (2022). OVRseen: Auditing network traffic and privacy policies in Oculus VR. In: *Proc. 31st USENIX Security Symposium*, Boston, MA, USA, pp. 3789–3806.

Device Tracking Error and Headset Slip: Two Error Sources in HMD-Reported Head Pose for Rehabilitation

Teresa Pérez de Arenaza Pozo[1], Carlos Plaza Cuenca[2], Javier Andión Jiménez[1], Abraham Otero Quintana[2], Rodrigo García Carmona[1]

[1] *Information Processing and Telecommunications Center (IPTC) Universidad Politécnica de Madrid, ETSI Telecomunicación, Madrid, Spain*

[2] *Department of Information Technologies, Institute of Technology, Universidad San Pablo-CEU, CEU Universities, Urbanización Montepríncipe, 28660 Boadilla del Monte, Madrid, Spain*

Corresponding autor: t.perezdearenaza@alumnos.upm.es

Keywords: Virtual reality; head-mounted display; motion capture; tracking accuracy; headset slippage; physical therapy

Abstract

When a clinician uses a Virtual Reality (VR) head-mounted display (HMD) to measure the position, angle or other derived biomechanical quantities, like range of motion (ROM), this measurement can accumulate errors due to two limitations: the ability of the device to record its own pose accurately, and the relative displacement of the HMD against the user's head. We have measured both error sources on a Meta Quest 3S headset worn by a person against an optical motion capture (OMC) system, that was considered the source of truth. Preliminary results indicate that the tracking accuracy of a worn headset is comparable to values previously reported for HMDs rigidly mounted onto tripods or mannequin heads. Under the tested conditions, headset slip relative to the head stayed below the device tracking error but varied by more than an order of magnitude depending how well the headset was fitted.

Introduction

The use of head-mounted displays (HMDs) in rehabilitation has gradually evolved with time. First used to deliver engaging therapeutic content, they are now measuring instruments in their own right: the pose the HMD reports can be recorded and analyzed to assess the patient's progress (Santos-Paz et al. 2022; Forsberg et al. 2023). Cervical range of motion (ROM), one of the most common assessments in neck rehabilitation, is usually measured with a universal goniometer: small, portable and low-cost, but dependent on an experienced practitioner for a reliable reading (Santos-Paz et al. 2022). The alternative, an optical motion capture laboratory, is accurate but costly, fixed in place and available to very few clinics. A consumer HMD sits between the two, offering instrumented, operator-independent measurement at a cost an ordinary hospital (or even a patient) can absorb; recent work reports excellent reliability for VR-based cervical assessment, with no significant difference between high-end and low-cost standalone headsets (Santos-Paz et al. 2026), and high concurrent validity against a gold-standard optical system (Forsberg et al. 2023).

Multiple studies have evaluated HMD tracking accuracy, reporting errors on the order of millimeters (Niehorster et al. 2017; Banaszczyk et al. 2024). In these evaluations the HMD is held rigidly in place on a tripod or a mannequin head, to isolate the device from any other source of error. For a clinical purpose these measurements are therefore incomplete: slippage caused by head movements, facial expressions, speech or a loosening headstrap are not accounted for.

The stability of the mounting itself has been examined in prior work, but in a different field and with a different outcome. With head-mounted eye-tracking, headset slip is known to degrade data quality. Even minor movements like speech or facial expressions can cause gaze tracking errors of 0.8–3.1° (Niehorster et al. 2020). The equivalent quantity for a VR HMD has not, to our knowledge, been measured, although it has been identified as a limitation of VR-based cervical assessment (Santos-Paz et al. 2022). The headset design, weight and fit have been found to affect the reliability of such assessments more than the tracking technology itself does (Santos-Paz et al. 2026), but a true measurement of the impact of these factors have not been published.

This work pursues two parallel lines of enquiry: whether the accuracy reported under rigid laboratory mounting persists when the same headset is worn by a freely moving individual, and how much the headset displaces relative to the head during use.

The contribution is a separate experimental characterization of both error sources in worn HMDs, using optical motion capture as an external reference.

Methods

The investigation consisted of two experiments. The first compared the headset's reported pose against the optical ground truth in order to quantify the tracking error of the system; the second tracked the headset and the participant separately in order to quantify the movement of the headset relative to the head. The same motion capture system was used in both experiments (OptiTrack, 12 cameras, 100 Hz). 3 volunteers took part and all participants gave written informed consent.



Figure 1. *Experimental setup for Experiment B. Retroreflective markers are attached to the headset shell and to a close-fitting cap worn by the participant, defining two rigid bodies tracked independently by the optical system.*

Experiment A compared the headset's own reported pose against the optical ground truth while a participant wore a Meta Quest 3S device while moving freely. The pose the headset reports refers to an internal virtual point, the center eye anchor, displaced from the markers on the shell by an unknown but constant rigid transform, so the two streams cannot be compared directly. This transform was recovered by solving the hand-eye calibration problem $A \cdot X = X \cdot B$, a standard robotics formulation for the fixed relationship between two rigidly connected frames whose motions are observed separately; we used the dual-quaternion solution (Daniilidis 1999), which in comparable HMD work was found to converge more consistently than alternative formulations (Banaszczyk et al. 2024). Once the transform was known, the reported trajectory could be expressed in the frame of the optical system and the remaining divergence reported as absolute pose error (APE). We depart from Banaszczyk et al. (2024) in that the two systems were started independently; the offset in recording start times was recovered before pairing by correlating velocity magnitude in each stream, a quantity invariant to the reference frame (Mair et al. 2011).

Experiment B placed a second rigid body on a close-fitting cap worn by the participant. The cap was taken as a proxy for the skull, and the headset pose was expressed in the cap's own reference frame, in which head motion cancels. Slip was defined as the deviation from a baseline established at the beginning of each recording, with the participant stationary and the placement of the HMD verified by the operator. Measurements were carried out with two HMDs (HTC Vive Pro, Meta Quest 3) under conditions in which the participant rotated their head about all three axes, produced exaggerated facial expressions, and deliberately displaced the headset, at fitting levels ranging from tight to loose.

Both experiments were summarized with root-mean-square error and percentiles, computed for Experiment A on the APE (Banaszczyk et al. 2024); we report the median and 95th percentile to characterize typical and upper-range errors. All comparisons between the two error sources are made at the 95th percentile.

Results

Over a 30 s recording, the reported pose diverged from optical ground truth by 3.15 mm RMSE in translation (p50 2.65, p95 5.35, maximum 10.71 mm) and 0.45° RMSE in rotation. This falls within the range previously reported for rigidly mounted headsets (Banaszczyk et al. 2024), providing early evidence that the accuracy measured under rigid mounting is retained when the headset is worn.

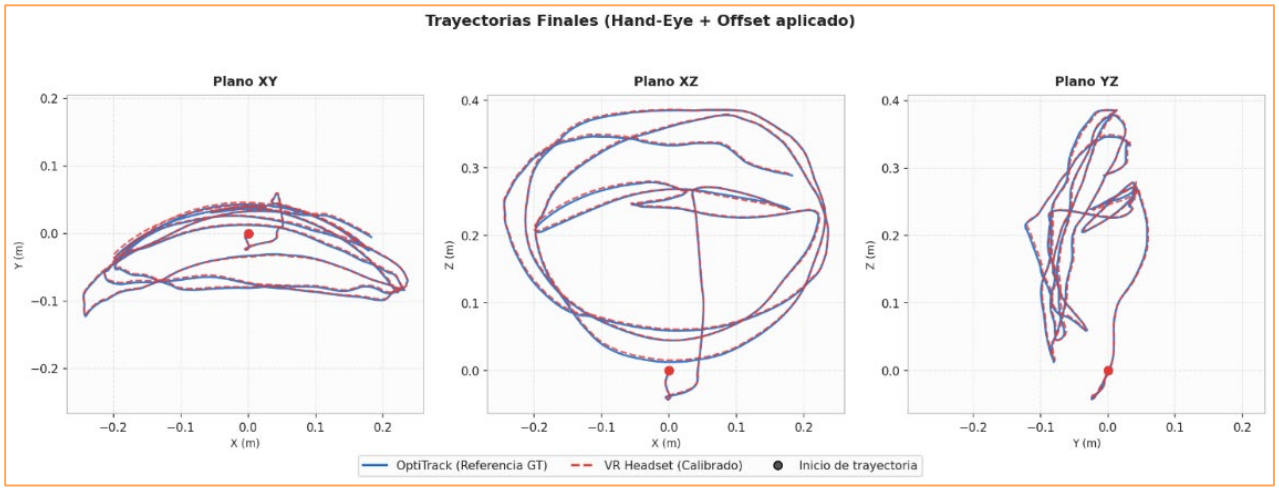


Figure 2. *OptiTrack ground truth (blue) and the headset-reported trajectory (red) after hand-eye calibration and temporal alignment, projected onto the three coordinate planes.*

As can be seen in Figure 2, fit mattered more than movement. With the participant sitting still, slip reached 0.13 mm at the 95th percentile, close to the measurement noise of the optical system. A carefully seated headset held to 0.78 mm even while the participant produced exaggerated facial expressions. Rotating the head about all three axes with a slightly loose fit gave 1.29 mm on the

HTC Vive Pro and 1.37 mm on the Meta Quest 3; a very loose fit reached 1.80 mm; and deliberately repositioning the headset mid-session, the worst case tested, reached 2.92 mm (all values at the 95th percentile).

How the headset was worn mattered more than how hard the participant moved: in two sessions in which the participant simply sat still, strap tightness alone changed slip by more than an order of magnitude (0.13 mm against 1.80 mm at the 95th percentile). Referred to the same physical point on the headset, device error exceeded slip in every condition tested, including deliberate repositioning, and was roughly seven times the slip of a properly seated headset (5.35 mm against 0.78 mm at the 95th percentile).

Limitations and Future Work

The two errors come from separate recording sessions, because our current setup cannot track the cap, the headset and the headset's own recordings at the same time; we can therefore say how large each error is relative to the other, but not how the two combine. A further caveat is that the cap stands in for the skull: if cap and headset slid together the analysis would report no slip. Recordings lasted between 30 and 60 s, so we make no claim about how much error accumulates over a full therapy session, and the results rest on a small sample (3 participants), so the absolute magnitudes may not generalize. All these are being addressed: simultaneous capture of all three streams, separating the two errors within a single recording; markers fixed over the mastoid processes in place of the cap; longer recordings spanning a full therapy session; and a larger, more varied sample.

References

Banaszczyk, A., Łysakowski, M., Nowicki, M. R., Skrzypczyński, P. & Tadeja, S. K. (2024). How accurate is the positioning in VR? Using motion capture and robotics to compare positioning capabilities of popular VR headsets. In: 2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 79–86. https://doi.org/10.1109/ISMAR-Adjunct64951.2024.00027

Daniilidis, K. (1999). Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research* 18(3), 286–298. https://doi.org/10.1177/02783649922066213

Forsberg, K., Jirlén, J., Jacobson, I. & Röijezon, U. (2023). Concurrent validity of cervical movement tests using VR technology — taking the lab to the clinic. *Sensors* 23(24), 9864. https://doi.org/10.3390/s23249864

Mair, E., Fleps, M., Suppa, M. & Burschka, D. (2011). Spatio-temporal initialization for IMU to camera registration. In: 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 557–564. https://doi.org/10.1109/ROBIO.2011.6181345

Niehorster, D. C., Li, L. & Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research. *i-Perception* 8(3), 1–23. https://doi.org/10.1177/2041669517708205

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E. & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods* 52(3), 1140–1160

Santos-Paz, J. A., Sánchez-Picot, Á., Rojo, A., Martín-Pintado-Zugasti, A., Otero, A. & García-Carmona, R. (2022). A novel virtual reality application for autonomous assessment of cervical range of motion: development and reliability study. *PeerJ* 10, e14031. https://doi.org/10.7717/peerj.14031

Santos-Paz, J. A., García-Carmona, R., Sánchez-Picot, Á. & Otero, A. (2026). Comparative evaluation of virtual reality setups for assessing cervical range of motion: a reliability study. Virtual Reality 30(1), 35.

One Twin, Many Realities: A Collaborative Physics AI Digital Twin Across XR Virtual Worlds

Afonso Ribeiro de Oliveira, Carmen Díaz-Marín, and Carlos Peña-Monferrer*

SimZero, Ciudad Politécnica de la Innovación (CPI), UPV Campus de Vera, Valencia, Spain

*Corresponding author: carlos@simzero.com

Keywords: Digital Twin; Virtual Worlds; Physics AI; Collaborative XR; Heterogeneous XR

Introduction

Extended reality (XR) provides spatial interfaces to cyber-physical systems and digital twins, enabling users to interact with physical assets while accessing operational, sensor, and predictive information (Tu et al., 2023). Combined with Physics AI, including physics-informed and Scientific Machine Learning (SciML) approaches (Cicirello, 2024), digital twins can further incorporate real-time predictions of physical phenomena and support what-if exploration in virtual space.

In previous work, we demonstrated Physics-Informed XR Telepresence, in which a remote operator controlled a physical autonomous mobile robot (AMR) through a VR environment augmented with real-time ML predictions derived from computational fluid dynamics (CFD) simulations (Peña-Monferrer et al., 2025). While demonstrating the feasibility of combining physical-system operation with real-time scientific predictions in XR, it considered a single operator and XR environment.

Collaborative digital twins can extend these capabilities to multiple participants (Dewberry et al., 2025), who may have different roles, information needs, and interaction contexts. For example, an on-site operator, remote technician, and domain expert may benefit from different interfaces, viewpoints, and levels of immersion. Moreover, while some users need to remain synchronized with the live operational state, others may need to explore alternative Physics AI scenarios without affecting ongoing physical operation.

Building on our previous infrastructure, this work introduces a collaborative digital twin architecture connecting heterogeneous XR virtual worlds to a single authoritative digital twin state. It combines: (1) synchronization of the authoritative state across XR clients; (2) on-demand deployment of virtual worlds; (3) local Physics AI what-if exploration; and (4) transferable operational authority. The concept is demonstrated with three participants using tablet-based augmented reality (AR), a desktop virtual environment (VE), and immersive virtual reality (VR) while interacting with an AMR and real-time SciML predictions of airborne contaminants.

System Architecture

The presented architecture, illustrated in Fig. 1, is organized around the authoritative state shared across multiple heterogeneous XR virtual worlds. Rather than maintaining independent twins for each interface, the connected XR clients access the authoritative state, which integrates the

physical asset state, sensor information, operational Physics AI inputs and predictions, collaborative information, and control authority. Each virtual world can provide its own representation, scale, viewpoint, and interaction mechanisms while remaining connected to the same underlying cyber-physical system. Virtual worlds can be deployed on demand from predefined templates developed using platforms such as Unity, Unreal Engine, or Babylon.js, enabling different XR experiences to be instantiated as needed.

A synchronization layer propagates changes in the authoritative state across the connected clients. Updates originating from the physical asset, sensor information, operational Physics AI inputs and predictions, or collaborative interactions are reflected across the virtual worlds. This allows participants using different interfaces to maintain a consistent view of the operational state while interacting with the digital twin according to individual roles and device capabilities.

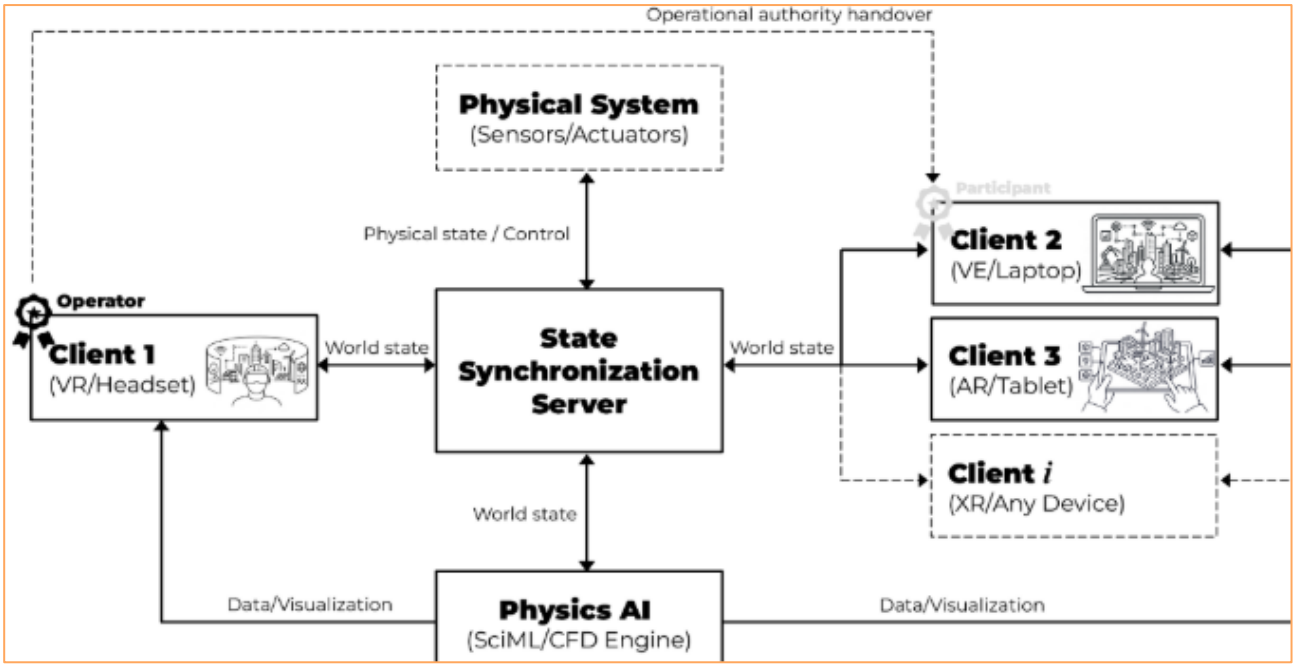


Figure 1. *Collaborative digital twin architecture across heterogeneous XR interfaces.*

Alongside the authoritative state, the architecture supports local state exploration. A participant can instantiate alternative Physics AI inputs and predictions within their own virtual world without modifying the authoritative state. For example, while one participant operates the physical system using the current operational parameters, another can modify model inputs locally and inspect the resulting alternative predictions. The physical asset and authoritative state remain unaffected, thereby providing an isolated environment for what-if analysis.

Finally, operational authority determines which participant can modify the live cyber-physical system. At a given time, a single participant acts as the active operator and is authorized to command the physical asset and modify the operational inputs of the Physics AI model. Other participants can monitor the shared state, interact with collaborative information, and perform local exploration without issuing conflicting operational commands. Because operational authority is maintained within the authoritative state, it can be transferred between participants and XR virtual worlds during the collaborative session. Following a handover, the new active operator can continue controlling the same physical system from a different interface while the remaining participants stay synchronized.

Demonstrator

The proposed architecture is instantiated in an industrial hazard-mitigation scenario combining a physical AMR with a SciML surrogate of CFD-based airborne contaminant-dispersion simulations, enabling real-time prediction of the contaminant field. The demonstration involves a contamination-mitigation mission in

which the active operator guides the physical AMR from A to B (Figure 2). The robot and predictive model form the cyber-physical system shared across the connected virtual worlds.

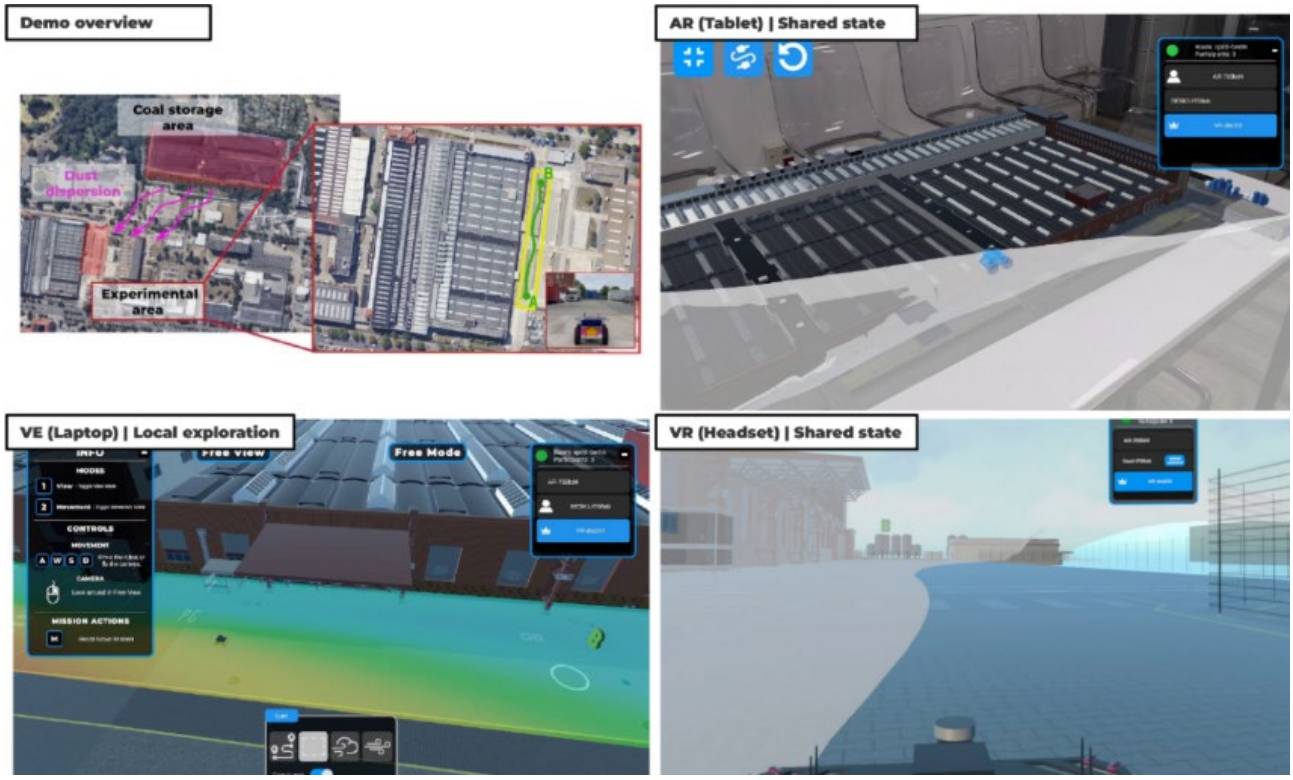


Figure 2. *Demonstrator views across tablet-based AR, desktop VE, and immersive VR interfaces.*

Three participants simultaneously access the digital twin through a tablet-based AR interface, a desktop virtual environment, and a Meta Quest headset. The tablet provides a tabletop representation of the environment for global inspection and planning, the desktop provides a human-scale virtual environment for monitoring and analysis, and the headset provides an immersive spatial perspective for interaction (e.g. from an onboard perspective). This heterogeneity may also broaden access to XR-enabled collaboration by accommodating different levels of technological familiarity, device availability, and user preferences. Despite these different representations, all three interfaces remain connected to the same authoritative state and visualize the physical asset, operational information, and Physics AI predictions according to the capabilities of each virtual world.

During the demonstration, one participant acts as the active operator and remotely commands the AMR while also being able to modify the operational inputs of the SciML model. Changes in the robot state or operational model inputs update the authoritative state, and the resulting state and predictions are propagated across the connected virtual worlds. The other participants can simultaneously monitor the evolving operation from their respective interfaces.

A participant can initiate local state exploration to investigate an alternative Physics AI scenario. For example, the participant can modify SciML inputs locally and visualize the resulting alternative contaminant-dispersion prediction without modifying the authoritative state. Meanwhile, the live operation and its current prediction remain unchanged for the other participants. The participant can subsequently return to the shared state, using the authoritative digital twin as the common reference for continued collaboration.

Finally, the demonstration includes operational authority handover between virtual worlds. Control of the physical AMR and the operational SciML inputs can be transferred from the current active operator to another participant during the ongoing collaborative session. Following the handover, the new operator continues controlling the same physical system from a different XR interface, while all participants remain synchronized.

End-to-end latency across the three XR clients, including Physics AI prediction and visualization, remained below 300 ms, supporting responsive synchronization throughout the collaborative session. A systematic performance evaluation is reserved for future work, including detailed latency characterization, synchronization performance, scalability with larger numbers of participants, and the impact of heterogeneous devices and network conditions.

Conclusions

This work extends our previous physics-informed XR Telepresence approach from a single remote operator to a collaborative digital twin shared across heterogeneous XR virtual worlds. The proposed architecture maintains a single authoritative state across participants while supporting modality-specific representations and interactions, local Physics AI exploration without affecting ongoing operation, and transferable operational authority between users and interfaces. The demonstrator combines an AMR, real-time SciML predictions of airborne contaminant dispersion, and simultaneous tablet-based AR, desktop VE, and immersive VR interfaces, illustrating how heterogeneous representations, shared physical operation, predictive simulation, and independent what-if exploration can coexist around the authoritative state. More broadly, this approach could enable new forms of human-centric collaboration across complex cyber-physical infrastructures in Industry 5.0 and Smart Cities.

Acknowledgements

The authors gratefully acknowledge the T-Systems team for providing the 3D textured model of the Deutsche Telekom 5G Testbed outdoor environment at Siemensstadt Square.

References

Cicirello, A. (2024). Physics-Enhanced Machine Learning: A position paper for dynamical systems investigations. arXiv preprint arXiv:2405.05987. https://doi.org/10.1088/1742-6596/2909/1/012034

Dewberry, N. K., AlHmoud, I., Benton, K., Suarez, D., Chen, Y.-P., Karkaria, V., Tsai, Y.-K., Brock, M., Alazzawi, N., Chowdhury, S., Chen, W., Cao, J. & Gokaraju, B. (2025). A real-time VR-enabled digital twin framework for multi-user interaction in Industry 4.0. *Manufacturing Letters*, 44, 1486–1497. https://doi.org/10.1016/j.mfglet.2025.06.168

Peña-Monferrer, C. & Díaz-Marín, C. (2025). Physics-Informed XR Telepresence: Real-Time ML-Driven Scientific Simulations at 5G Edge for Human-Machine Interaction. *Poster and demo presented at EuroXR 2025*, SPIRIT Track, Winterthur, Switzerland, 3–5 September 2025.

Tu, X., Autiosalo, J., Ala-Laurinaho, R., Yang, C., Salminen, P. & Tammi, K. (2023). TwinXR: Method for using digital twin descriptions in industrial eXtended reality applications. *Frontiers in Virtual Reality*, 4, 1019080. https://doi.org/10.3389/frvir.2023.1019080

Immersive FVV Live Platform: enabling Multiuser Cultural Experiences with AI and Volumetric Video

Javier Usón, Isabel Rodríguez, Celia Ibañez, Carlos Cortés, César Díaz, and Julián Cabrera

Grupo de Tratamiento de Imágenes, Information Processing and Telecommunications Center (IPTC), ETSIT Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

Corresponding author: j.usonp@upm.es

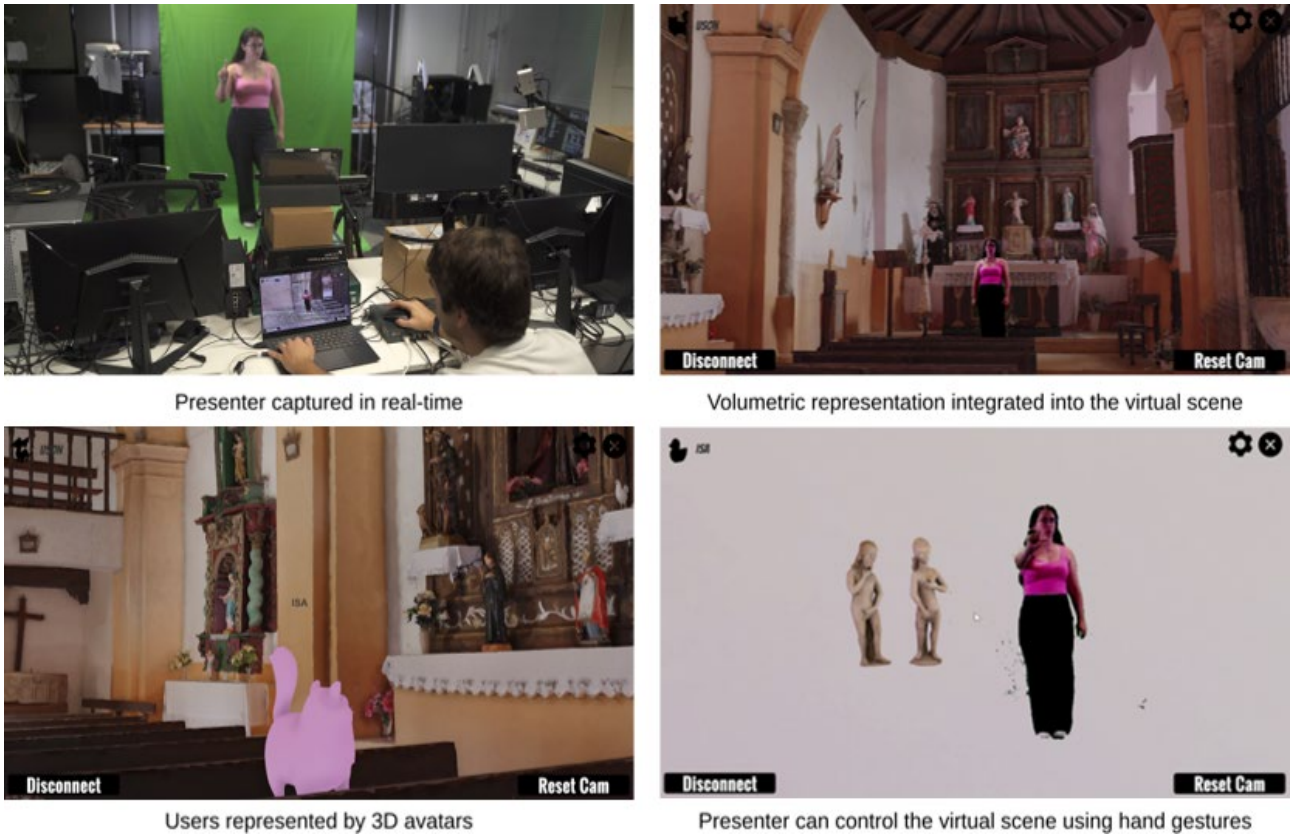
Keywords: Volumetric video, immersive technologies, cultural heritage, real-time systems, 6DoF

Abstract

Extended Reality (XR) has the potential to reshape the way users access cultural heritage, allowing them to explore virtual reconstructions of places and objects. Multiuser XR has received increasing attention as a means of introducing social interaction into cultural experiences, with studies showing that shared experiences can strengthen users' sense of presence and engagement (Skowronek et al., 2022; Yousefdef & Oyelere, 2024; Freeman et al., 2022). Volumetric video provides a promising means of incorporating realistic representations of people, making these interactions more natural and supporting collaboration while reinforcing feelings of presence and embodiment (Sylaiou & Fidas, 2022).

In this demo, we showcase our Immersive Free-Viewpoint Video (FVV) Live Platform, a platform designed to support the development of real-time multiuser cultural experiences by combining our volumetric video system (FVV Live (Carballeira et al., 2022), virtual scenarios built with Unity, and AI-based hand gesture control. The demo presents a virtual cultural visit to a Romanesque church in Pozancos (Guadalajara, Spain). The experience integrates a real-time volumetric representation of a presenter with a 3D reconstruction of the church and other related objects. Users are provided with bidirectional audio communication with the presenter and with each other, allowing them to ask questions and discuss topics of interest. Since the presenter needs to remain in front of the cameras, their movement range is limited. To overcome this limitation, the volumetric video feed is analyzed by a deep learning model to detect hand gestures, enabling the presenter to seamlessly control aspects of the virtual scene, such as navigating through the environment and showcasing specific objects.

Participants can experience the system in both roles: as presenters, being captured by the volumetric capture setup and integrated into the church scene while controlling the experience through hand gestures; and as users, freely exploring the virtual environment and interacting with the presenter through either a traditional 2D interface using mouse and keyboard controls or a VR interface using a Meta Quest headset. The demo illustrates how the combination of volumetric video, multiuser interaction, and gesture-based control can enable more natural and engaging forms of collaborative access to cultural heritage in XR.



Presenter captured in real-time

Volumetric representation integrated into the virtual scene

Users represented by 3D avatars

Presenter can control the virtual scene using hand gestures

Figure 105. *Capture setup used for the demo, alongside examples of the immersive experiences and of the features provided by the Immersive FVV Live Platform.*

Immersive FVV Live Platform

The demo presents a use case developed with our Immersive FVV Live Platform, combining volumetric video delivery and hand-gesture-based user interaction. Both features are implemented following a remote rendering approach, which offloads the computationally intensive tasks (namely virtual view synthesis and deep-learning-based gesture recognition) to an edge server. This architecture is highly compatible with 5G networks, since they provide high bandwidth and low latency capabilities for real-time volumetric video transmission and smooth user interaction. Furthermore, 5G network infrastructures increasingly incorporate Multi-access Edge Computing (MEC), which brings computational resources closer to end users through powerful edge servers, thereby further reducing latency. Details about the Immersive FVV Live Platform are provided in our previous work [6].

As a summary, the platform provides the following features, illustrated in Figure 1.

- Real-time integration of a presenter captured using FVV Live into a virtual scene populated with topic-related 3D models.
- Multiple simultaneous users, each represented by a 3D avatar.
- Bidirectional audio communication between all participants (users and presenter).
- An AI-based gesture recognition pipeline that processes the FVV Live camera feeds to detect hand gestures from the presenter, allowing the presenter to control several aspects of the virtual experience for all participants.
- A scalable deployment architecture based on Kubernetes with automatic resource allocation, tailored for 5G network environments.

The basic architecture of the platform is presented in Figure 2.

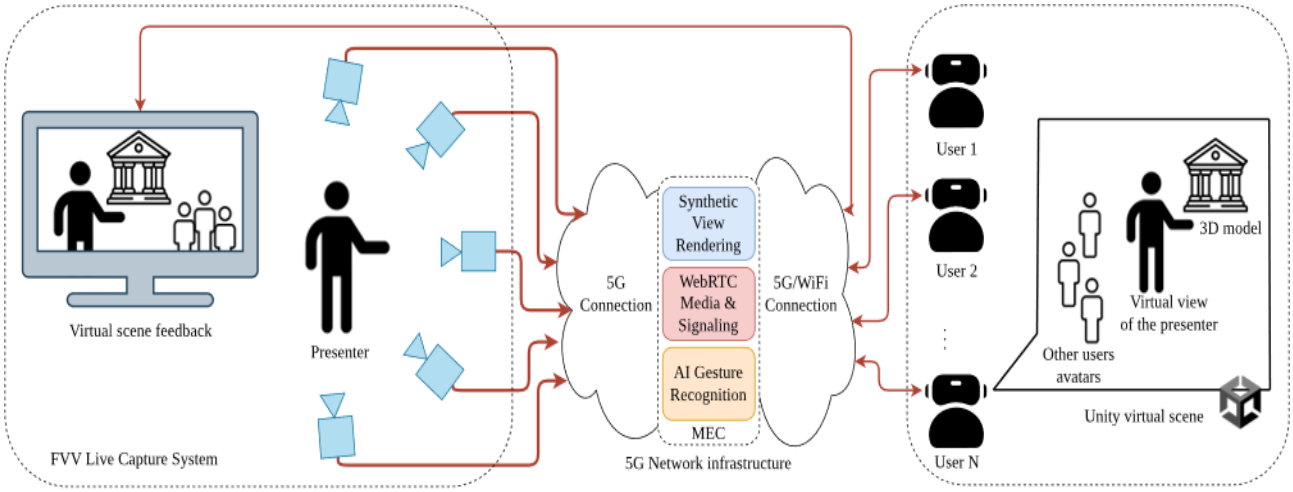


Figure 106. *Architecture diagram of the Immersive FVV Live Platform. A presenter is captured by an array of stereo sensors, and the resulting volumetric data is transmitted to an edge server (MEC), where virtual view rendering and gesture recognition are performed. The rendered views and signaling data are then transmitted to users through WebRTC connections. The presenter receives real-time feedback from the scene, indicating what the users are seeing.*

Demo equipment

The demo includes the following equipment:

- A subset of the FVV Live volumetric capture system, including three Stereolabs ZED 2i cameras and one GPU-equipped server.
- A second GPU-equipped server that will act as the edge server (MEC in Figure 2).
- A network switch and a Wi-Fi router to provide a local area network (LAN) for the transmission to be independent from the venue network.
- A portable green screen setup to facilitate the segmentation of the presenter.
- A laptop to showcase the role of the presenter.
- A Meta Quest 3 headset to showcase the role of the user.

This is an extension to our previous demo showcased in QoMEX 2025 [7], integrating the latest developments of the virtual church visit use case.

Acknowledgments

This work was partially supported by the following projects: PID2023-148922OA-I00 (EEVOCATIONS) funded by MCIN/AEI/10.13039/501100011033 of the Spanish Government, TSI-063000-2021-80 (DISRADIO – Pilotos) funded by the Spanish Government and NextGenerationEU/PRTR, and TEC-2024/COM-322 (IDEALCV-CM) funded by Comunidad de Madrid.

References

Skowronek, J., Raake, A., Berndtsson, G. H., Rummukainen, O. S., Usai, P., Gunkel, S. N., ... & Toet, A. (2022). Quality of experience in telemeetings and videoconferencing: a comprehensive survey. *IEEE Access*, *10*, 63885-63931. https://doi.org/10.1109/ACCESS.2022.3176369.

Yousefdeh S.A,G., & Oyelere, S.S. (2024) Investigating co-presence and collaboration dynamics in realtime virtual reality user interactions. *Front. Virtual Real*. 5:1478481. https://doi.org/10.3389/frvir.2024.1478481

Freeman, G., Acena, D., McNeese, N. J., & Schulenberg, K. (2022). Working together apart through embodiment: Engaging in everyday collaborative activities in social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, *6*(GROUP), 1-25. https://doi.org/10.1145/3492836

Sylaiou S, Fidas C. (2022). Virtual Humans in Museums and Cultural Heritage Sites. *Applied Sciences*. 12(19):9913. https://doi.org/10.3390/app12199913

Carballeira, P., Carmona, C., Díaz, C., Berjón, D., Corregidor, D., Cabrera, J., ... & García, N. (2021). FVV live: A real-time free-viewpoint video system with consumer electronics hardware. *IEEE Transactions on Multimedia*, *24*, 2378-2391. https://doi.org/10.1109/TMM.2021.3079711

Usón, J., Díaz, C., Rodríguez, I., Muñoz, V., Cañellas, F., Montagud, M., & Cabrera, J. (2026). Multiuser immersive cultural experiences enabled by 5G, AI and free-viewpoint video. *Multimedia Systems*, *32*(9), 558. https://doi.org/10.1007/s00530-026-02637-8

Usón, J., Muñoz, V., Cortés, C., Rodríguez, I., Díaz, C., Gutierrez, J., & Cabrera, J. (2025, September). Enabling Real-Time Collaborative Cultural Experiences with Free Viewpoint Video. In *2025 17th International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-4). IEEE. https://doi.org/10.1109/QoMEX65720.2025.11219989

Embodied Safety Moderation in Social VR

Haoran Yun, Arthur Audrain, Katja Zibrek, Ferran Argelaguet

Inria, Rennes, France

Corresponding author: haoran.yun@inria.fr

Keywords: Social virtual reality; harassment; moderation; safety tools; 3D user interfaces

Introduction

Social VR platforms allow users to experience embodied social interactions in real time through virtual avatars, voice, and movement. However, these features can amplify the sense of physical threat when harassment occurs, even though it happens in a virtual world without physical harm (Blackwell et al., 2019). Previous work conducted a virtual field study in AltspaceVR and observed potential harm at 45% of public events, yet moderators addressed only 24% of these incidents (Sabri et al., 2023). This gap is partly due to moderators' limited sight and hearing in VR. In this demo, we address this gap by extending moderators' perceptual reach: miniature previews of multiple rooms at once, with alerts by severity level. Attendees take the role of a moderator, monitoring simulated harassment scenarios across multiple rooms from a single control room and intervening with embodied safety tools.

The demo experience

We present a VR demonstration in which a participant takes the role of a safety moderator. From a virtual control room, the participant surveys miniature social environments, responds to a prioritized incident alert, and enters the selected environment at human scale. There, the participant observes replayed avatars and applies safety tools before returning to the control room. **Demo video:** https://youtu.be/UIzLJhbuYis

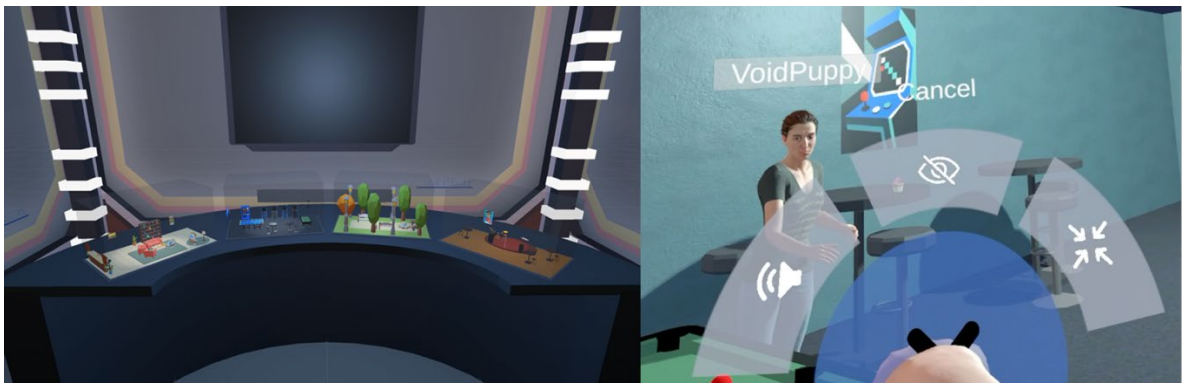


***Figure 107**. Left: Control room with four miniature social environments and a wall screen. Right: Radial safety menu.*

The demonstration is designed for one participant using a Meta Quest 3 with two controllers. A session is intended to last about ten to fifteen minutes, including onboarding and four scenarios. Participants first calibrate their avatar and complete a short tutorial introducing avatar selection, muting, reducing, and hiding. They then enter the control room, where four miniature environments are arranged on a console.

Each miniature presents a social setting with simplified representations of the replayed actors (i.e., colored capsules). Moving props are mirrored in the preview, letting participants observe actions such as carrying or throwing an object. A screen display presents incident information and session progress. An alert dispatcher selects the highest-severity unreviewed room and activates its beacon. Only one room rises at a time, making the currently recommended incident explicit.

When the user points at a miniature, it rises. When the user presses the trigger, the view fades, and the participant is relocated to a predefined position inside the corresponding full-scale environment. The incident replay restarts from the beginning. At this scale, the participant can inspect the avatars' movements and interpersonal distances and hear their speech where the scenario includes dialogue. The user observes the scene as a moderator.

Inside the full-scale room, the user can point at an avatar to open a radial safety menu. The menu includes three actions drawn from a taxonomy of mitigation techniques for social VR (Audrain et al., 2026)**:** *Mute*, *Hide*, and *Reduce*. *Mute* sets the selected avatar's voice volume to zero. *Hide* makes the target avatar invisible while keeping the name tag visible. *Reduce* changes the avatar to 30 percent of its original size and raises its voice pitch. This makes bodily scale an explicit part of the intervention interface. All the actions can be reversed.

Each room has a completion rule (i.e., applying a specified action to a designated actor). Meeting the rule ends the incident after a short delay and returns the participant to the control room. It updates the reviewed-room count and advances the alert queue. These completion rules structure a short demonstration. They should not be interpreted as an objective measure of correct moderation judgment.



***Figure 108**. The four full-scale social environments: social lounge, game room, music room, and park.*

Implementation

The project is implemented in Unity 6000.3.6f1 with Meta XR interaction components, and uses AvatarReady for avatar handling, recording, and replay. The system has two workflows that share the same virtual environments: recording and gameplay.

The recording workflow captures avatar motion and voice from the Meta Quest 3 headset and controllers. A single actor can perform multiple avatars to create a full scenario. It extends the one-

man-crowd approach (Yin et al., 2022) to social VR scenario capture. A custom object-motion recorder captures designated props alongside each avatar take. It preserves actions involving environment manipulation, such as throwing a ball.

The gameplay workflow replays the recorded incidents. Thus, no live actors are needed during the demo. To make the replay feel closer to a live social VR session, the scene responds visibly when the participant applies a moderation action (e.g., “Good job, thank you” and “That didn’t help”). The radial safety menu is adapted from a tool developed by Christos Lougiakis within the META-TOO project.

The virtual environments include a social lounge, a game room, a music room, and a park. Each room has one type of harassment adapted to the environmental context. The four scenarios cover physical, verbal, and environmental harassment. Researchers perform the roles in VR for subsequent replay.

Social lounge: A combination of verbal pressure, spatial intrusion, and object attack. A harasser intrudes on a conversation and persists in close proximity with an unwanted object attack.

Game room: Object attack. A harasser responds to other actors' friendly requests with violent behavior, throwing balls and waving a stick around.

Music room: Verbal harassment. A harasser repeatedly interrupts a birthday performance by shouting, so the performer cannot continue.

Park: Spatial intrusion. A harasser persistently follows another person and intentionally blocks their path.



***Figure 109**. The four harassment scenarios: spatial intrusion and object attack in the social lounge, object attack in the game room, verbal harassment in the music room, and spatial intrusion in the park.*

Research Outlook

Beyond its current demo form, the system is intended to serve as a testbed for studying how interface design shapes moderators' perception of and response to incidents, and as a training environment for new moderators before they handle live events.

Acknowledgements

This work was supported by the European Union's Horizon Europe research and innovation programme through the project META-TOO (grant agreement No. 101160266) and the project SAFE-VR (Marie Skłodowska-Curie grant agreement No. 101209380).

References

Audrain, A., Zibrek, K. & Argelaguet, F. (2026). Taxonomy of Interaction Techniques to Mitigate Inappropriate Social Interactions in Social Virtual Reality. I*EEE Transactions on Visualization and Computer Graphics* 32(5), 3444–3454. https://doi.org/10.1109/TVCG.2026.3679113

Blackwell, L., Ellison, N., Elliott-Deflo, N. & Schwartz, R. (2019). Harassment in Social Virtual Reality: Challenges for Platform Governance. Proceedings of the ACM on Human-Computer Interaction 3(CSCW), Article 100. https://doi.org/10.1145/3359202

Sabri, N., Chen, B., Teoh, A., Dow, S. P., Vaccaro, K. & ElSherief, M. (2023). Challenges of Moderating Social Virtual Reality. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–20. https://doi.org/10.1145/3544548.3581329

Yin, T., Hoyet, L., Christie, M., Cani, M.-P. & Pettré, J. (2022). The One-Man-Crowd: Single User Generation of Crowd Motions Using Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 28(5), 2245–2255. https://doi.org_/10.1109/TVCG.2022.3150507

From Digital Resources to Immersive Learning: The Alliance4XR Platform

Electra Tsaknaki[1], Theodoros Gkiosis[2], Zisis Tsouloukidis[2], Vasilis Petrou[1], Monica Gonzalez Galvez[3], and Konstantinos Smagas[1]

Geoimaging[1], Virtual Solutions[2], EuroXR[3]

Corresponding author: tsaknaki@geoimaging.com.cy

Keywords: eXtended Reality; XR Education; Immersive Learning; Higher Education;

Abstract

eXtended Reality (XR) technologies offer new opportunities for experiential learning, collaborative training and the development of sector-specific digital skills. However, their integration into Higher Education Institutions and Vocational Education and Training remains challenging. Existing solutions can be difficult to configure, require specialised technical knowledge, depend on costly hardware, or provide limited support for the creation and management of educational content.

Previous research has highlighted that immersive learning environments differ considerably in their levels of interactivity, openness and instructional design. This suggests that the successful integration of XR in education depends not only on the technological capabilities of a system, but also on accessibility, usability and the alignment of XR activities with specific learning objectives (Won et al., 2023). In vocational education and training, XR has also been associated with opportunities to support practical skill development and connect theoretical knowledge with simulated work-related activities (Chiang et al., 2022).

These challenges are particularly relevant in engineering-related domains, where learning activities may require three-dimensional objects, realistic environments, collaborative interaction and practical simulations. Alliance4XR addresses these challenges by supporting the integration of XR into education and training through practical methodologies, learning resources and digital tools.

Platform Requirements and Design Approach

The Alliance4XR digital platform[10] was designed following a requirements analysis and benchmarking process. The analysis considered the needs of Higher Education Institutions, Vocational Education and Training centres, and professionals working in structural engineering, construction, grids and energy, and maritime professions.

[10] *https://app.alliance4xr.eu/*

The design process identified several requirements for a usable and adoptable XR learning platform:

1. accessible workflows for educators and learners;
2. support for different types of digital and three-dimensional content;
3. reduced dependence on specialised technical expertise;
4. compatibility with different devices and infrastructure configurations;
5. tools for organising learning materials into structured courses;
6. support for collaborative and immersive learning activities;
7. an intuitive interface for creating, editing and previewing 3D objects.

Based on these requirements, the platform follows a modular, web-based approach that enables users to progressively introduce XR into existing teaching and training practices.

eXtended Reality Platform

The Alliance4XR platform provides an integrated environment for creating, organising, sharing and presenting educational content. Its main functions include:

1. uploading and previewing documents, videos and 3D objects;
2. creating and editing three-dimensional objects;
3. organising materials into courses and lessons;
4. preparing live interactive sessions in an immersive 3D environment;
5. managing files, objects, courses, lessons and sessions through a central dashboard;
6. supporting scheduled activities;
7. accessing content through different devices and interaction configurations.

The platform is not limited to a single immersive use case. Instead, it supports a gradual learning workflow, beginning with conventional digital resources and extending to interactive 3D content and immersive sessions. This allows educators to select the appropriate level of XR integration according to their pedagogical objectives, technical confidence and available equipment.

Usability-Oriented Learning Workflow

Usability is a central aspect of the platform. The system brings content management, 3D creation, learning organisation and immersive session preparation into one coherent workflow.

The dashboard provides a single entry point for managing learning resources and activities. The 3D editor and preview functions allow users to work directly with spatial content, while the course and lesson tools support the organisation of materials into structured educational sequences. Live interactive sessions extend these capabilities by enabling users to prepare and deliver learning activities in an immersive three-dimensional environment.

This integrated workflow aims to reduce the fragmentation that can occur when educators need to combine separate tools for content storage, 3D creation, course management and immersive delivery. It also supports users with different levels of XR experience, from educators who primarily use digital learning materials to users who are ready to incorporate more advanced 3D and immersive activities.

The platform's usability-oriented design is particularly important for education and training providers that may not have dedicated XR development teams. By providing ready-to-use functions and a consistent interface, the platform aims to make XR-supported learning more approachable and easier to incorporate into existing educational practices.

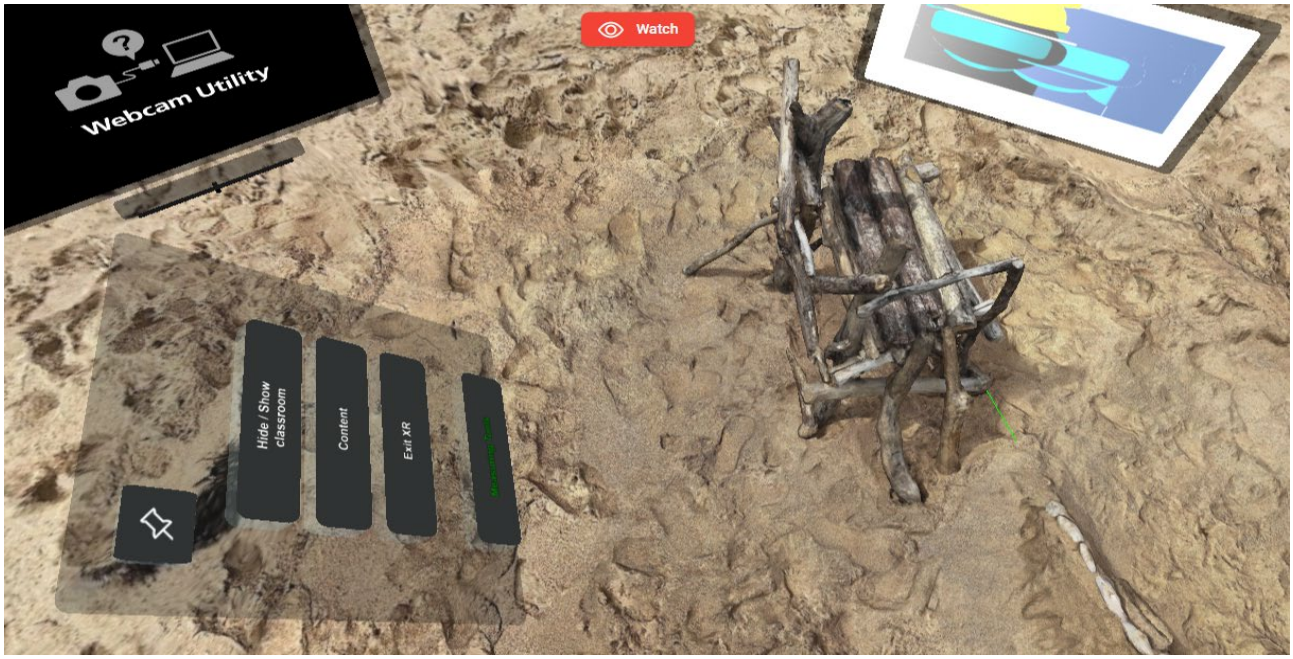


***Figure 110**. The virtual reality platform as viewed from the headset*

Alliance4XR Ambassador Board and Operational Ambassadors

Technical tools alone are not sufficient to support the long-term adoption of XR in education and training. Continued communication, knowledge exchange and community engagement are also necessary. For this reason, Alliance4XR is establishing an Ambassador Programme consisting of a Strategic Ambassador Board and a wider network of Operational Ambassadors.

The Strategic Ambassador Board brings together experienced professionals and experts from the XR, education and innovation ecosystems, including representatives of higher education institutions (HEIs), vocational education and training (VET), industry, XR practice and European policy. Board members provide strategic guidance, validate project results, identify trends, challenges and opportunities, and help connect Alliance4XR with relevant organisations, initiatives and networks.

Operational Ambassadors have a more practical role and support Alliance4XR through three profiles: Adoption Ambassadors, who use and test the project in real learning environments; Integration Ambassadors, who build collaborations and connections with relevant ecosystems; and Visibility Ambassadors, who help communicate the project and its results to wider audiences.

The ambassador model is designed to be simple, flexible and adaptable. It does not impose a minimum communication commitment, allowing participants to choose when and how they contribute, tailor messages to their audiences and local contexts, and create content independently. Ambassador activities focus on four priorities:

The ambassador community will focus on four communication priorities:

1. **Platform and learning resources:** showing how educators and learners can use the platform, tutorials and teaching resources;
2. **Results and use cases:** sharing project results, pilots and examples of XR-supported education and training;
3. **Community and collaboration:** connecting educators, researchers, industry representatives and XR professionals;
4. **Events and opportunities:** promoting workshops, demonstrations, training activities and related opportunities.

The programme provides a structured yet flexible model for linking strategic guidance, real-world adoption, ecosystem integration and dissemination to support the sustainable uptake of XR in education and training.

Conclusion

Alliance4XR provides a practical approach to integrating XR into Higher Education and Vocational Education and Training. Through its web-based platform, users can manage learning resources, create and preview 3D objects, organise courses and lessons, and prepare immersive interactive sessions within one environment.

By combining accessible platform functions with a planned external Ambassador Board and Operational Ambassador community, Alliance4XR addresses both the technical and organisational dimensions of XR adoption. The resulting ecosystem supports educators, learners, researchers, industry professionals and other stakeholders in discovering, using and continuing to develop XR-supported learning practices.

References

Chiang, F. K., Shang, X., & Qiao, L. (2022). Augmented reality in vocational training: A systematic review of research and applications. *Computers in Human Behavior*, 129, 107125. https://doi.org/10.1016/j.chb.2021.107125

Won, M., Ungu, D. A. K., Matovu, H., Treagust, D. F., Tsai, C. C., Park, J., ... & Tasker, R. (2023). Diverse approaches to learning with immersive Virtual Reality identified from a systematic review. *Computers & Education*, *195*, 104701. https://doi.org/10.1016/j.compedu.2022.104701

Virtual Embodiment in Healthy Participants and Individuals with Spinal Cord Injury: Preliminary Evidence from a Multimodal VR Protocol

Giorgia Malerba[1,2], Emilia Biffi[1], Chiara Barà[4], Roberta Nossa[1], Fabio Alexander Storm[1], Riccardo Riboni[1], Mattia Chiappini[1,5], Valentina Grasso[1], Marco Sacco[3], Giovanna Rizzo[4], Alfonso Mastropietro[4], and Sara Arlati[3]

[1] *Scientific Institute, IRCCS Eugenio Medea, Bosisio Parini (LC), Italy*

[2] *Department of Electronics Information and Bioengineering, Politecnico di Milano, Milan, Italy*

[3] *Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council, Lecco, Italy*

[4] *Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council, Milan, Italy*

[5] *Department of Psychology, Università degli Studi di Torino, Turin, Italy*

Corresponding author: Giorgia.malerba@lanostrafamiglia.it

Keywords: Virtual embodiment; Spinal cord injury; Electrodermal activity; Heart rate variability.

Background

Virtual reality (VR) is increasingly used in rehabilitation (Bargeri et al. 2023; Wang et al. 2024) because it allows users to interact with controlled and ecologically meaningful environments while receiving multisensory feedback. An important mechanism underlying the effectiveness of VR is the embodiment (Makransky and Petersen 2021), namely the experience of owning, controlling and being located in a virtual body. Embodiment is commonly described as a multidimensional construct including body ownership, self-location and sense of agency (Kilteni, Groten, and Slater 2012) and can be influenced by visuo-motor and visuo-tactile synchrony, action congruence, interoceptive feedback, and avatar-related characteristics (Fabritius, Stein, and Chuang 2025; Pratviel et al. 2022).

Most studies, however, investigate individual embodiment manipulations separately and predominantly in healthy populations. Less is known about whether the same cues are equally effective in people with neurological conditions, in whom body representation, somatosensory information, and motor experience may be altered (Fabritius et al. 2025).

We previously developed a three-day multimodal VR protocol to systematically manipulate the sense of embodiment in healthy participants and clinical populations (Malerba et al. 2026). The platform combines the Virtual Embodiment Questionnaire (VEQ) with electrodermal activity (EDA), electrocardiogram (ECG), skin temperature, electroencephalography (EEG) in selected conditions, and behavioral assessment.

The present work reports preliminary results from the ongoing study, focusing on healthy participants (HLT) and participants with spinal cord injury (SCI).

Methods

The experimental protocol comprised three sessions (Malerba et al. 2026). Day 1 compared four multisensory conditions: visuo-tactile stimulation applied to areas with impaired/absent (Altered) or preserved (Perceivable) sensitivity, cardio-visual heartbeat feedback (HR), and active avatar hand control (Hand Control). Day 2 manipulated motor control and visuo-motor congruence during arm and leg cycling with a cycle ergometer, including VR-only movement, active and passive matched cycling (with passive movement imposed by the ergometer), and an arm-to-leg mismatch in which participants cycled with their arms while observing the avatar cycling with its legs. EEG was also recorded during Day 2 conditions, but its analysis is still ongoing and the results are therefore not included in the present preliminary report. Day 3 compared active and passive walking and wheelchair locomotion, with and without cardio-visual HR feedback. Embodiment was assessed after each condition using the VEQ, while EDA, skin temperature, and ECG were continuously recorded.

VEQ Ownership, Agency, and Change scores were computed as the mean of the corresponding items according to (Roth and Latoschik 2020).

EDA was processed using a custom Python pipeline including filtering, down-sampling, artifact removal, and cvxEDA decomposition (Greco et al. 2016); tonic and phasic features were extracted and expressed as absolute changes from baseline. ECG was filtered and R-peaks were detected using NeuroKit2, with visual correction when required; HRV features were then extracted and normalized relative to baseline (Costantini, Storm, and Bianchi 2026).

Statistical analyses were performed in MATLAB 2024b. Day 1 within-group effects were assessed using Friedman tests followed, when significant, by Bonferroni-corrected Wilcoxon post-hoc comparisons. For Days 2 and 3, a priori Wilcoxon signed-rank comparisons tested the main experimental contrasts. On Day 2, comparisons tested the effects of active versus passive motor execution, VR-only versus real movement and arm-to-leg mismatch. On Day 3, contrasts tested active versus passive control, walking versus wheelchair locomotion, and the effect of cardio-visual biofeedback. Differences between HLT and SCI were assessed using Mann–Whitney U tests for each experimental condition. Spearman rank correlations were used to investigate associations between VEQ components and physiological features. Statistical significance was set at $p = 0.05$ for all analyses.

Preliminary Results

Twenty-six participants were enrolled: 14 HLT (age 29.8 [7.2] years; 7 females), 10 SCI (29.8 [13.4]; 3 females), and 2 participants with post-stroke sequelae (STK) (17.7 [1.9]; 1 female). Inferential analyses were restricted to HLT and SCI because of the limited STK sample.

The main within-group findings across the three experimental days were:

- Day 1: In HLT, Hand Control produced higher Agency than Altered ($p=.004$), HR ($p<.001$), and Perceivable ($p=.003$), and higher Ownership than HR ($p=.015$). It was also associated with significantly greater phasic EDA responses and lower average normal-to-normal interval (AVNN). In SCI, no significant VEQ differences emerged.
- Day 2: In HLT, active compared with passive cycling increased Agency for both arms ($p=.009$) and legs ($p=.002$). Matched active-arm movement also produced higher Agency ($p=.003$) and Ownership ($p<.001$) than the arm-to-leg mismatch. Phasic EDA and AVNN also varied consistently across motor conditions, with lower AVNN during active than passive cycling. In SCI, Agency increased from VR-only to both active and passive arm cycling (both $p=.008$), and AVNN was lower during active than passive arm cycling ($p=.031$). Unlike HLT, no significant Agency or Ownership reduction emerged for the arm-to-leg mismatch.
- Day 3: In HLT, Agency was higher during active than passive walking and wheelchair locomotion (both $p<.001$), while Ownership was higher during active than passive walking

(p=.001). Phasic EDA showed condition-dependent changes that were generally consistent with the VEQ pattern across locomotor conditions. In SCI, Agency was higher during active than passive wheelchair locomotion (p=.031), while AVNN differed between passive wheelchair conditions with and without HR feedback (p=.016). Overall, HR biofeedback did not consistently increase VEQ scores.

Significant HLT–SCI differences were condition-specific:

- VEQ: HLT showed higher Ownership during Day 1 Altered (p=.022), Day 2 active arm cycling (p=.024), and Day 3 active walking (p=.042), and higher Agency during Day 1 Hand Control (p=.028) than SCI.
- Physiology: HLT showed greater phasic EDA responses during Day 1 Hand Control and Day 3 active walking (all p≤.028) than SCI.

Finally, Agency and Ownership were positively correlated with phasic EDA peak number and amplitude (ρ=.20–.29, all p<.001), while AVNN was negatively correlated with Agency (ρ=−.145, p=.003) and Ownership (ρ=−.101, p=.037).

Discussion and Conclusion

Preliminary findings identify active sensorimotor control as the most consistent facilitator of embodiment, particularly Agency, across the three experimental sessions. In HLT, Hand Control on Day 1, active cycling on Day 2, and active locomotion on Day 3 consistently increased Agency. Similar effects were observed in SCI, particularly for active upper-limb control and wheelchair locomotion. Interestingly, the Perceivable condition produced comparatively high Agency and Ownership in SCI, suggesting that multisensory stimulation may support embodiment even in the absence of direct fine-motor control.

Subjective changes were accompanied by coherent autonomic modulation. Phasic EDA emerged as the physiological measure most consistently associated with embodiment: the number and amplitude of phasic responses correlated positively with both Agency and Ownership, and their modulation across experimental conditions generally paralleled changes in VEQ scores. ECG provided complementary evidence. AVNN, the mean interval between consecutive normal heartbeats, was negatively associated with both VEQ components and tended to decrease in conditions showing stronger Agency, indicating a higher heart rate with greater subjective embodiment. Thus, both EDA and ECG showed physiological changes consistent with the subjective embodiment response.

Differences between HLT and SCI were strongly dependent on the body segment and sensorimotor context involved. Notably, significant group differences emerged in scenarios requiring direct control or accurate virtual reproduction of the upper limbs: during Hand Control, HLT showed greater Agency and a larger phasic EDA response than SCI, and during active arm cycling HLT also showed higher Ownership. This pattern may indicate that SCI participants are more sensitive to inaccuracies in the virtual reproduction of well-preserved body parts, where even small discrepancies between real and virtual upper-limb movements may become more salient and reduce embodiment. A similar pattern emerged on Day 3. During active walking, HLT showed higher Ownership and greater phasic EDA responses than SCI, whereas these group differences were not evident during active wheelchair locomotion. Participants performed the task while physically seated; therefore, wheelchair locomotion provided a virtual representation more compatible with their actual seated posture. These findings suggest that, in SCI participants, embodiment may be influenced not only by avatar control but also by the similarity between virtual movement and their real sensorimotor experience. Finally, cardio-visual heartbeat biofeedback produced little consistent enhancement of Agency or Ownership, suggesting a more limited influence on conscious embodiment than direct motor control and visuo-motor congruence within the present protocol. These preliminary results

highlight the importance of tailoring embodied VR not only to users' motor capabilities but also to the sensorimotor information they can reliably perceive. Ongoing recruitment will allow these observations to be tested in larger clinical samples.

Acknowledgments

This work was funded by Inail (PR23-SV-P2 Rientr@Returns - Consolidamento tecnologico e ampliamento della destinazione d'uso della piattaforma di simulazione Rientr@) and partially by the Italian Ministry of Health (Ricerca Corrente 2025/2026 to E. Biffi)

References

Bargeri, S., Scalea, S., Agosta, F., Banfi, G., Corbetta, D., Filippi, M., ... & Gianola, S. (2023). Effectiveness and safety of virtual reality rehabilitation after stroke: an overview of systematic reviews. *EClinicalMedicine*, *64*.

Costantini, S., Storm, F. A., & Bianchi, A. M. (2026). Benchmarking ECG R-peak Detectors in Offline and Real-time Settings across Heterogeneous Datasets. https://doi.org/10.15203/99106-194-6-07

Fabritius, P.S., Stein, J.P., & Chuang, L. L. (2025). A Systematic Review of the Impact of a Standardized Embodiment on Virtual Reality Research. MuC '25: *Proceedings of the Mensch Und Computer* 2025 10:232–41. https://doi.org/10.1145/3743049.3743077 .

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2015). cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE transactions on biomedical engineering*, *63*(4), 797-804. https://doi.org/10.1109/TBME.2015.2474131 .

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 373-387. https://doi.org/10.1162/PRES_a_00124

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational psychology review*, *33*(3), 937-958. https://doi.org/10.1007/s10648-020-09586-2 .

Malerba, G., Biffi, E., Barà, C., Nossa, R., et al. (2026). Design and Development of a Multimodal Virtual Reality Protocol to Investigate Embodiment-Enhancing Techniques in Healthy Participants and Clinical Populations. In *Extended Reality*, edited by L. T. De Paolis, P. Arpaia, and M. Sacco. Cham: Springer Nature Switzerland.

Pratviel, Y., Bouni, A., Deschodt-Arsac, V., Larrue, F., & Arsac, L. M. (2022). Avatar embodiment in VR: Are there individual susceptibilities to visuo-tactile or cardio-visual stimulations?. *Frontiers in Virtual Reality*, *3*, 954808. https://doi.org/10.3389/frvir.2022.954808 .

Roth, D., & Latoschik, M. E. (2020). Construction of the virtual embodiment questionnaire (veq). *IEEE transactions on visualization and computer graphics*, *26*(12), 3546-3556. https://doi.org/10.1109/TVCG.2020.3023603 .

Wang, L., Zhang, H., Ai, H., & Liu, Y. (2024). Effects of virtual reality rehabilitation after spinal cord injury: a systematic review and meta-analysis. *Journal of NeuroEngineering and Rehabilitation*, *21*(1), 191. https://doi.org/10.1186/s12984-024-01492-w.

Experiencing the Binaural Rendering Toolbox: An Interactive Demonstration of 3D Binaural Audio for XR

Daniel González-Toledo, María Cuevas-Rodríguez, Pablo Gutierrez-Parera and Arcadio Reyes-Lecuona

Telecommunication Research Institute (TELMA), Universidad de Málaga, Málaga, Spain

Corresponding author: dgonzalezt@uma.es

Keywords: Binaural audio; spatial audio; virtual reality; augmented reality; psychoacoustics; auralisation; HRTF; room acoustics

Introduction

Spatial audio is an essential component of immersive virtual and augmented reality experiences. In addition to increasing realism and presence, auditory spatialisation can convey information about the position, distance and movement of objects outside the user's current field of view. To produce a coherent multisensory scene, the rendered audio must respond in real time to changes in the positions and orientations of the listener, sound sources and virtual environment.

Binaural rendering creates the signals to be reproduced at the listener's ears, normally through headphones. The transformation introduced by the listener's head, torso and outer ears is described by a Head-Related Transfer Function (HRTF). Applying the appropriate pair of HRTFs to a sound signal can reproduce the acoustic cues associated with a particular direction and therefore generate the perception that the sound originates from a location outside the listener's head. HRTFs vary considerably between individuals, meaning that the selected dataset can affect localisation accuracy, externalisation and the overall plausibility of the resulting auditory scene.

Although binaural rendering is widely used in XR, experimental and application-oriented developments often require researchers to integrate several heterogeneous components: HRTF processing, interpolation between measured directions, distance effects, room-acoustic simulation, source directivity and real-time scene control. Comparing algorithms across separate tools and incompatible data formats can also be difficult.

The Binaural Rendering Toolbox (BRT) addresses these requirements through an open and extensible framework for real-time binaural audio rendering (González-Toledo et al., 2023). It was developed in the framework of the SONICOM[11] project (Picinali et al., 2022) and the SONIX[12] project as an evolution from the 3D Tune-In Toolkit (Cuevas-Rodríguez et al., 2019) and was originally designed as a virtual laboratory for spatial psychoacoustic experimentation. Its development is currently being continued within the framework of the SENSORAMA project[13]. In this context, a

[11] https://www.sonicom.eu
[12] https://www.diana.uma.es/sonix/
[13] https://cordis.europa.eu/project/id/101297303

virtual laboratory is understood as a configurable rendering environment in which researchers can construct auditory scenes, select the models and datasets used in each part of the processing chain, and control experimental variables in real time. The same characteristics make BRT suitable as the audio engine of VR and AR experiments or applications, where it can operate in synchrony with an independently rendered visual scene.

The demonstration presented in this extended abstract allows EuroXR participants to experience the main perceptual effects produced by different BRT models and to understand how these models can be combined when designing interactive immersive applications.

The Binaural Rendering Toolbox

The BRT library is open-source and distributed under the GNU General Public License v3.0[14] and it is organised around a modular processing architecture comprising source, environment and listener models, followed by an optional bilateral filtering stage. Rather than implementing a single predefined spatialisation algorithm, the framework allows the rendering chain to be assembled from models with complementary functions.

Source models describe the properties and behaviour of the sound emitter. They receive real-time information about its position and orientation and may represent either an omnidirectional or directional source. In the latter case, the signal changes according to the orientation of the source relative to the listener, reproducing the fact that real sound emitters do not generally radiate energy equally in every direction. Environment models represent sound propagation between sources and listeners. In a free-field configuration, only the direct propagation path is considered. Alternative models introduce reflections and reverberation through measured impulse responses or room-acoustic algorithms. Because the environment is represented independently of the source and listener models, different room-rendering techniques can be selected without redesigning the complete application. Listener models generate the binaural signals corresponding to the sound arriving at the listener. The direct sound can be spatialised using HRTFs, while room effects can be incorporated through Binaural Room Impulse Responses (BRIRs) or spatial room impulse responses. Listener position and orientation can be updated continuously, allowing the auditory scene to respond to head or body movements reported by an XR tracking system.

The BRT Application (BeRTA) is a standalone application built upon the BRT Library, providing a graphical interface to its binaural rendering capabilities. As shown in Figure 1, BeRTA allows users to configure and monitor the audio scene, including sound sources, environment and listener. The application supports the Spatially Oriented Format for Acoustics (SOFA), an Audio Engineering Society standard for storing spatial acoustic data such as HRTFs, BRIRs and source directivity measurements (Majdak et al., 2013). SOFA provides a common description of the measured signals, spatial coordinates and associated metadata. Its use facilitates the exchange of datasets and allows the same rendering framework to operate with measurements obtained from different listeners, rooms or sound sources. BeRTA can be controlled in real time through Open Sound Control (OSC) messages. This separation is useful in experimental XR systems, allowing audio and visual rendering to operate independently, even on separate computers, while exchanging real-time scene updates.

[14] https://github.com/GrupoDiana/BRTLibrary/releases



Figure 111. *BeRTA user interface. The left-hand panel provides model configuration and rendering information, while the right-hand panel displays and controls the spatial arrangement of the listener and sound sources.*

Interactive Demonstration

The demonstration uses short listening examples to illustrate perceptually relevant BRT functionalities. The first part demonstrates three-dimensional source positioning. A sound source moves around the participant in azimuth and elevation, including positions above and below the horizontal plane. Further examples vary the source distance, allowing participants to hear transitions between distant and nearby positions, where the rendering incorporates near-field effects. A further example introduces source directivity. As the virtual emitter rotates, the received spectrum and level change according to its directional radiation pattern. This capability is particularly relevant for the auralisation of musical instruments, loudspeakers, human talkers and other emitters whose orientation contributes to the plausibility of an XR scene. The demonstration also compares several BRT room-acoustics rendering approaches:

- **Reverberant Virtual Loudspeakers (RVL):** the source signals are encoded into an Ambisonics representation and decoded through a set of virtual loudspeakers characterised by BRIRs (Engel et al., 2021).
- **Hybrid ISM–RVL rendering:** perceptually relevant early reflections are generated using the Image Source Method (ISM) (Gutierrez-Parera et al., 2026), while the later reverberant field is reproduced using RVL.
- **Scattering Delay Networks (SDN):** room reflections and reverberation are generated through a delay-network structure associated with the geometry and surfaces of the simulated enclosure (De Sena et al., 2015).
- **Long impulse-response convolution:** this example demonstrates the capacity of the framework to process the long responses associated with highly reverberant spaces, such as churches or large halls.

The demonstration allows participants to experience how BRT operates and to appreciate the range of spatial and acoustic cues that it can reproduce within a common real-time framework. These capabilities make BRT particularly suitable for use in the SENSORAMA project, where it can provide the binaural audio component of multimodal XR experiences. Its real-time control mechanisms allow the auditory scene to be synchronised with the visual and haptic components of the virtual environment, while its modular architecture makes it possible to select and combine the source, listener and room-acoustic models best suited to each experimental scenario.

Acknowledgements

This research has been funded by European Union's Horizon 2020 (SONICOM project under grant 101017743), and the Spanish National Project SONIX (PID2023-152547NB-I00). This demo has also been supported by the SENSORAMA project, funded by the European Union's Horizon Europe under grant agreement 101297303. Additional support from the EIDIA program (DGP_POST_2024_00447) Junta de Andalucía and ESF+.

Disclaimer**:** Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

References

Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., & Reyes-Lecuona, A. (2019). 3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation," *PLOS ONE*, vol. 14, no. 3, e0211899. https://doi.org/10.1371/journal.pone.0211899

De Sena, E., Hacıhabiboğlu, H., Cvetković, Z. & Smith, J. O. (2015). Efficient synthesis of room acoustics via scattering delay networks, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1478–1492. https://doi.org/10.1109/TASLP.2015.2438547

Engel, I., Henry, C., Amengual Garí, S. V., Robinson, P. W. & Picinali, L. (2021). Perceptual implications of different Ambisonics-based methods for binaural reverberation. *The Journal of the Acoustical Society of America*, vol. 149, pp. 895–910. https://doi.org/10.1121/10.0003437

González-Toledo, D., Molina-Tanco, L., Cuevas-Rodríguez, M., Majdak, P. & Reyes-Lecuona, A. (2023). The Binaural Rendering Toolbox: A virtual laboratory for reproducible research in psychoacoustics. In *Proceedings of the 10th Convention of the European Acoustics Association, Forum Acusticum* 2023, Turin, Italy. https://doi.org/10.61782/fa.2023.1042

Gutierrez-Parera, P., Arrebola, F., Cuevas-Rodriguez, M., Gonzalez-Toledo, D., Molina-Tanco, L., & Reyes-Lecuona, A. (2026). Hybrid ISM+BRIR convolution reverberation acoustic model with tuning of absorption coefficients based on single-position measurements. *Applied Acoustics*, 254(December), 111448. https://doi.org/10.1016/j.apacoust.2026.111448

Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H. & Noisternig M. (2013). Spatially Oriented Format for Acoustics: A data exchange format representing head-related transfer functions. In *Proceedings of the 134th AES Convention*, Rome, Italy.

Picinali, L., Katz, B. F., Geronazzo, M., Majdak, P., Reyes-Lecuona, A., & Vinciarelli, A. (2022). The SONICOM Project: Artificial Intelligence-Driven Immersive Audio, From Personalization to Modeling [Applications Corner]. *IEEE Signal Processing Magazine*, 39(6), 85–88. https://doi.org/10.1109/MSP.2022.3182929

Towards Dynamic Auralisation of Historic Church Organs in Virtual Environments

Pablo Gutierrez-Parera, Daniel Gonzalez-Toledo, Maria Cuevas-Rodriguez, and Arcadio Reyes-Lecuona

Telecommunication Research Institute (TELMA), Universidad de Málaga, Málaga, Spain

Corresponding author: areyes@uma.es

Keywords: Binaural rendering, 6DoF audio, church organ, auralisation, Ambisonics, acoustic simulation, virtual reality

Introduction

The MusicSphere project is developing digital technologies for documenting, preserving and experiencing historic musical instruments and their acoustic environments. One of its use cases focuses on church organs, whose sound is closely linked to the acoustics of the buildings in which they are installed. Virtual Reality offers a particularly attractive way of presenting these instruments, since users can explore both the organ and its architectural context while listening to a spatial reproduction of the sound.

However, a fully navigable experience requires the acoustics to change as the listener moves through the virtual environment. While head rotation can be handled relatively naturally with spatial audio techniques, listener translation introduces a more difficult problem: the propagation paths between the organ, the room and the listener change continuously with position.

This work presents an ongoing prototype for dynamic binaural auralisation of church organs in 6DoF virtual environments. The Binaural Rendering Toolbox (BRT) (Gonzalez-Toledo et al. 2025) is used as the underlying auralisation framework, providing the basis for the new developments required to implement the approaches explained in this work. Two alternative strategies are being investigated, each addressing the spatial variation of room acoustics in a different way.

Background and Audio Rendering Strategies

The first strategy relies on a spatial grid of Ambisonic room impulse responses. At each position, an Ambisonic response already provides a directional description of the sound field and therefore supports 3DoF interaction, allowing the user to rotate their head while preserving the apparent direction of the acoustic scene. To support 6DoF navigation, however, the renderer must move from one spatial response to another as the listener changes position.

The second strategy uses a hybrid acoustic model. Instead of storing complete acoustic responses at many points in the room, the direct sound and early reflections are generated dynamically from the scene geometry, while a spatially static reverberant component is reproduced by convolution.

The two approaches therefore distribute the complexity differently. The first relies mainly on spatial sampling and interpolation, whereas the second relies on the real-time update of sound propagation paths.

Interpolated Ambisonic Impulse-Response Approach

In the first approach, Ambisonic room impulse responses are associated with different positions within the navigable area. During user movement, the renderer must progressively move between these responses.

A direct switch between two neighbouring responses may produce audible discontinuities in timbre, reverberation or spatial impression. One possible solution is therefore to use a sufficiently dense grid so that consecutive responses differ only slightly (De Muynke et al. 2024). However, acquiring such a dense set of responses through measurement or detailed acoustic simulation may be impractical.

An alternative is to interpolate between responses obtained at more widely separated positions. This is not a trivial problem, since an impulse response contains temporal, spectral and directional information at the same time. Reflections may change their arrival time, amplitude and direction as the listener moves, making simple interpolation techniques potentially unsuitable.

More advanced signal-processing or data-driven methods, including machine-learning approaches, could be explored to estimate intermediate responses (Politis et al. 2023; Das et al. 2025). Importantly, this interpolation does not necessarily need to be performed in real time. A sparse set of measured or simulated responses could be processed offline to generate a much denser grid of intermediate responses. The VR application could then operate with a comparatively simple real-time strategy, selecting or switching between closely spaced responses during navigation.

Hybrid ISM+Room Impulse Response Approach

The second approach reduces the need for dense spatial sampling by dynamically modelling the part of the room response that is most strongly affected by listener movement. The direct sound and early reflections are generated using geometrical acoustics, while the later reverberant component is reproduced through a spatially static convolution (Gutierrez-Parera et al. 2026).

From the point of view of an interactive VR application, this allows the most relevant propagation paths to react continuously to user movement without requiring a complete precomputed response at every possible position. However, it also introduces a different type of artefact.

Each geometrical reflection can be understood as a delayed copy of the source signal. As the listener moves, the propagation distance changes, and therefore the corresponding delay must also vary continuously. If these time-varying delays are not handled carefully, they can introduce subtle modulation, phase or spectral effects (De Sena et al. 2015).

These artefacts may be difficult to detect with many common signals. Speech, for example, is continuously changing in pitch, spectral content and amplitude, which can mask small variations produced by the rendering process. Pipe-organ sounds are potentially much more sensitive. Sustained organ notes can remain highly stable for several seconds and contain strong, well-defined harmonic components. Small variations in reflection delay can therefore modify the phase relationship between direct and reflected sound, producing audible beating, modulation or changes in spectral coloration. The broad frequency range of the instrument and the long reverberation times typical of churches may further increase the relevance of these effects.

Conclusions and Future Work

The current prototype is intended as a common framework for implementing and comparing these two strategies for 6DoF church-organ auralisation. The work is still in progress and no conclusions can yet be drawn regarding their relative perceptual performance.

The next stages will investigate different spatial densities and interpolation strategies for the Ambisonic-response approach, including offline generation of intermediate responses. For the hybrid approach, attention will be given to the update of time-varying reflection delays and to the combination of dynamically generated early reflections with static reverberation. A subsequent perceptual evaluation will examine which artefacts are relevant during realistic user navigation and under which conditions each approach offers the most suitable compromise between acoustic quality, computational cost and real-time performance.

Acknowledgements

This work has received funding from the European Union under the Horizon Europe programme through the MusicSphere[15] project (Grant Agreement No. 101233618).[16] Additional support was provided by the EIDIA postdoctoral program (grant No. DGP_POST_2024_00447) funded by Junta de Andalucía and the ESF+.

References

Das, O., Santo, G. D., Schlecht, S. J., & Cvetković, Z. (2025). Neural-network based interpolation of late reverberation in coupled spaces using the common slopes model. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. https://doi.org/10.1109/WASPAA66052.2025.11230987

De Muynke, J., Poirier-Quinot, D., & Katz, B. F. G. (2024). Effect of Interpolation Artifacts on Perceived Stability of Nearby Sources in a Navigable Reverberant Virtual Environment. *Journal of the Audio Engineering Society*, 72(10), 664–678. https://doi.org/10.17743/jaes.2022.0158

De Sena, E., Antonello, N., Moonen, M., & van Waterschoot, T. (2015). On the Modeling of Rectangular Geometries in Room Acoustic Simulations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(4), 774–786. https://doi.org/10.1109/TASLP.2015.2405476

Gonzalez-Toledo, D., Cuevas-Rodriguez, M., Molina-Tanco, L., Picinali, L., & Reyes-Lecuona, A. (2025). New Release of the Binaural Rendering Toolbox: Introducing Version 2.0. *Proceedings of the 11th Convention of the European Acoustics Association Forum Acusticum / EuroNoise* 2025, 6457–6464. https://doi.org/10.61782/fa.2025.0605

Gutierrez-Parera, P., Arrebola, F., Cuevas-Rodriguez, M., Gonzalez-Toledo, D., Molina-Tanco, L., & Reyes-Lecuona, A. (2026). Hybrid ISM+BRIR convolution reverberation acoustic model with tuning of absorption coefficients based on single-position measurements. *Applied Acoustics*, 254 (December), 111448. https://doi.org/10.1016/j.apacoust.2026.111448

Politis, A., Pajunen, L., Leppänen, J., Mate, S., & Eronen, A. (2023). Wide-Area 6DOF Rendering of Multi-Point Ambisonic Recordings Based on Interpolation of Spatial Parameters. *2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2023-Oct, 1–5. https://doi.org/10.1109/WASPAA58266.2023.10248142

[15] https://musicsphere-eccch.eu/

[16] Disclaimer: *Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.*

Towards Bioadaptive VR for Autism: An EDA-Driven Adaptation of a Conversational Virtual Human

Leonardo Matteo Dinoi[1], Giovanni D'Errico[2], Ersilia Vallefuoco[3], Alessandro Pepino[3], Javier Marín-Morales[1], and Lucio Tommaso De Paolis[4]

[1]*Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, Universitat Politècnica de València, Valencia, Spain,*

[2]*Cultural Heritage Department, University of Salento, Lecce, Italy,*

[3]*Department of Electrical Engineering and Information Technology, University of Naples Federico II, Naples, Italy,*

[4]*Department of Engineering for Innovation, University of Salento, Lecce, Italy*

Corresponding author: lmdindin@doctor.upv.es

Keywords: Virtual Reality; Bioadaptive Interaction; Virtual Human; EDA

Abstract

This paper presents a virtual reality escape room in which real-time electrodermal activity modulates the verbal behaviour of a conversational virtual human. The system was developed as an initial feasibility step towards bioadaptive VR support for children on the autism spectrum. Electrodermal activity is processed in overlapping five-second windows and classified into three participant-specific EDA-derived activation levels, which are used to adjust the virtual human's conversational behaviour. 4 adults completed adaptive and non-adaptive conditions in a counterbalanced within-subject feasibility study. The system obtained a mean System Usability Scale score of 75.5. Self-reported arousal and baseline-adjusted skin conductance showed descriptive patterns favouring the adaptive condition, although the between-condition comparisons were not statistically significant. Given the small non-clinical sample, the results support the technical feasibility and usability of the interaction loop but do not establish the effectiveness of the adaptation.

Introduction

Virtual Reality (VR) provides controlled and repeatable environments in which tasks and stimuli can be adapted to different user profiles. These characteristics have encouraged its use in learning, rehabilitation, and social-skills training for people with Autism Spectrum Disorder (ASD) (Shahmoradi & Rezayi, 2022; Chen et al., 2022). However, many VR applications rely on predetermined interaction sequences and cannot respond to changes in the user's state during the experience. Physiological measures such as electrodermal activity (EDA), which reflects changes in sympathetic nervous system activation, offer a way to address this limitation by providing continuous information through wearable sensors (Critchley, 2002; Boucsein, 2013). Importantly, EDA is not treated here as a direct estimate of a specific emotional state, but as a participant-specific physiological cue for adaptation. Although physiological signals have been used to adapt virtual environments, both in

ASD-oriented adaptive systems (Bekele et al., 2016) while virtual humans have been explored for clinical support (De Dominicis et al., 2026) and bioadaptive arousal regulation (D'Errico et al., 2025), but their use for EDA-driven conversational support in ASD remains underexplored. For the intended application, the goal is to modulate the amount and style of conversational support according to changes in physiological activation, without relying solely on explicit requests for assistance.

Building on this approach, based on previous studies (Vallefuoco et al., 2025) we developed a VR escape room in which real-time EDA modulates the verbal behaviour of a conversational virtual human. The system implements an end-to-end interaction loop that converts participant-specific changes in skin conductance into discrete EDA-derived activation levels and uses them to adjust the virtual human's conversational behaviour. The contribution of this work is the integration and preliminary validation of this closed loop, in which physiological information is used to adapt conversational support rather than the virtual environment or task difficulty. We conducted an initial within-subject study with 4 adults to evaluate the technical feasibility and usability of the system. As this preliminary study did not involve the intended population, evaluating the effectiveness of the approach with children on the autism spectrum remains a subsequent stage of the research.

Bioadaptive VR System

The application is a single-room VR escape game developed in Unity for the Meta Quest 2. It comprises three cognitive tasks presented in a fixed sequence within an environment designed to limit unnecessary sensory stimulation. Throughout the experience, a conversational virtual human named Luca introduces the scenario, explains the tasks, and assists the user (Figure 1).



Figure 112. *Overview of the sci-fi lab environment.*

The bioadaptive interaction combines real-time EDA acquisition, signal processing, and a cloud-based conversational service (Figure 2). EDA is recorded using a Shimmer3 GSR+ and processed in Python on an external computer. Before gameplay, a three-minute resting baseline is collected to account for individual differences in skin conductance. During the experience, the first derivative of the skin conductance level is calculated over overlapping five-second windows and compared with participant-specific thresholds derived from the baseline to classify activation as low, medium, or high.

The resulting activation level is sent to Unity and provided to the conversational service as contextual information. It adjusts the personality traits of the Convai character that influence the wording and content of Luca's responses. At higher activation levels, these traits are configured to

elicit more structured, sensitive, and supportive verbal responses, whereas lower activation leads to a less directive conversational style. The adaptation affects verbal content only and does not alter task difficulty or the character's non-verbal behaviour. In the non-adaptive condition, the same game and conversational system are used, but the character's personality traits remain unchanged rather than being adapted according to the participant's EDA, and the recorded EDA does not influence Luca's responses.

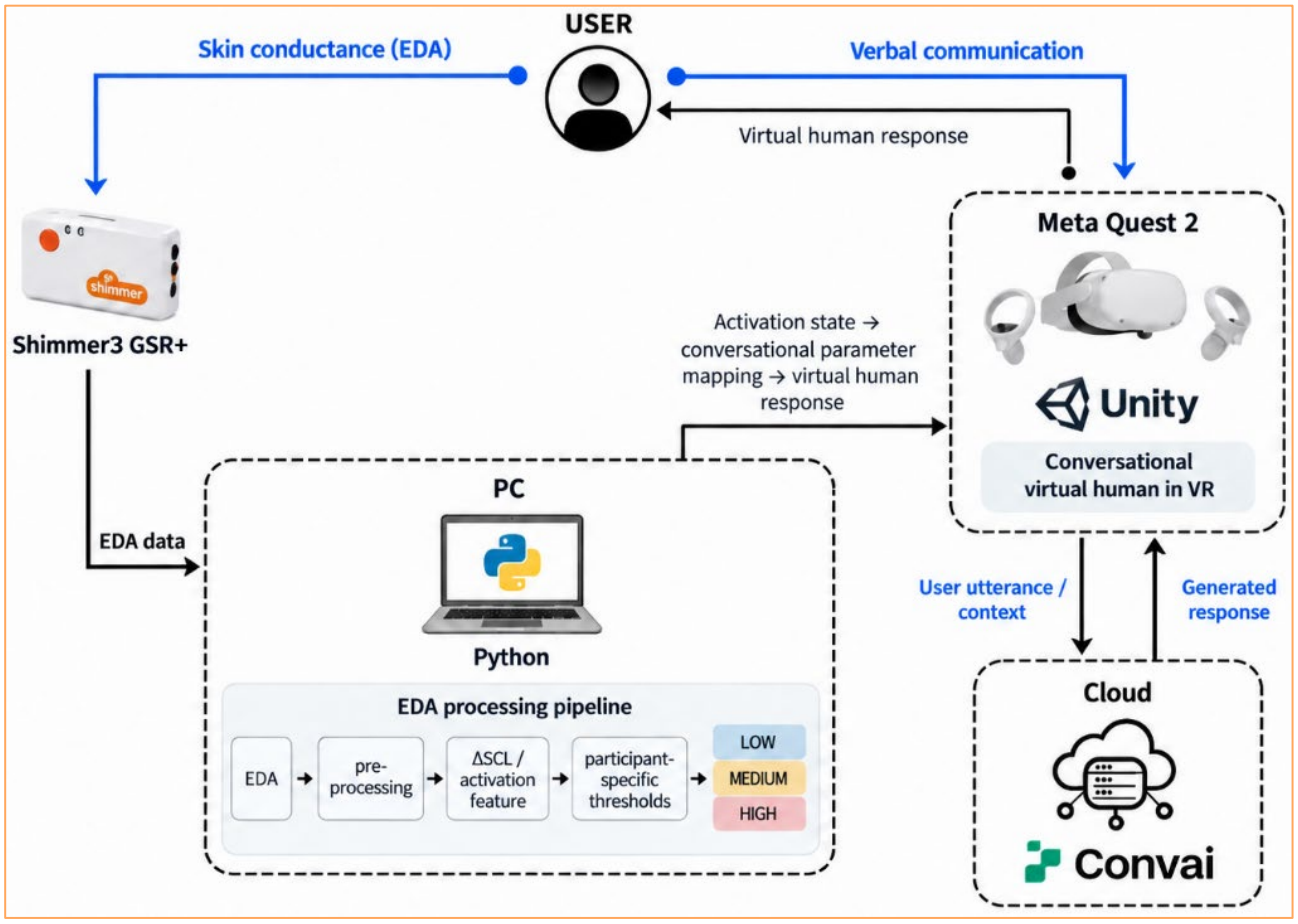


Figure 2. *High-level system architecture diagram.*

Preliminary Evaluation

4 adult volunteers recruited among students, researchers, and collaborators at the AVR Lab of the University of Salento participated after providing informed consent. This preliminary phase was limited to technical and procedural validation with consenting adults and did not involve members of the intended vulnerable population. Testing was conducted individually in a quiet university laboratory. Each participant completed adaptive and non-adaptive versions of the escape room, which shared the same narrative and interaction structure but used functionally matched, non-identical tasks to reduce learning effects. Game version and condition order were counterbalanced across 4 experimental sequences.

Before each condition, participants watched a three-minute nature video while an individual EDA baseline was recorded. Self-reported arousal was measured using the Self-Assessment Manikin after the baseline and after gameplay, while EDA was recorded continuously during both conditions. After completing the two versions, participants completed the System Usability Scale. The evaluation examined the operation of the physiological and conversational interaction loop, overall system usability, and descriptive differences in self-reported arousal and skin conductance between conditions.

Results and Discussion

All 4 participants completed both conditions. System Usability Scale scores ranged from 70 to 82, with a mean of 75.5, suggesting satisfactory overall usability in this small sample. Self-reported arousal decreased relative to baseline after the adaptive condition and increased after the non-adaptive condition. Baseline-adjusted skin conductance showed a similar pattern, with smaller changes during adaptive interaction. However, between-condition Wilcoxon signed-rank tests found no statistically significant differences in either measure (both $p = .25$). These patterns should therefore be treated as descriptive observations rather than evidence of an adaptation effect. The evaluation confirmed that physiological acquisition, real-time processing, the ambiguity of EDA as a measure of activation, and an implementation-driven adaptation policy based on broad personality-trait manipulation. Future work will instead investigate evidence-informed tutoring strategies. Interpretation is limited by the small non-clinical sample, the use of non-identical tasks whose equivalence was not independently validated, and the ambiguity of EDA as a measure of activation. The results support the technical feasibility of the system but do not establish adaptation effectiveness or clinical benefit.

Conclusion

The prototype demonstrates the technical feasibility of integrating real-time EDA processing with adaptive conversational behavior in an immersive virtual human. The next evaluation should use a larger adult sample, validated task equivalence, and a manipulation check to determine whether users perceive the intended changes in conversational support before extending the study to children on the autism spectrum.

References

Shahmoradi, L., & Rezayi, S. (2022). Cognitive Rehabilitation in People with Autism Spectrum Disorder: A Systematic Review of Emerging Virtual Reality-Based Approaches. *Journal of NeuroEngineering and Rehabilitation* 19, Article 91. https://doi.org/10.1186/s12984-022-01069-5.

Chen, J., Wang, G., Zhang, K., Wang, G., & Liu, L. (2022). Virtual Reality Enhances the Social Skills of Children with Autism Spectrum Disorder: A Review. *Interactive Learning Environments*, pp. 1–22. https://doi.org/10.1080/10494820.2022.2146139.

Critchley, H. D. (2002). Electrodermal Responses: What Happens in the Brain? *The Neuroscientist* 8(2), pp. 132–142. https://doi.org/10.1177/107385840200800209.

Boucsein, W. (2012). Electrodermal Activity. 2nd ed. *New York, NY: Springer* https://doi.org/10.1007/978-1-4614-1126-0.

Bekele, E., Wade, J., Bian, D., Fan, J., Swanson, A., Warren, Z., & Sarkar, N. (2016). Multimodal Adaptive Social Interaction in Virtual Environment (MASI-VR) for Children with Autism Spectrum Disorders (ASD). In: *2016 IEEE Virtual Reality (VR*), pp. 121–130. https://doi.org/10.1109/VR.2016.7504695.

D'Errico, G., Minissi, E., Arpaia, P., Altozano, A., Moccaldi, N., Calero, C., ... & Marin-Morales, J. (2025, June). 0.61782/fa.2023.1042. In *International Conference on Extended Reality* (pp. 277-288). Cham: Springer Nature Switzerland.

De Dominicis, A., D'Errico, G., Dinoi, L. M., Pallara, C., Rizzo, R., Scardapane, A., & De Paolis, L. T. (2026, June). Design of a Virtual Human-Based Digital Reception System for a Radiology Department: A Requirement-Driven Approach. In *International Conference on Extended Reality* (pp. 509-519). Cham: Springer Nature Switzerland.

Vallefuoco, E., Ponticelli, E., Gison, G., & Pepino, A. (2025). Virtual Reality Escape Room for Adolescents with Autism Spectrum Disorder. In: *Extended Reality: International Conference, XR Salento* 2025, Proceedings, Part VI, Lecture Notes in Computer Science, vol. 15742, pp. 475–483. https://doi.org/10.1007/978-3-031-97778-7_36.

Forgenie: A multi-input augmented reality interface for creating 3D models

James Cox, Elza Ibragimov, Chathuni Kulathunga, Sean Banerjee, Natasha Kholgade Banerjee

Terascale All-sensing Research Studio, Wright State University, Dayton, OH, USA

Corresponding author: natasha.banerjee@wright.edu

Keywords: Augmented Reality (AR), 3D modelling, extended reality (XR), computer-aided design (CAD), generative AI, text-to-CAD

Abstract

In this paper we present Forgenie, an augmented reality (AR) application that allows users to use multiple inputs such as text, speech, and sketches to create 3D models. Forgenie integrates off-the-shelf AI speech-to-text and text/image-to-3D model generation tools to create 3D models. Once the initial 3D model is created, we provide users with customization options such as color changes, rescaling, and repositioning. We also provide users with an export tool to save the generated models and print them using a consumer-grade 3D printer.

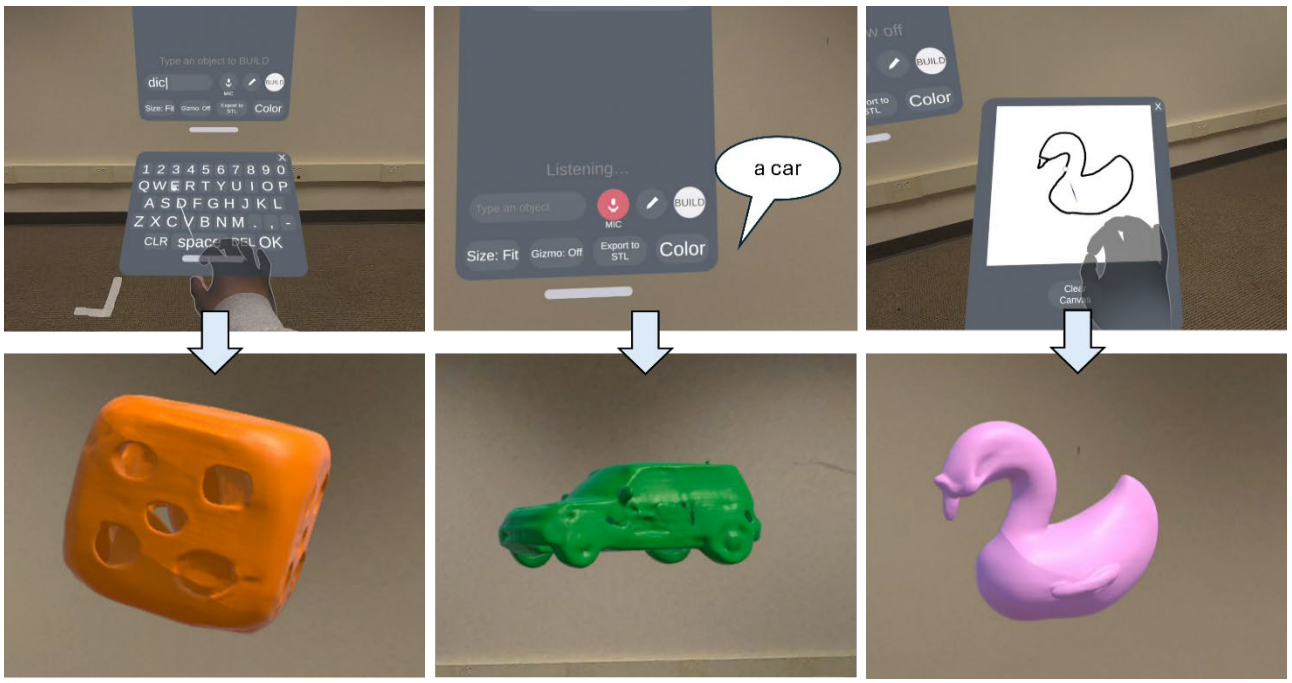


Figure 113. *We show a user building objects with Forgenie's input methods. Left: Arrow shows the result of a user typing "dice" into the application. Center: Arrow shows the result of a user speaking "a car" into the application. Right: Arrow shows the result of a user sketching on a canvas into the application.*

Introduction

The continuous growth of augmented (AR) and virtual reality (VR) devices and applications facilitates a greater focus on these tools being integrated in personal and industrial settings. Work by Subramanian et al. (2025) notes that AR and VR applications streamline operations in industrial workspaces by optimizing workflows and minimizing errors. We contribute to the growth and development of AR technology by creating an application, Forgenie, that allows users to build 3D objects and manipulate them in their environment. Forgenie allows users with varying knowledge and skill levels on AR, AI, and CAD tools to create and build their ideas.

Related Work

The approach of Weng et al. (2024) implements speech to 3D generated models into mixed reality (MR), but only provides one input modality for the user and does not allow color customization or exporting features for 3D printing. Work by Zhihao et al. (2026) introduces a pipeline that uses a combination of speech and gesture or speech and sketch to generate and place interactable props in a VR environment. Work by Tong et al. (2025) combines XR sketching with a descriptive verbal prompt to create virtual assets. Their approach relies heavily on drawing skill and description accuracy of the voice prompt. Unlike prior approaches, Forgenie allows the user to build using text, speech, or sketch. Forgenie allows customization of color and scaling within the AR environment with an additional ability to export the built objects for 3D printing and further development in the real-world.

System Design

We developed Forgenie as an AR application that can run on a consumer headset, such as a Meta Quest, and only requires access to a backend host computer with a high-end GPU, such as an NVIDIA RTX 4090, and a local network shared with the headset. The Forgenie application runs on the headset while generation, transcription, and processing run on the backend host computer. We use the Whisper AI speech recognition model (Radford et al., 2023) for speech-to-text transcription and WaLa (Sanghi et al., 2024), an approach that uses diffusion models for 3D model generation. As shown in Figure 1, Forgenie enables a user to create a 3D model by describing an object using text, speech, or sketch. During generation, we provide a status panel that indicate that generation is in progress. Once the 3D model has been generated, we allow the user to inspect and manipulate the object in the room around them. In Forgenie, we overlay the object on the user's view of the real room through the headset's passthrough feature.

Text-Based Generation

We use WaLa to generate a mesh from a text prompt in the panel in two stages. In the first stage we produce a set of depth maps from the text and in the second we reconstruct a mesh from the depth maps. We then return the finished mesh to the headset and import it into the scene as a .glb file. Our initial testing shows that it takes on average 35-47 seconds for the entire process from submission to seeing the generated object in the headset.

Speech-Based Generation

We capture spoken input by using the microphone on the headset. Once we capture the speech, we send it to Whisper to extract the text. Using Whisper, we normalize the generated text by converting it to lowercase and removing any punctuation marks before returning it to the headset for the user to review. Once the user is satisfied with the speech-to-text conversion, we follow the same pipeline as the text-based generation. Thus, our speech-based generation adds an extra overhead step to convert the spoken content into text using Whisper. However, this overhead is minimal and the overall processing time remains the same as the pure text-based approach.

Sketch-Based Generation

Unlike the text- and speech-based generation, for the sketch-based workflow, we pass the user's drawing directly to WaLa's sketch model. Using this approach, we produce a mesh in a single stage rather than the two stages used for text. Once the model is generated, Forgenie returns and imports the mesh as a .glb as with the text- or speech-based generation. We find that on average the sketch-based generation takes between 12-21 seconds.

User Interface Design

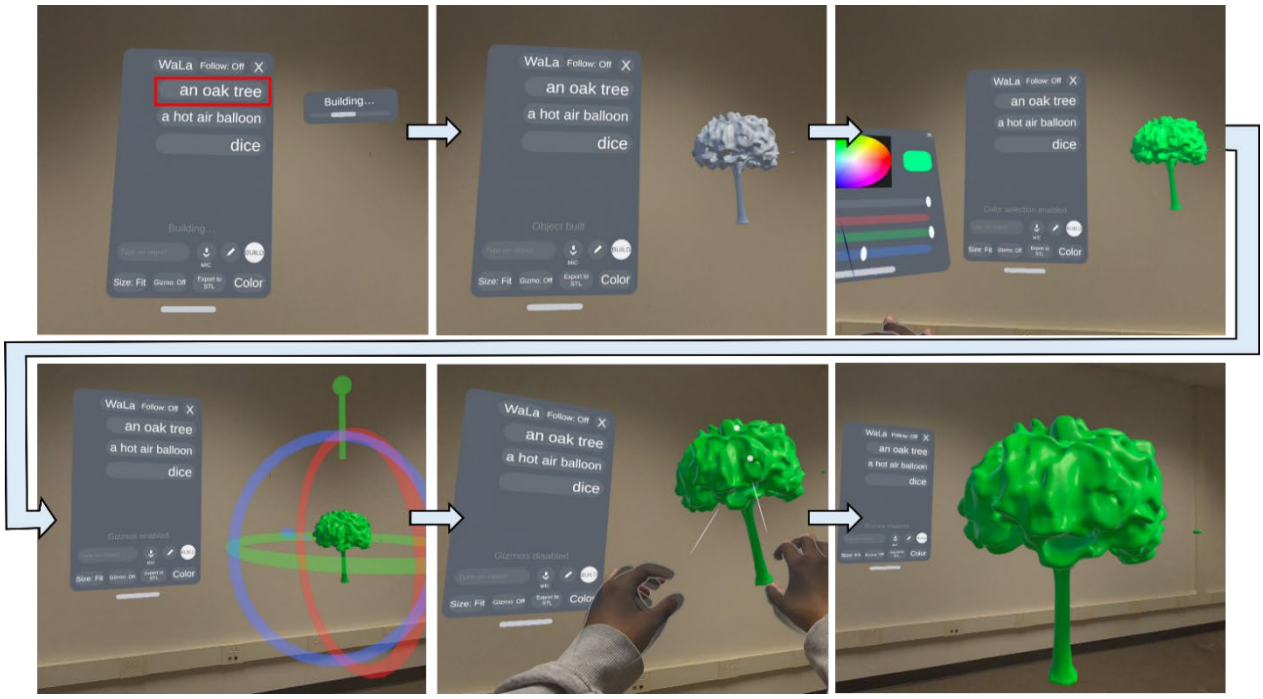


***Figure 114**. Using indicator arrows, we show the process of spawning "an oak tree" into the environment, the user coloring the object, the user moving the object using the gizmo, the user scaling the object using hand interactions, and the final scaled object in the passthrough environment.*

As shown in Figure 2, the user interface for Forgenie uses a minimalist approach to reduce clutter and minimize task overload. We allow the user to perform all interactions using a floating control panel in front of their field of view. We design the panel so that it follows the user with a dead zone so that small head movements do not cause the panel to drift. The panel only moves when the user's head moves far enough to trigger the following motion or when the user grabs the panel's handles to manually reposition it in the room. We also offer an anchor mode so users can choose where the panel move with their head or stays in the same position always. Our control panel contains all input controls, the BUILD button, and a status line that reports the current state of the system. The panel additionally displays the user's three most recent prompts as history entries that can be selected to reload a prior prompt into the input field without the user entering it in again.

In Forgenie, we allow the user to use the control panel to provide their input using text, speech, or sketch and then submit it using the BUILD button. In the text mode, the user uses a floating keyboard to enter the text. In the speech mode, the user presses on the microphone button, labeled as MIC, before speaking. Finally, in the sketch mode, the user uses a drawing pad and virtual pen to draw their design. In all modalities, pressing the BUILD button routes the content to the corresponding generation path.

Once the 3D model has been built, we import the result back into the headset scene and place it in the room beside the user's main control panel. Within Forgenie, we lock the model in the world space so that it does not follow the user's eye or head movement. By default, we scale each model so that its largest dimension is about 30 cm. This gives each object a consistent size regardless of

the actual generated scale. We allow the user to grab the model with one hand to move it or by pinching with the index and thumb fingers to transform the object around their space. We allow the user to also use both hands to pinch or grab the object and either stretch their hands out or inwards to scale the object in the corresponding direction. This same grabbing/pinching pattern can also be used to rotate the object. To allow the user to change the color of the generated model, we provide a color wheel that can be opened from the control panel. Finally, we provide users with an Export to STL option to export their model for 3D printing.

Conclusion

Through the design of Forgenie, we provide users with an AR application that allows them to describe objects using text, speech, or sketches and obtain a 3D model. The user can change the color and scale and export the generated model for 3D printing. As part of future work, we plan to add ranking features that allows a user to rank the input modalities and generated outputs, thereby allowing the AR interface to adapt to the user's preferred input mechanisms. In our current version of Forgenie we only allow the creation of a single 3D model. As part of future work, we plan to include multiple .glb files in one scene allowing for a user to see past creations alongside their newly generated objects and provide approaches to merge different objects together to create complex results. Forgenie currently only supports a single language, English, however, future versions can use multiple language inputs to allow the user to use their preferred language to interact with the application. We also plan on an updated sketch pad with artist tooling like pre-drawn shapes and basic drawing tools like selection/move, eraser, and fill/paint bucket to expand the capabilities of the freehand sketching function.

Acknowledgment

This work was partially supported by the U.S. National Science Foundation Award CMMI-2537030.

References

Chen, T., Ding, C., Zhang, S., Yu, C., Zang, Y., Li, Z., ... & Sun, L. (2024, June). Rapid 3d model generation with intuitive 3d input. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 12554-12564). IEEE. https://doi.org/10.1109/CVPR52733.2024.01193

Radford, A., Kim, J.W., Xu, T., Brockman, G., McLeavey, C. & Sutskever, I., (2023). Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pp. 28492-28518.

Sanghi, A., Khani, A., Reddy, P., Rampini, A., Cheung, D., Malekshan, K.R., Madan, K. & Shayani, H., (2024). Wavelet latent diffusion (wala): Billion-parameter 3d generative model with compact wavelet encodings. *arXiv preprint arXiv:2411.08017*. https://doi.org/10.48550/arXiv.2411.08017

Subramanian, S., Sampath, B., Loganathan, D.P.E., Natarajan, E., & Radhakrishnan, E. (2025). The integration of augmented and virtual reality in modern manufacturing. In *Manufacturing Strategies and Systems* (pp. 156-173). CRC Press.

Tong, Y., Qiu, Y., Li, R., Qiu, S. & Heng, P.-A. (2025). MS2Mesh-XR: Multi-Modal Sketch-to-Mesh Generation in XR Environments. 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), Lisbon, Portugal, 2025, pp. 272-276. https://doi.org/10.1109/AIxVR63409.2025.00052

Weng, S. C.-C., Chiou, Y.-M. & Do, E. Y.-L. (2024). Dream Mesh: A Speech-to-3D Model Generative Pipeline in Mixed Reality. 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), Los Angeles, CA, USA, 2024, pp. 345-349. https://doi.org/10.1109/AIxVR59861.2024.00059

Yao, Z., Yao, X., Xiong, H., Feng, Y. L., Sun, Q., Guo, Y., & Mi, H. (2026, April). GestuProp: 3D virtual reality prop generation with co-speech gestures. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems* (pp. 1-17). https://doi.org/10.1145/3772318.3790491.

From Voice-Coaching Metaphors to Integrated Breathing, Voice and Gesture Practice in Virtual Reality

Magdalena Majdak[1], Artur Cybulski[3], Sławomir Tadeja[4], Daniela Hekiert[5], and Magdalena Igras-Cybulska[2,3]

[1] *Institute of Polish Language, Polish Academy of Sciences, Warsaw, Poland*

[2] *AGH University of Krakow*

[3] *Interactive Crafts sp. z o.o., Lublin, Poland*

[4] *University of Cambridge, Cambridge, UK; 5: SWPS University, Warsaw, Poland*

Corresponding author: magda.igras@gmail.com

Keywords: Virtual reality; voice practice; breathing; embodied interaction; training

Abstract

We present a work-in-progress VR application that adapts voice-coaching scenarios combining breathing, phonation, gesture, and imagery. A modular pipeline progresses from relaxation and breathing to coordinated phonation, spatial creation, and expressive speech. The feedback loop combines acoustic events with controller-derived gesture descriptors. The contribution is a scenario-to-interaction specification and a protocol together with a formative evaluation plan. Interaction performance and training outcomes are to be validated in the future work.

Motivation and related work

Traditional voice coaching combines verbal cues, imagery and physical exercises to guide breathing and phonation (Majdak et al., 2014; Majdak, 2023/2024). VR can extend this practice through immersive, responsive visual feedback. Commercial applications such as Maloka and TRIPP offer immersive breathing and mindfulness practice (Ghosal et al., 2026). Research spans VR-based vocal training (Doganyigit and Islim, 2021), voice-therapy design (Hansa and Hansen, 2026), and combined voice-emission and public-speaking practice (Igras-Cybulska et al., 2020, 2023). Related approaches include voice-responsive forests (Bartyzel et al., 2024, 2025), pitch- and loudness-based game control (Godula et al., 2026), breathing games (Patibanda et al., 2017), and physiological research informing mixed reality breathing interfaces (Chen et al., 2025). Recent literature regarding the effectiveness of VR in breathing and speech training is inconclusive. For instance, Remacle et al. (2023) reported greater gains in self-reported communication self-efficacy in noise after training incorporating VR than after an information session alone, although VR's contribution was not isolated. Bostyn et al. (2026) found reduced vocal effort in both VR-supported and imagery-based practice, without demonstrating an additional short-term VR benefit.

Our work-in-progress extends this design space by a step-by-step progression from audible breathing to expressive speech for public-speaking preparation. Events detected from microphone

and gestures control vegetation, flower growth, tracing and painting. VR therefore enacts coaching imagery while avoiding excessive vocal effort.

Built in Unity 6.5 (URP/OpenXR) for standalone Meta Quest, primarily Quest 3, the prototype synchronises microphone audio with headset and controller tracking. Short-window energy, spectral and periodicity features detect audible breathing, /s/-like activity, and vocal onsets/offsets after calibration to ambient sound and comfortable practice; they do not estimate respiratory volume or technique. Design follows evidence that VR sensory conditions can affect vocal effort (Daşdöğen et al., 2026) and that users value personalisation and simple optional interactions (Wrońska et al., 2024). Feedback is brief and bounded, does not scale with vocal loudness or duration, and discourages extremes (NIDCD, 2025). Gesture dynamics, range and hand separation follow Skrzek et al. (2024) and will be coach-reviewed before the pilot.

Scenario adaptation and exercise repertoire

Table 1 specifies a modular repertoire adapted from the coaching scenarios based on Magdalena Majdak's programme.

Table 1. *Proposed scenario-to-interaction specification*

Exercise	Coaching task and user action	Planned input	VR response and adaptation
Body scan	Attend to body regions in sequence, from face to feet, and notice unnecessary tension.	User-controlled progression through spoken cues.	A schematic body guide highlights the cued region while the meadow remains still. Highlighting guides attention.
Embodied tree	Feel rooted feet and in the same time head suspended from the sky; explore gentle, comfortable sway; you shift your body weight, yet your relaxed, stable legs planted on the ground and your head suspended toward the sky form an axis that prevents your balance from being disrupted.	Headset displacement relative to a neutral position.	Roots remain fixed while a tree crown follows small head movements.
Air spheres	Feel a floral scent; bring an air sphere towards the mouth and let it permeate your inner self . Notice exhaled warmth on a free hand.	Controller approach; audible breath events.	The sphere dissolves before reaching the headset. Breath events can release butterflies near a symbolic chest location.
Imagined resistance and opening	Move the hands down, sideways and forwards along with the exhalation, then make an opening gesture.	Controller direction, excursion, dynamics and separation; audible breath events.	A translucent volume follows the hands; breath events add ripples. A mountain backdrop supports relief imagery.
Vowel-shaped exhalation	Shape the lips for vowels while exhaling and moving the arms, then add the voice.	Controller motion; audible breath events	Directional forms follow the hands, with brief breath-triggered ripples. Spoken cues guide articulation.

Exercise	Coaching task and user action	Planned input	VR response and adaptation
Vowels with gesture	Add comfortable a ([a]), o ([ɔ]), u ([u]) and e ([ɛ]) to the movement sequence.	Vocal onsets and offsets; controller motion.	Each onset starts a brief hand-following trail. Gesture descriptors modulate its spread and movement within calibrated limits.
Exhalation on /s/	Release a comfortable exhalation on /s/. You can feel them even in the lower parts of your torso.	Acoustic detection of /s/-like activity, distinguished from other noise.	A brief air-stream animation acknowledges detection. No duration target or compulsory breath hold.
Parting grain	Move both hands together left, then right, pairing each sweep with an exhalation.	Controller trajectories, dynamics and separation; audible breath events.	Grain bends along the hand paths; detected breaths add ripples. Spatial spread and animation speed reflect movement within the user's calibrated range.
Awakening the voice	Sigh, stretch and voice a gentle "five more minutes".	Vocal activity and controller movement.	A resting flower stirs after a vocal event; gestures open nearby foliage. Responses acknowledge activity without assessment of vocal quality.
Syllables and bell imagery	Try mam, mem, mom and mum; attend to the vowel and trailing /m/ (mammmm, mammmm, memmmm, memmmm, mommmm, mommmm, mummmm, mummmm-[mamː], [mɛmː], [mɔmː] and [mumː], respectively). You can feel the vibrations by touching the body with your free hand.	Vocal onsets and offsets.	Vocal onset starts a gentle bell-swing animation. The bell illustrates the coaching metaphor.
Tracing a flower	Follow the stem, leaves and petals with breath and attention, then with voice and attention.	Controller-pointer path; breath or vocal events according to the exercise phase.	An event enables one bounded contour segment, filled through pointer movement. Progress persists across pauses. Controller pointing explicitly replaces the source's gaze-guided tracing.
Growing and coloring a flower	Select a location; create a stem, buds and bloom with calm, long sounds from deep within the body, then color the flower.	Confirmed pointer target; separate vocal onsets; pointer dwell time.	Each onset advances one growth stage. During coloring, an onset enables a brief interval; pointer dwell increases saturation up to a fixed cap.
Painting the meadow	Paint with hand and voice while watching the emerging spatial image.	Vocal onsets; controller trajectory, excursion, dynamics and separation.	Each onset enables one short stroke. Movement determines its path and bounded

Exercise	Coaching task and user action	Planned input	VR response and adaptation
			spread. Strokes persist, allowing the image to develop across comfortable vocalizations and pauses.
Storytelling through numbers	Count to approximately 30 with a fairy-tale melody, addressing virtual listeners.	Speech activity; confirmed listener selection through the controller pointer.	The selected listener nods or smiles when speech begins. The response acknowledges selection and speech activity; it does not assess prosody, vocal direction or emotional impact.



Figure 1. *Proposed VR interactions: (A) air spheres; (B) gesture-responsive vegetation; (C) a flower at a controller-selected location.*

Conclusion and planned formative evaluation

This work contributes a coach-derived repertoire of breathing, phonation and expressive-speech exercises, together with explicit mappings from acoustic events and controller-derived gesture characteristics to bounded VR responses. The prototype is being implemented in Unity using OpenXR for Meta Quests. Future evaluation will compare matched VR and non-VR practice to test the added value of responsive visual feedback. Technical testing with 30 adults will assess detection precision, recall, false activations and latency using leave-one-participant-out validation. A separate group of 30 adults will evaluate instruction comprehension, comfort, user experience and feedback interpretation using questionnaires, logs and interviews.

Acknowledgements

This work disseminates results of the glossoVR 2.0 project, funded by the National Centre for Research and Development (NCBR), Poland, under the LIDER XI programme (grant no. LIDER/55/0230/L-11/19/NCBR/2020). The exercise design draws on voice-coaching scenarios authored by Magdalena Majdak. M. Majdak gratefully acknowledges the contribution of her teachers: These exercises would not have been created without the contribution of my teachers. I would like to extend special thanks to Swietłana Bucka and Aldona Krasucka, whose classes inspired the creation of these scenarios.

References

Bartyzel, P., Tadeja, S. K., Majdak, M., Łukawski, G., & Hekiert, D. (2024, March). Voice-responsive virtual reality training environment for occupational and non-professional voice users. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 208-213). IEEE. https:doi.org/10.1109/VRW62533.2024.00042

Bartyzel, P., Igras-Cybulska, M., Hekiert, D., Majdak, M., Łukawski, G., Bohné, T., & Tadeja, S. (2025). Exploring user reception of speech-controlled virtual reality environment for voice and public speaking training. *Computers & Graphics*, *126*, 104160. https:doi.org/10.1016/j.cag.2024.104160

Bostyn, L., Leyns, C., Saka, E., Vanhee, L., D'haeseleer, E., & Rombouts, E. (2026). Contextual Voice Training for Student Teachers: Exploring the Role of Virtual Reality. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2026.04.037

Bucka, S. (2025). Filozofia głosu. *Warszawa*.

Chen, K., Chang, Z., Zou, Q., & Kunze, K. (2025, October). Exploring Singing Breath: Physiological Insights and Directions for Breath-Aware Augmentation in Mixed Reality Design. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 702-706). https://doi.org/10.1145/3714394.3756159

Daşdöğen, Ü., Awan, S. N., Bottalico, P., Iglesias, A., Getchell, N., & Abbott, K. V. (2026). The influence of multisensory input on voice perception and production using immersive virtual reality. *Journal of Voice*, *40*(1), 253-e29. https://doi.org/10.1016/j.jvoice.2023.07.026

Doganyigit, S. & Islim, O. F. (2021). Virtual reality in vocal training: a case study. *Music Education Research*, 23(3), 391–401.

Ghosal, S., Zhang, M., Bogosian, A., Marsh, E., Edginton, T., Stanmore, E., & O'Connor, S. (2026). Virtual reality-based mindfulness applications: a commercial health app review. *BMJ Health & Care Informatics*, *33*(1), e101510. https://doi.org/10.1136/bmjhci-2025-101510

Godula, K., Tadeja, S., Hekiert, D., & Igras-Cybulska, M. (2026). Endless Runner F0: Loudness-Based and Pitch-Based Voice Control for VR Arcade Games. *IEEE Transactions on Games*. https://doi.org/10.1109/TG.2026.3692944

Hansa, J. & Hansen, H. (2026). User-centered qualitative evaluation of a fully immersive, head-mounted virtual reality application prototype to facilitate real-life transfer in voice therapy. *International Journal of Speech-Language Pathology*, 28(2), 303–315. https://doi.org/10.1080/17549507.2025.2473075

Igras-Cybulska, M., Cybulski, A., Hekiert, D., Majdak, M., Słomka, P., Szumiec, A., ... & Kacprzak, S. (2020, March). glossoVR-voice emission and public speech training system. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 832-833). IEEE. https://doi.org/10.1109/VRW50115.2020.00267

Igras-Cybulska, M., Hekiert, D., Cybulski, A., Tadeja, S., Witkowski, M., Nakonieczny, K., ... & Majdak, M. (2023, March). Towards multimodal VR trainer of voice emission and public speaking-work-in-progress. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 355-359). IEEE. https://doi.org/10.1109/VRW58643.2023.00079

Krasucka, A. (2006). O sposobach trenowania głosu. Z moich doświadczeń. In: *Emisja głosu nauczyciela. Wybrane zagadnienia*. M. Przybysz-Piwko ed., Warszawa. CODN, 61-71.

Majdak, M. (2023). Voice Self-Assessment in the Light of Acoustic Analysis – Case Studies. *Logopaedica Lodziensia*, 8, 65–78.

Majdak, M. (2023/2024). Voice exercise scenarios. *Unpublished teaching material*.

Majdak, M., Igras, M., & Domeracka-Kołodziej, A. (2014, April). Looking for natural voice—the effectiveness of the program of postgraduate studies of voice and speech training. In *XXII Annual Pacific Voice Conference (PVC)* (pp. 1-6). IEEE. https://doi.org/10.1109/PVC.2014.6845425

National Institute on Deafness and Other Communication Disorders (NIDCD). (2025). Taking Care of Your Voice. Accessed 7 September 2026. https://www.nidcd.nih.gov/health/taking-care-your-voice

Patibanda, R., Mueller, F. F., Leskovsek, M., & Duckworth, J. (2017, October). Life tree: understanding the design of breathing exercise games. In *Proceedings of the annual symposium on computer-human interaction in play* (pp. 19-31). https://doi.org/10.1145/3116595.3116621

Remacle, A., Bouchard, S., & Morsomme, D. (2023). Can teaching simulations in a virtual classroom help trainee teachers to develop oral communication skills and self-efficacy? A randomized controlled trial. *Computers & Education*, *200*, 104808. https://doi.org/10.1016/j.compedu.2023.104808

Skrzek, T., Malawski, F., Tadeja, S. K., Hekiert, D., Hemmerling, D., & Igras-Cybulska, M. (2024, March). Automatic Recognition of Gesture Qualities in Speech Delivery in VR-Work-in-Progress. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 624-627). IEEE. https://doi.org/10.1109/VRW62533.2024.00122

Wrońska, G., Sterna, R., Łukawski, G., & Igras-Cybulska, M. (2024, March). Exploring user preferences of VR relaxation experiences: A comparative mixed-methods study. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 877-878). IEEE. https://doi.org/10.1109/VRW62533.2024.00235

Extending the Surgical View: A Collaborative XR Interface for Spatial Communication in Robotic Surgery

García-Arocas, D.[1]; Nocera, A.[2]; Palma-Ruano, B. [1]; Pérez-Folgado, I. [1]; Riera, J.V. [1]; Gilabert-Estelles, J.[3]; Gimeno, J[1].

[1]*IRTIC, University of Valencia;*

[2]*Department of Information Engineering and EndoCAS, University of Pisa, Pisa, Italy;*

[3]*University General Hospital of Valencia, University of Valencia*

Corresponding author: jesus.gimeno@uv.es

Keywords: Extended reality; Collaborative interfaces; Robotic surgery; Spatial communication; Stereoscopic visualization.

Abstract

Robotic surgery places the console surgeon and other participants under different viewing and interaction conditions, which can hinder spatial communication. We present a collaborative XR interface combining shared stereoscopic surgical video with independently controlled virtual instruments based on robotic tool geometry and kinematics. Participants can use these instruments to point to structures, demonstrate and reproduce movements, and discuss intended actions within a common visual space. The interface supports controller and hand-tracking input and uses shared visual cues to communicate control states such as clutch activation. We describe the interaction design and initial prototype and report informal feedback from an expert-led session with postgraduate robotic surgery students. Participants valued reproducing observed movements and using their virtual instruments to formulate questions, particularly when discussing clutch operation. These preliminary observations suggest potential for surgical training and collaboration, although structured studies are required to assess usability and effects on communication and learning.

Introduction and related work

In robotic surgery, the console surgeon and other participants interact with the procedure under different viewing and interaction conditions. This separation can make it difficult to establish shared spatial references when discussing anatomical structures, instrument orientations, and intended movements, particularly in training scenarios.

Previous work has explored visual guidance beyond verbal instruction. Wild et al. (2022) investigated augmented reality telestration for identifying surgical targets. Jarc et al. (2016) introduced three-dimensional ghost tools for robotic surgery training, while Shabir et al. (2022) compared interfaces for controlling virtual surgical instruments in telementoring.

Building on these approaches, we present a collaborative XR interface combining stereoscopic surgical video with independently controlled virtual instruments. Unlike interfaces centred on

annotations or instructor-controlled guidance, our system enables both participants to point, demonstrate, reproduce, and discuss surgical movements through mutually visible tools. This work describes the interaction design and initial prototype, together with preliminary observations from an educational session.

System overview

The proposed system combines a shared stereoscopic environment, independently controlled virtual instruments, and reciprocal interaction to support spatial communication around surgical video.

Shared Collaborative Environment

All participants observe the same stereoscopic surgical video together with the virtual instruments controlled by each user (Figure 1). This common visual context allows explanations and demonstrations to be related to the same anatomical structures and surgical actions. The system supports both live and prerecorded video without requiring an explicit three-dimensional reconstruction of the surgical scene. Preliminary depth-estimation tests on static surgical images have explored the potential for improving this integration through depth-aware compositing. Estimated depth support more consistent occlusion between virtual instruments and visible structures; extending this approach to real-time video remains for future work.

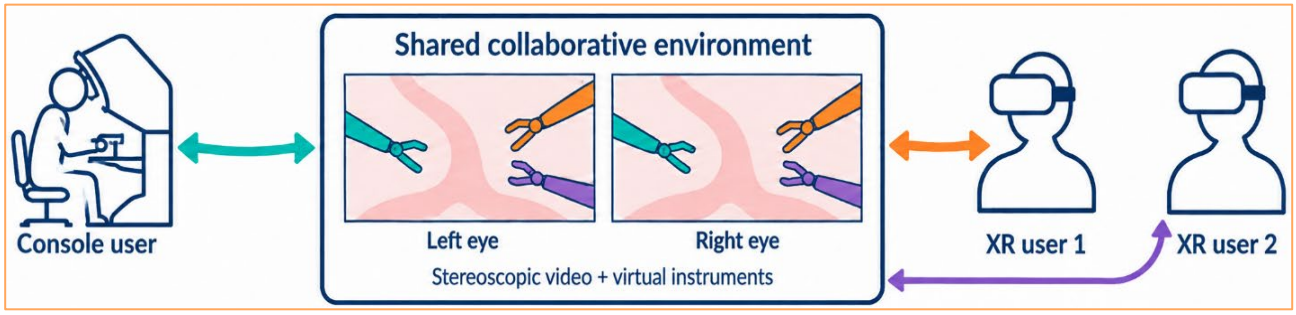


Figure 115. *Overview of the collaborative XR environment. Participants share stereoscopic surgical video and interact through independently controlled, colour-coded virtual instruments.*

Virtual instruments and input devices

The virtual instruments reproduce the geometry and main kinematic constraints of robotic surgical tools. Users can control their position, orientation, and gripper motion using either tracked controllers or hand gestures. The interface therefore supports video see-through devices, such as Meta Quest 3, and optical see-through devices, such as Microsoft HoloLens.

Visual cues, including colour changes, communicate control states such as clutch engagement and simulated electrosurgical activation. These cues make actions that may not be apparent from instrument movement visible to both participants. The instruments serve as communication tools, without simulating tissue interaction or the physical effects of electrosurgery.

Collaborative Interaction

Each participant independently controls their instruments while observing the actions of the other users. Participants can indicate anatomical structures, demonstrate approach directions and trajectories, reproduce movements, and discuss alternative actions. This reciprocal interaction supports instructor–learner and peer-to-peer exchanges without requiring users to transfer control of a single instrument.

Preliminary results

An initial educational session was conducted with an expert and students enrolled in the Master's programme in Robotic Surgery and Advanced Endoscopy at the University of Valencia. Participants explored the interface using prerecorded surgical videos. Figure 2 shows two different groups of

users collaborating during these initial tests. Informal feedback was positive regarding both educational use and spatial communication.

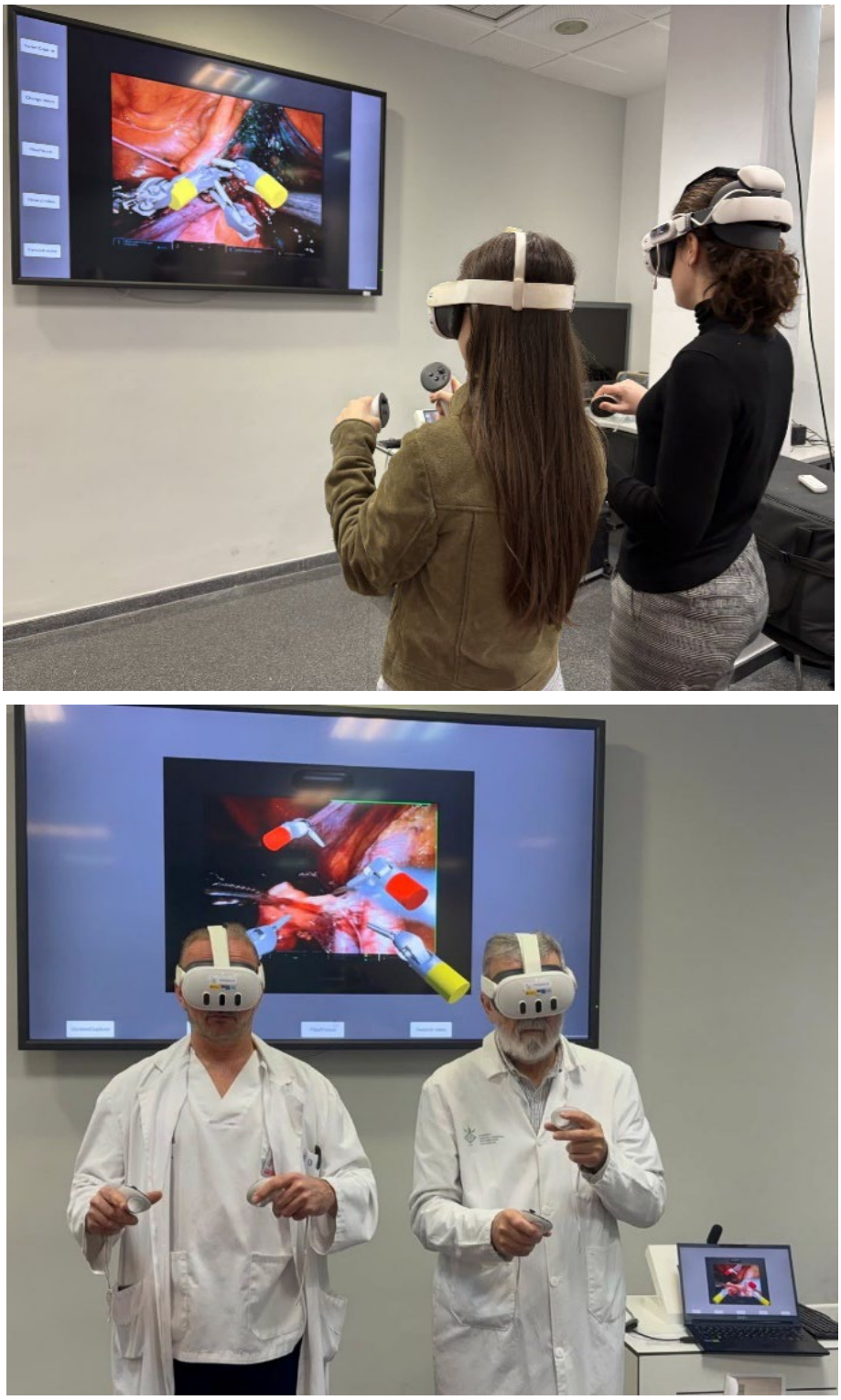


Figure 116. *Images captured during initial testing, showing two users collaborating over two prerecorded surgical videos.*

Students highlighted that virtual instruments remained useful even when physical instruments were visible in the video. They valued being able to attempt the observed movements and use their own virtual tools to formulate questions to the expert. Shared visual cues were also relevant: students specifically highlighted the colour change indicating clutch engagement as useful for understanding and discussing instrument control.

These observations provide preliminary feedback on the interface's potential. They do not establish improvements in communication or learning, which require a structured evaluation.

Conclusions and future work

This work presented a collaborative XR interface combining stereoscopic surgical video with independently controlled virtual instruments. Shared instrument movements and visual cues provide a means of expressing spatial information and control states, supporting reciprocal demonstrations and discussion. Informal feedback from an initial educational session suggests potential for this approach, although its effects on communication and learning remain to be systematically evaluated.

Future work will extend the preliminary depth-estimation tests on static images towards real-time video processing, exploring depth-aware compositing to improve occlusion between virtual instruments and the surgical scene. Further development will refine device-specific input mappings and visual cues, alongside a structured evaluation of how participants interpret, reproduce, and discuss actions through the interface.

Acknowledgments

This work was partially funded through the MIRACLE project (TSI-100123-2024-0056), supported by the Spanish Ministry for Digital Transformation and the European Union – NextGenerationEU.

References

Wild, C., Lang, F., Gerhäuser, A. S., Schmidt, M. W., Kowalewski, K. F., Petersen, J., ... & Nickel, F. (2022). Telestration with augmented reality for visual presentation of intraoperative target structures in minimally invasive surgery: a randomized controlled study. *Surgical Endoscopy*, *36*(10), 7453-7461. https://doi.org/10.1007/s00464-022-09158-1.

Jarc, A. M., Shah, S. H., Adebar, T., Hwang, E., Aron, M., Gill, I. S., & Hung, A. J. (2016). Beyond 2D telestration: an evaluation of novel proctoring tools for robot-assisted minimally invasive surgery. *Journal of Robotic Surgery*, *10*(2), 103-109. https://doi.org/10.1007/s11701-016-0564-1.

Shabir, D., Anbatawi, M., Padhan, J., Balakrishnan, S., Al-Ansari, A., Abinahed, J., ... & Navkar, N. V. (2022). Evaluation of user-interfaces for controlling movements of virtual minimally invasive surgical instruments. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *18*(5), e2414. https://doi.org/10.1002/rcs.2414.

Design and Evaluation of an XR-Based Immersive Teleoperation System with Multimodal Operator Monitoring

David Barbero[1], Ignacio Benito[2], Pablo Pérez[2], and Jesús Gutiérrez[1]

[1] Grupo de Tratamiento de Imágenes (GTI), Universidad Politécnica de Madrid, Madrid, Spain

[2] Nokia XR Lab, Madrid, Spain

Corresponding author: jesus.gutierrez@upm.es

Keywords: eXtended Reality, teleoperation, operator monitoring, mental workload, stress, quality of experience, presence

Introduction

Teleoperation has advanced considerably in recent years, driven by the latest technological developments, and has gained particular relevance in fields such as telesurgery, remote driving or intervention in hazardous environments. Some technical limitations still persist in these systems, such as communication latency or available bandwidth, which are being progressively addressed thanks to the deployment of 5G and the evolution towards 6G (Moniruzzaman et al. 2022). There is, however, another major limitation that no network improvement can solve: the operator is not physically present in the remote environment, and two problems of a different nature arise from this absence.

The first one is perceptual, since the operator accesses the remote environment only partially and indirectly, through an intermediate device. This degrades their situational awareness and, with it, their ability to anticipate and react. This is where interfaces based on eXtended Reality (XR) come in, replacing 2D monitors and multi-camera views with a three-dimensional representation that surrounds the operator (Roldán et al. 2017).

The second one concerns the state of the operator while performing the task, an issue that has generally received considerably less attention in similar works. The lack of presence mentioned above, together with the demands of the task, raises the operator's mental workload and stress, and both directly condition performance and the safety of the operation. The operator's state, however, is usually collected after the task by means of questionnaires, which do not reflect what was happening at each moment.

By contrast, continuous monitoring of physiological signals offers a way to track that state throughout the whole execution of the task, with the temporal resolution that questionnaires cannot provide (Iarlori et al. 2024).

The system presented in this work addresses both problems at once. The operator drives a small-scale robotic vehicle from a virtual cockpit, using a VR headset and a real steering wheel and pedals. Simultaneously, a set of biosensors continuously records their physiological activity without

interfering with the task. Preliminary evaluations confirm that the immersive experience is satisfactory and that the scenarios induce progressively higher levels of mental workload and a clearly different stress response.

System Overview

The system consists of two subsystems operating in parallel: the immersive teleoperation subsystem and the operator monitoring subsystem. The former (Figure 1) aims to recreate a real driving experience within a virtual environment. To this end, it uses a robotic vehicle (*Hiwonder JetAcker*) as the teleoperated device, which carries three USB cameras mounted on its chassis that capture what happens around it. The operator sees these images in real time through a virtual reality headset (*HTC Vive Pro Eye*) and drives the vehicle with a steering wheel and pedals (*Logitech G29*). The virtual environment recreates the interior of a car: the images from the three cameras are projected onto the windscreen and the side windows, and the steering wheel, the dashboard and the engine sound respond in real time to the operator's actions. A laptop, in turn, acts as the operator station, running the virtual environment, reading the wheel and pedal commands and sending the motion orders to the robot. All of this communicates over a local Wi-Fi network.

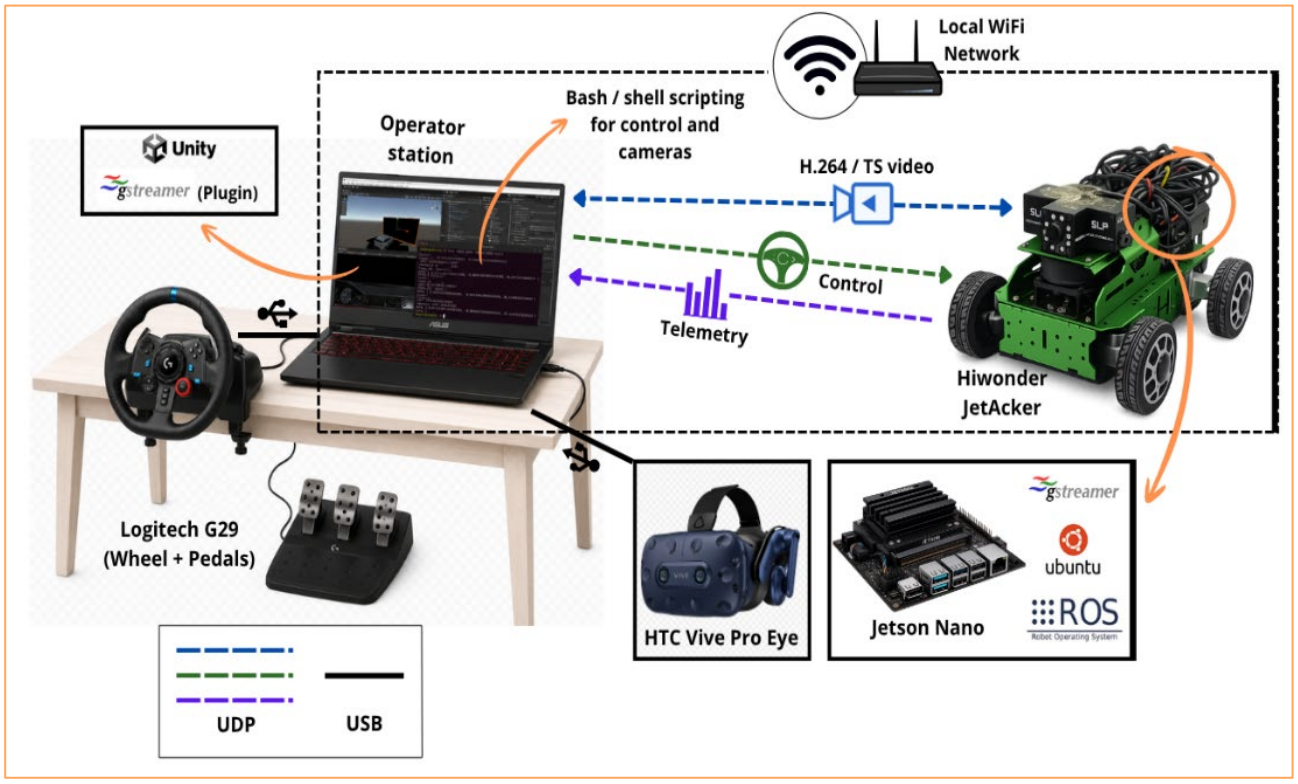


Figure 1. *Overall architecture of the immersive teleoperation system*

The second subsystem (Figure 2) is the operator monitoring one, which continuously and synchronously records the operator's physiological and ocular response while driving. It is built around three elements sharing a common time reference: the set of biosensors worn by the operator, the eye-tracking system built into the VR headset itself, which provides gaze information without the need for additional equipment, and an application developed specifically for this work, which brings all the signals together in a single interface and makes it possible to check that they are being recorded correctly.

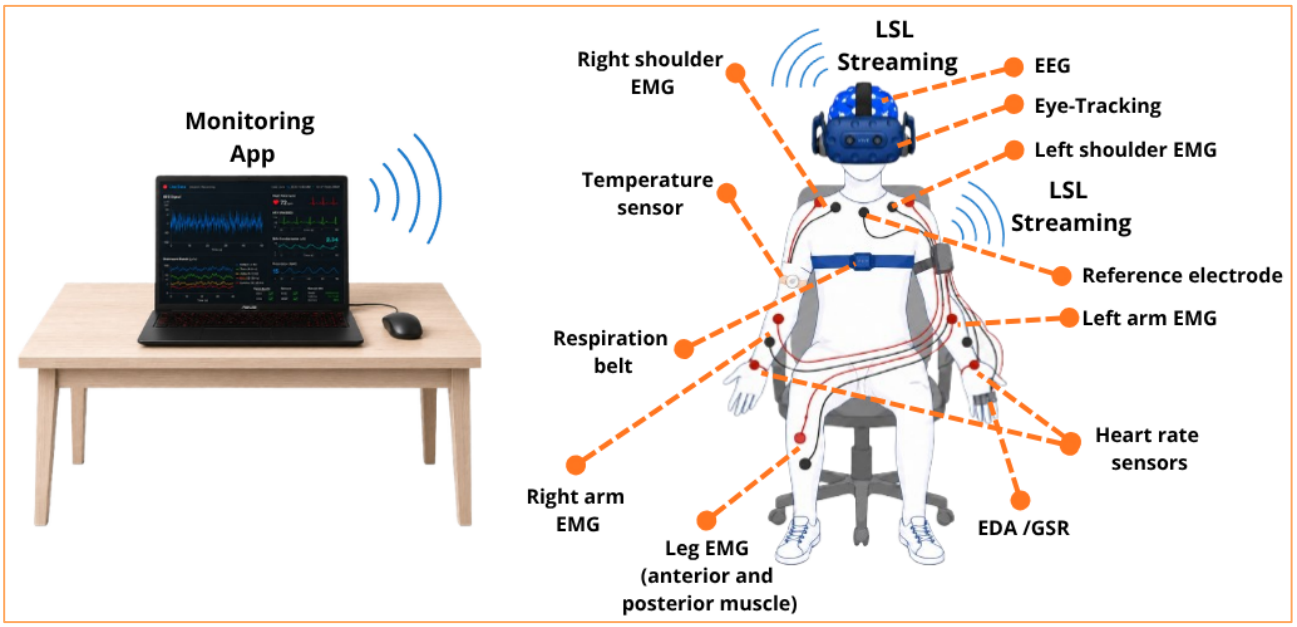


Figure 2. *Diagram of the operator monitoring system*

Experimental Design and Evaluation

The system has been evaluated with 27 participants (17 men and 10 women; mean age 25.4 years), mostly students and staff of the university. Most of them had barely used virtual reality before, although almost all of them could drive. Each test session lasted about 90 minutes and always followed the same protocol: instructions and an initial questionnaire, sensor placement, a training phase, the three proposed scenarios (each with its associated questionnaire) and a final questionnaire.

Those scenarios were set up on a physical circuit built on the floor of a laboratory room, where the operator station was also installed (Figure 3), and were designed with a deliberately increasing difficulty. The first one involved basic driving: two laps with the bends as the only demand. The second one followed exactly the same route but added typical urban situations: obstacles to avoid, a traffic light, a pedestrian crossing, a car unexpectedly cutting across, a stop sign and a parking manoeuvre. The third one was a time trial: a single lap of the full circuit, to be completed as fast as possible, with the elapsed time in sight and a secondary visual-search task that forced attention to be divided. The order was the same for all participants. After each scenario, participants rated their mental workload with the NASA-TLX (Hart 2006) and their stress with the SSSQ (Helton 2004). Cybersickness was also tracked throughout the whole session, and the final questionnaire collected an overall assessment of the immersive experience.

Results

Preliminary results confirmed that the three scenarios induced the demand levels they had been designed for (Figure 4). Global mental workload increased steadily across scenarios. Stress, by contrast, followed a different pattern depending on the dimension: *worry* and *engagement* remained high and practically constant, whereas *distress* was the only one that varied, rising appreciably in the time trial. Participants were therefore equally engaged in all three scenarios, but tension responded specifically to the time pressure of the last one.

The system, for its part, received a positive assessment. Quality of experience obtained a mean of 2.96 out of 4, practically in the “good” category, and the presence and immersion index reached 3.9 out of 5, with the items related to the feeling of being in the remote environment above 4. Cybersickness remained at tolerable levels, and the biosensors were also very well accepted, with participants not perceiving them as interfering with driving.

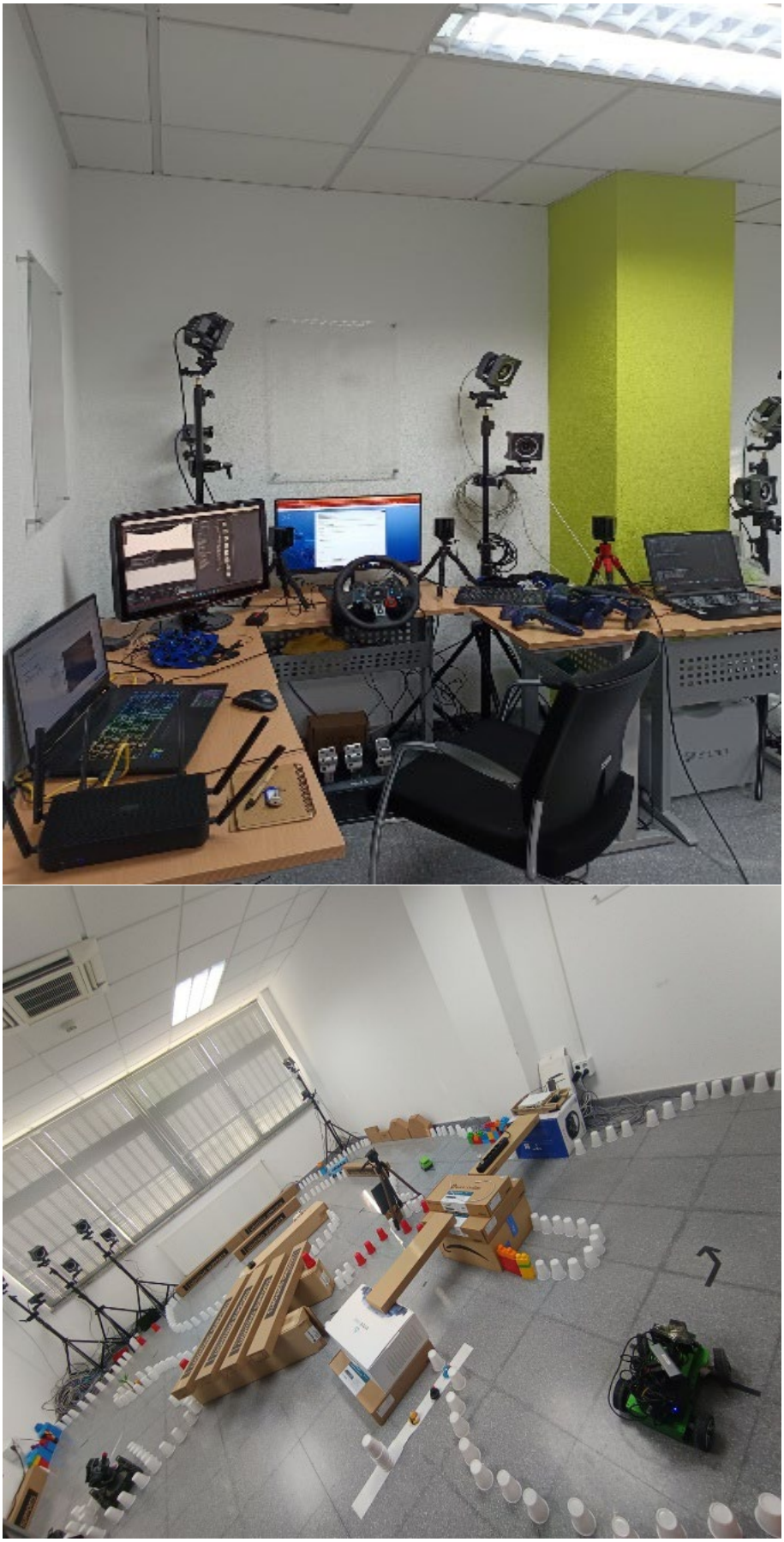


Figure 3. *Operator station (up) and scale circuit used in the three scenarios (down)*

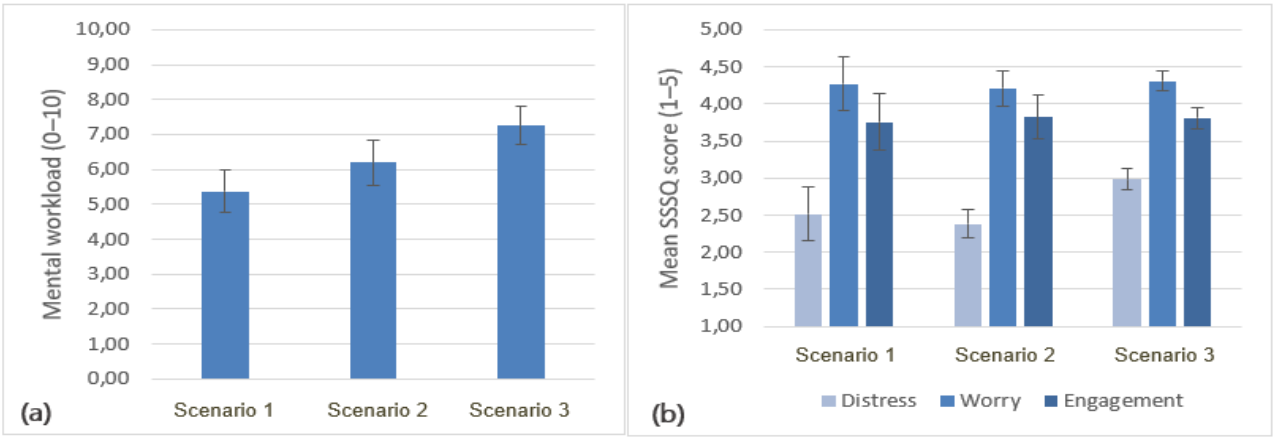


Figure 4. *Global mental workload (a) and SSSQ dimensions (b) per scenario (mean ± 95% CI)*

Conclusions and Future Work

This paper has presented an immersive teleoperation platform that combines a virtual cockpit with continuous monitoring of the operator. The preliminary evaluation shows that the system delivers a satisfactory immersive experience and, most notably, that its scenarios induce clearly differentiated levels of mental workload and stress response. The platform therefore constitutes a valid testbed for studying the operator's state in immersive teleoperation. Building on this, ongoing work focuses on exploiting the recorded physiological signals to automatically infer that state, a line still under development for which no results are reported in this paper. Further ahead, the aim is to move towards real-time estimation, add objective driving measures, study the effect of bitrate, resolution or latency on the operator, and keep improving the platform.

References

Moniruzzaman, M. D., Rassau, A., Chai, D., & Islam, S. M. S. (2022). Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey. *Robotics and Autonomous Systems*, *150*, 103973. https://doi.org/10.1016/j.robot.2021.103973

Roldán, J. J., Peña-Tapia, E., Martín-Barrio, A., Olivares-Méndez, M. A., Del Cerro, J., & Barrientos, A. (2017). Multi-robot interfaces and operator situational awareness: Study of the impact of immersion and prediction. *Sensors*, *17*(8), 1720. https://doi.org/10.3390/s17081720

Iarlori, S., Perpetuini, D., Tritto, M., Cardone, D., Tiberio, A., Chinthakindi, M., ... & Monteriù, A. (2024). An overview of approaches and methods for the cognitive workload estimation in human–machine interaction scenarios through wearables sensors. *BioMedInformatics*, *4*(2), 1155-1173. https://doi.org/10.3390/biomedinformatics4020064

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications. https://doi.org/10.1177/154193120605000909

Helton, W. S. (2004, September). Validation of a short stress state questionnaire. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 48, No. 11, pp. 1238-1242). Sage CA: Los Angeles, CA: Sage Publications. https://doi.org/10.1177/154193120404801107

Introducing an IEC 62559-Based Use Case Repository and Architecture Model for Virtual Worlds

Arjun Rai Gupta[1], Prof. Christoph Runde[1], and Prof. Rolf Kruse[2]

[1]*Virtual Dimension Center w. V., Fellbach, Germany*

[2]*Fachhochschule Erfurt, Germany*

Corresponding author: arjunrai.gupta@vdc-fellbach.de

Keywords: Virtual worlds; extended reality; use case repository; IEC 62559; architecture model; interoperability; standardization

Problem Context and Motivation

Standards development organisations (SDOs) repeatedly agree that virtual worlds need shared terminology and structured use cases. The IEC Market Strategy Board's 2025 trend report on the industrial metaverse closes with six recommendations, five of them in effect requests for structure: shared terminology and ontology, governance standards, digital asset management, a testing framework for use cases, and interoperability protocols (IEC, 2025). Each presupposes what none provides: a way to say where a use case sits, so a gap can be pointed at rather than argued about.

Saying where a use case sits is hard in this field. The Joint Research Centre found a broad range of domains but a narrow range of activities, describing virtual worlds as assemblies of modules recombined for different contexts (Boucher et al., 2026). In the four-dimensional StandICT.eu classification schema (StandICT.eu TWG Virtual Worlds, 2026), that is many verticals (dimension D3) against few application fields (D2). Any classification must also accommodate a dual nature: virtual worlds are a cross-sector layer applied in many domains, and a domain in their own right. The citiverse shows why: ITU treats it as virtual worlds applied to the city and develops its standards in the study group for smart sustainable cities, yet the city is no single domain, since the citiverse use cases it curates span urban planning, city administration, economic development and tourism, transport and mobility, and public safety and health (ITU et al., 2025).

This paper introduces a use case repository (UCR) for virtual worlds built on the IEC 62559 methodology, supported by classification schemas drawn from current standards-coordination work, a candidate Virtual Worlds Architecture Model (VWAM), and a roadmap for deploying the UCR as a collaborative platform.

State of the Art, Research Gap and Methodology

The methodology comes from the smart grid, a system of systems in which many vendors' components must act as one infrastructure (Uslar & Engel, 2015). To make its deployment possible, the European Commission mandated the European standardisation organisations to develop the standards it needs, through mandate M/490 (European Commission, 2011). Standards bodies can only close gaps they can see, so the mandate work adopted a use case methodology to elicit which

functions are required and which standards are missing (Gottschalk et al., 2017). IEC then turned its power-system use case specification, IEC PAS 62559:2008 (IEC, 2008), into the IEC 62559 series, meant for technical systems beyond the smart grid (IEC, 2015, 2017, 2019).

The template of IEC 62559-2 (IEC, 2015) captures actors, business and system use cases, scenarios and steps, information exchanges and requirements in predefined fields, so use cases from different contributors can be compared rather than read as prose. Three provisions beyond the template matter here: the repository is described as a working database with roles governing who may create, review or approve; a staged workflow ties access to lifecycle state; and an XML format moves use cases between tools (IEC, 2019, 6.1, 6.4; IEC, 2017).

For the smart grid, use cases are placed on the Smart Grid Architecture Model (SGAM; CEN-CENELEC-ETSI SG-CG, 2012; Gottschalk et al., 2017), built on the GridWise interoperability stack (Ambrosio & Widergren, 2007) and adding a plane of two axes: the electrical conversion chain as domain, an information management hierarchy as zone. It is a reference designation system, not a reference architecture (Uslar & Engel, 2015). Mapping use cases onto it shows which interfaces and information exchanges lack a standard: possible gaps for technical bodies to confirm (IEC, 2019). Derivatives followed, among them an architecture model for active assisted living (Gottschalk et al., 2017), the Smart City Infrastructure Architecture Model (SCIAM; DKE/DIN & VDE, 2015) and the Reference Architecture Model Industrie 4.0 (RAMI 4.0; DIN, 2016); Figure 1 shows SGAM, SCIAM and RAMI 4.0. Virtual worlds have no equivalent of any of these: no shared template, no repository, and no architecture model to place a use case on. The method adopted here is to take the description standard unchanged and build the domain adaptation around it.

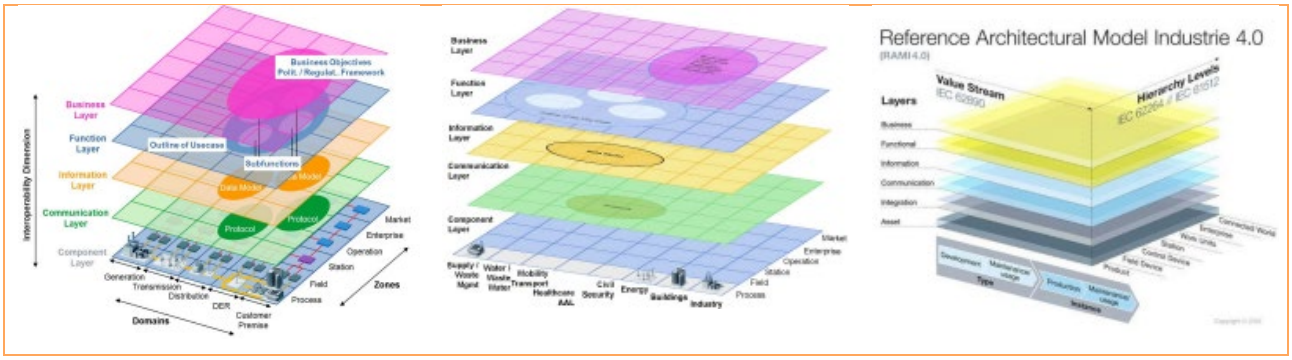


***Figure 1**. SGAM (CEN-CENELEC-ETSI SG-CG, 2012), SCIAM (DKE/DIN & VDE, 2015) and RAMI 4.0 (DIN, 2016), reproduced from the cited sources.*

IEC 62559-2 recommends a common glossary for all use cases and notes that a database lets definitions be filled automatically (IEC, 2015). For virtual worlds, terminology is further along than use cases but not yet settled: ISO/IEC 24931-1 is at final draft (ISO/IEC, 2026), the ITU-T Focus Group on metaverse has published 52 technical reports and specifications (ITU-T FG-MV, 2024), and IEEE P2048 is in development (IEEE, 2026), while many further standards define the same terms, sometimes in conflict. The Virtual Dimension Center (VDC) holds a terminological database structured to ISO 16642 (ISO, 2017) and ISO 26162-1 (ISO, 2019), which feeds the repository's glossary.

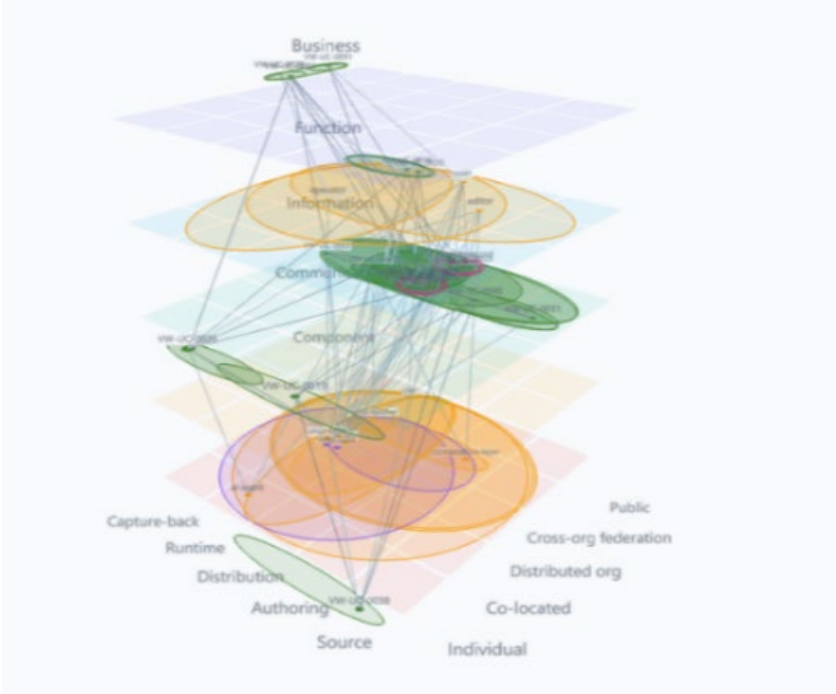
Implementation and Results

The Virtual Worlds UCR is a server-rendered repository with an addressable record for every object the methodology names. Form rules derived from IEC 62559-2 are enforced as deferred database constraints at commit, so a non-conforming record cannot be stored, and the export validates against the IEC 62559-3 schema (IEC, 2017).

Implementations are first-class records: IEC 62559 describes use cases, but the field's published material is overwhelmingly deployment description, so the repository ingests deployments and derives use cases from them. Each carries an evidence class, from operational deployment down to provider showcase, which travels with every claim derived from it.

To hold the dual nature described above, classification uses several independent axes rather than one tree, supplemented by keyword tagging. StandICT.eu classification schemas are applied where each fits: D3 (verticals) tags implementations, D1 (technology) tags systems and actors, and D4 (society, market, policy) tags use cases. Use cases also carry the concepts, domains and technologies of ISO/IEC 24931-1 (ISO/IEC, 2026) and eleven cross-framework canonical concepts, derived by Rolf Kruse from various metaverse taxonomies, classifications, and frameworks, and now proposed for ISO/IEC 24931-2 (ISO/IEC JTC 1/SC 24, 2026). A parallel classification follows the sectors of the Virtual Worlds Partnership's research agenda (Virtual Worlds Association, 2025).

a)



b)

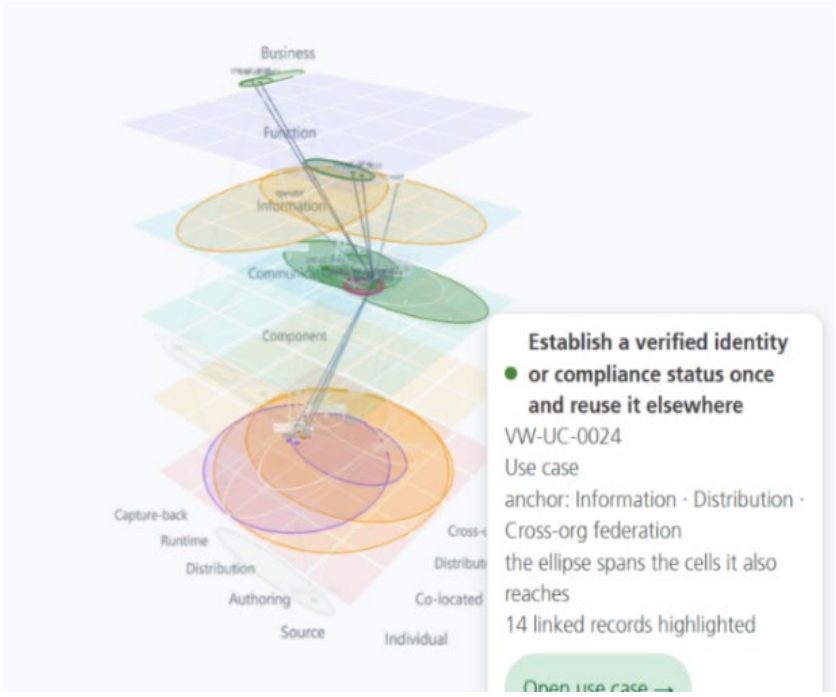


***Figure 2**. The proposed VWAM in the repository's model viewer. (a) All placed records of the cluster 'Portable identity and rights across worlds'. (b) One record selected with its linked records highlighted.*

Placement is separate from classification: each record is also placed on the VWAM (Figure 2), which has three axes. Layers were retained from SGAM, the interoperability stack describing degrees of abstraction in any system of systems. Domain and zone were rejected: the electrical conversion chain has no analogue, and the zones descend from the Purdue reference model by way of IEC 62264-1 (IEC, 2013), where levels are ordered by function and time frame above a physical process. Its ground-plane axes are instead a content pipeline (Source, Authoring, Distribution, Runtime, Capture-back) and participation reach (Individual, Co-located, Distributed organisation, Cross-organisational federation, Public). A record is placed by an anchor cell and a spread over the cells it also reaches (Figure 2a). Figure 2b selects ‘Establish a verified identity or compliance status once and reuse it elsewhere’, anchored at Information, Distribution and cross-organisational federation, with its 14 linked records. The record belongs to the cluster ‘Portable identity and rights across worlds’ and is derived in part from use cases of the Metaverse Standards Forum (2025).

Evaluation, Impact and Outlook

The repository's unit of work is a chain from a business argument down to testable requirements, for example: the business case ‘Train without exposing people to hazard’; under it the generic business use case ‘Bring an apprentice to competence on a task that carries a hazard’; beside it the individual use case ‘Qualify technicians to isolate high-voltage vehicles’; and below, ‘Train a procedural assembly or maintenance task on complex equipment’ and ‘Train an emergency response procedure under simulated hazard’, with 86 and 46 reference implementations respectively.

Actors are held once and reused, with business roles such as trainee kept apart from system roles such as learning record store; naming an application in a business use case is refused at commit. Information objects find existing data models: ‘Asset geometry and pose’ names OpenUSD (AOUSD, 2025), glTF (ISO/IEC, 2022) and Industry Foundation Classes (IFC; ISO, 2024), ‘Competence outcome’ names the Experience API (xAPI; ISO/IEC/IEEE, 2025), ‘Maintenance history’ names ISO 14224 (ISO, 2016).

The work has two limits. First, in ISO/IEC/IEEE 42010 terms (ISO/IEC/IEEE, 2022) the VWAM grid is a structuring formalism, like RAMI 4.0. An architecture description framework must also identify stakeholders, their concerns and the viewpoints framing them; the VWAM specifies none, so it is not yet such a framework.

Second, the corpus was derived with machine assistance, so every derived record is draft and every actor, area and tag provisional. Only domain experts contributing and reviewing records can make it authoritative. Kuchenbuch et al. (2026) found that language-model assistance roughly doubled the errors experienced modellers detected but did little for beginners, so planned AI assistance only proposes changes to a draft, which an editor accepts or rejects.

The second limit sets the next step. IEC 62559 describes the repository as a shared working platform, and none of its role or lifecycle provisions is exercised by a single-author corpus. The first public release will be a read-only citable repository. Contributions will then be accepted through a short structured submission, below the template skill IEC TR 62559-1 expects of a use case developer, with editors doing the template work (IEC, 2019).

The intended end state is federated rather than central. The system ships as self-hostable containers, so an organisation can run its own instance and still exchange records in IEC 62559-3. 433 implementation records already link out to showcases published elsewhere, whose publishers are asked to claim their record, and a planned read-only API will serve the structured description. The current invitation is narrow: bring a use case, place it on the model, and say where the model fails. Findings are proposed to feed ISO/IEC PWI 24931-3 (ISO/IEC JTC 1/SC 24, 2025) and the related SC 24 work on “Use case type classification for VR/AR/MR systems integration” (AWI 25689) and “Use cases and design guidelines for VR/AR/MR based learning, education, and training” (PWI 26952) (ISO/IEC JWG 12, 2026). The nearly empty Communication layer could draw on the VDC Standards Register (Runde et al., 2026) and the Metaverse Standards Forum's Spatial Systems

Atlas (Bilham, 2026), each system admitted only with a use case that uses it. Collaborators can join the closed alpha now at https://usecases.vdc-fellbach.de.

Acknowledgement

The authors are StandICT.eu 2029 fellows. The use case repository is developed within the StandICT.eu Technical Working Group Virtual Worlds. The StandICT.eu 2029 project is funded by the European Union under grant agreement no. 101213612.

References

Ambrosio, R. & Widergren, S. (2007). A framework for addressing interoperability issues. IEEE PES General Meeting. https://doi.org/10.1109/PES.2007.385817

AOUSD (2025). USD Core Specification, Version 1.0.1. *Alliance for OpenUSD*.

Bilham, G. (2026). Spatial Systems Atlas. Built for the Metaverse Standards Forum Infrastructure Working Group. https://github.com/openspatials/spatial-systems-atlas; https://openspatials.com/msf/map/

Boucher, P., Somia, T., Barreda Ángeles, M., Hupont Torres, I. & Schade, S. (2026). Observing Virtual Worlds: Structured assessment of virtual worlds applications. JRC145804. *Publications Office of the European Union*. https://doi.org/10.2760/2993248

CEN-CENELEC-ETSI SG-CG (2012). Smart Grid Reference Architecture. CEN-CENELEC-ETSI *Smart Grid Coordination Group.*

DIN (2016). DIN SPEC 91345:2016-04, *Reference Architecture Model Industrie* 4.0 (RAMI4.0). Beuth.

DKE/DIN & VDE (2015). The German Standardization Roadmap Smart City, Version 1.0.

European Commission (2011). M/490 Standardization Mandate to European Standardisation Organisations (ESOs) to support European Smart Grid deployment.

Gottschalk, M., Uslar, M. & Delfs, C. (2017). The Use Case and Smart Grid Architecture Model Approach. *SpringerBriefs in Energy*.

IEC (2008). IEC PAS 62559:2008, IntelliGrid methodology for developing requirements for energy systems.

IEC (2013). IEC 62264-1:2013, Enterprise-control system integration – Part 1: Models and terminology.

IEC (2015). IEC 62559-2:2015, Use case methodology – Part 2: Definition of the templates for use cases, actor list and requirements list.

IEC (2017). IEC 62559-3:2017, Use case methodology – Part 3: Definition of use case template artefacts into an XML serialized format.

IEC (2019). IEC TR 62559-1:2019, Use case methodology – Part 1: Concept and processes in standardization.

IEC 2025. Industrial metaverse. IEC STTR, Market Strategy Board.

IEEE (2026). IEEE P2048, Standard for Metaverse: Terminology, Definitions, and Taxonomy. Draft standard, active project. IEEE SA.

ISO (2016). ISO 14224:2016, Petroleum, petrochemical and natural gas industries – Collection and exchange of reliability and maintenance data for equipment.

ISO (2017). ISO 16642:2017, Computer applications in terminology – Terminological markup framework.

ISO (2019). ISO 26162-1:2019, Management of terminology resources – Terminology databases – Part 1: Design.

ISO (2024). ISO 16739-1:2024, Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries – Part 1: Data schema.

ISO/IEC (2022). ISO/IEC 12113:2022, Information technology – Runtime 3D asset delivery format – Khronos glTF 2.0.

ISO/IEC (2026). ISO/IEC FDIS 24931-1, Information technology – Metaverse – Part 1: Concepts, characteristics and technology. Final Draft International Standard.

ISO/IEC JTC 1/SC (24/2025). ISO/IEC PWI 24931-3, Information technology – Metaverse – Part 3: Use cases. Preliminary work item.

ISO/IEC JTC 1/SC 24 (2026). ISO/IEC PWI 24931-2, Information technology – Metaverse – Part 2: Framework and architecture. Preliminary work item.

ISO/IEC JWG 12 (2026). ISO/IEC AWI 25689, Use case type classification for VR/AR/MR systems integration; ISO/IEC PWI 26952, Use Cases and Design Guidelines for VR/AR/MR Based Learning, Education, and Training. Work items of the joint working group of ISO/IEC JTC 1/SC 24 and SC 36.

ISO/IEC/IEEE (2022). ISO/IEC/IEEE 42010:2022, Software, systems and enterprise – Architecture description.

ISO/IEC/IEEE (2025). ISO/IEC/IEEE 39274-1-1:2025 (IEEE 9274.1.1-2023), Learning technology – JSON data model format and RESTful web service for learner experience data tracking and access – Part 1-1: xAPI using JSON serialization and RESTful data transport.

ITU, UNICC & Digital Dubai (2025). Citiverse Use Case Taxonomy: Global Insights and Implementation Pathways. *Global Initiative on Virtual Worlds and AI*. ISBN 978-92-61-41611-9.

ITU-T FG-MV (2024). Focus Group on metaverse: technical reports and technical specifications FGMV-01 to FGMV-52. *ITU*.

Kuchenbuch, R. A., Vereno, D., Niemann, L., Werth, O., Lehnhoff, S. & Sauer, J. (2026). The Impact of Artificial Intelligence on Quality in Smart Grid Requirements Engineering. IEEE Access 14. https;//doi.org/10.1109/ACCESS.2026.3654443

Metaverse Standards Forum (2025). Enabling Shared KYC/AML Verification Across the Metaverse (MSF2024-001-REPSKA); Identity Verification for Digital Asset Creators (MSF2024-001-IDVER); Interoperable User Agreement Compliance across Metaverse Platforms (MSF2024-001-UAC). *Use cases, version* 1.0.

Runde, C., Gupta, A. R. & Khanna, S. (2026). Navigating XR Standardization: A Living, Multidimensional Standards Register for XR and Virtual Worlds. *Submitted to EuroXR 2026*. Register: https://www.vdc-fellbach.de/infrastruktur-assets/xr-metaverse-normenregister

StandICT.eu TWG Virtual Worlds (2026). Multi-Dimensional Classification System for XR, Metaverse and Immersive Technology Standards (v2.0); Virtual Worlds Standards Landscape Report (v28). *Internal working documents*.

Uslar, M., & Engel, D. (2015). Towards generic domain reference designation: How to learn from smart grid interoperability. *DA-Ch Energieinformatik*, *1*, 1-6.

Virtual Worlds Association (2025). Strategic Research and Innovation Agenda for the Virtual Worlds Partnership, Version 1.0.

A Virtual Reality System for Pericardiocentesis Training

Marta Goyena[1], Matteo Dal Magro[1], Lucía Curiel[2], Sergio Rico[2], María Angustias Peinado[2], Alberto Rubio[3], Manuel Quintana[2], Jesús Gutiérrez[1], and Rodrigo García[1]

[1]*ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain*

[2]*Centro de Simulación Clínica SimuPaz, Hospital Universitario La Paz, Madrid, Spain*

[3]*Hospital Universitario HM Montepríncipe, HM Hospitales, Madrid, Spain*

Corresponding author: marta.goyena@upm.es

Keywords: Virtual reality, medical simulation, interaction

Introduction

Pericardiocentesis is a medical procedure used to remove fluid from the pericardial space and is commonly performed when a pericardial effusion causes significant symptoms or cardiac tamponade, which is a potentially life-threatening condition (Vakamudi et al. 2017, Alerhand et al. 2022). The procedure involves several steps, including patient preparation, needle insertion, and catheter insertion. Correct execution of these steps requires both knowledge of the procedure and precise manipulation of the instruments (Alerhand et al. 2022, Flint et al. 2020).

Despite its clinical relevance, pericardiocentesis is a relatively infrequent procedure, which may result in limited opportunities for healthcare professionals to acquire and maintain the required procedural skills (Lord et al. 2021). Simulation-based training provides an opportunity to practice these skills in a safe and controlled environment without risk to patients. Virtual Reality (VR) has emerged as a promising approach for medical simulation, enabling immersive and repeatable training environments while reducing some of the infrastructure requirements associated with high-fidelity physical simulators (Barteit et al. 2021, Liaw et al. 2023).

Previous research has explored the use of VR for pericardiocentesis training and has demonstrated its potential for supporting procedural learning and familiarization with the steps of the procedure (Rubio-López et al. 2025). However, the way users interact with the virtual environment may influence both the performance of the procedure and the overall training experience. Different interaction technologies provide different levels of physical interaction and feedback, which may be especially relevant for procedures requiring precise motor actions.

In this context, this work presents a VR-based pericardiocentesis training system supporting three interaction modalities: VR controllers, hand tracking, and haptic gloves. The system reproduces the procedural workflow within the same virtual environment, providing a common environment for comparing the interaction techniques. An upcoming experimental study will evaluate the effects of these modalities on procedural performance, workload, stress, presence, usability, and overall user experience.

VR Pericardiocentesis Training System

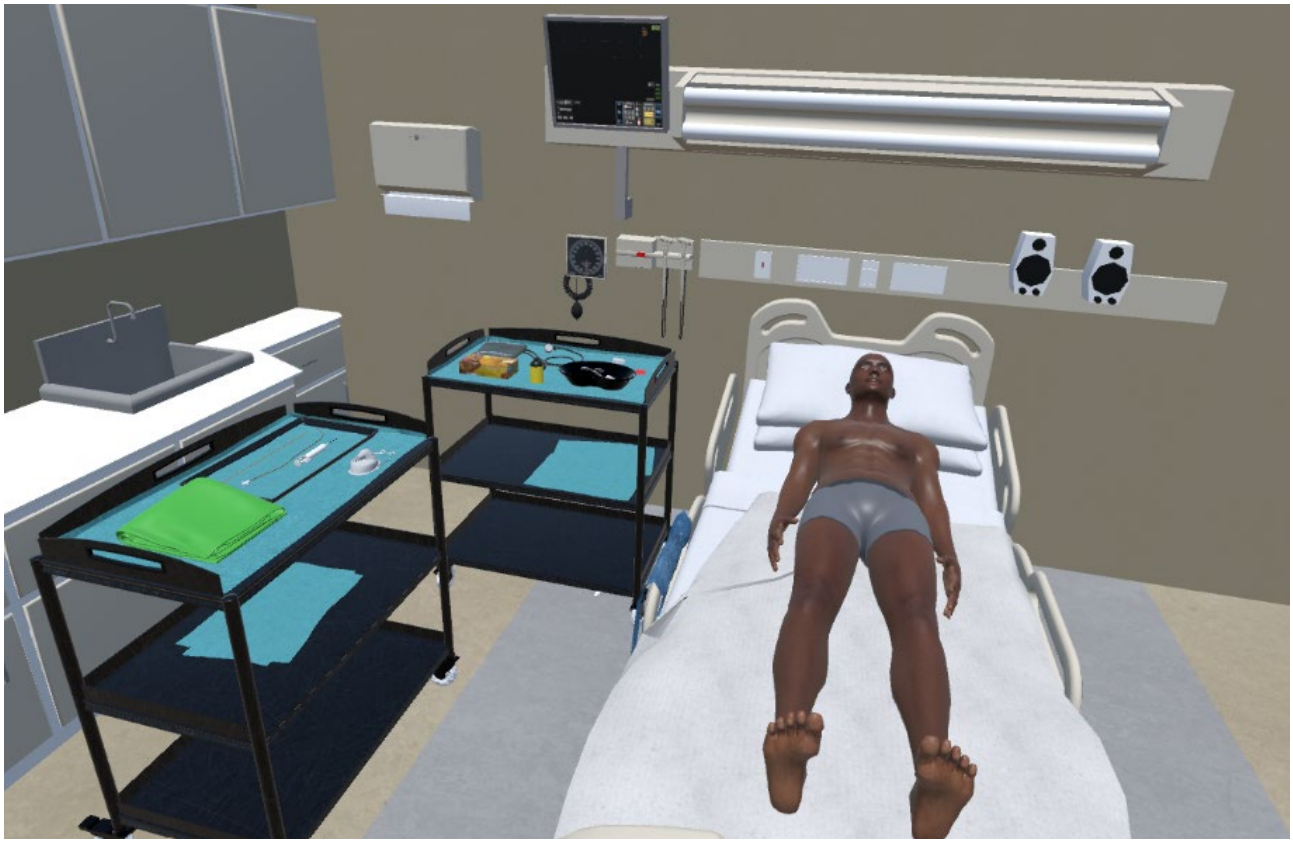


***Figure 117**. Virtual training scenario with a patient on a bed and materials required for the procedure in the trolleys.*

The VR-based training system for pericardiocentesis is based on a simulation presented in previous work (Rubio-López et al. 2025). It has been developed using Unity and deployed on an HTC VIVE XR Elite head-mounted display. Fig. 1 presents the main elements of the virtual scenario. The patient is positioned on the hospital bed. Two instrument trolleys contain the materials required for patient preparation and pericardiocentesis. A monitoring screen provides information about the patient's physiological state throughout the simulation, allowing the user to monitor the patient's condition during the procedure.



***Figure 118**. A user performing key tasks of the pericardiocentesis procedure, including patient preparation, patient monitoring, and needle insertion.*

As shown in Fig. 2, the simulation reproduces the main steps of the pericardiocentesis procedure, including patient preparation, positioning of the monitoring devices and required materials, monitoring of the patient's health state, and insertion of the needle, guidewire, and drainage catheter.

The virtual environment allows users to perform the procedure in a safe and controlled setting without real-world consequences.

The system supports three interaction modalities:

- **VR controllers:** Provide tracked-controller interaction with vibrotactile feedback at selected points of the procedure.
- **Hand tracking:** Allows users to interact directly with virtual objects using their hands.
- **Haptic gloves**: Use SenseGlove Nova 2 gloves to enable hand interaction with virtual objects and provide force and vibrotactile feedback during procedural interactions.

Future work: Planned Evaluation

A comparative user study is planned with healthcare professionals recruited from two hospitals. The study will use a between-subject design, with participants assigned to one of three interaction modalities: VR controllers, hand tracking, or haptic gloves. The study will consist of the following phases:

1. **Pre-study assessment:** Participants will complete informed consent procedures and a pre-study questionnaire.
2. **System introduction:** Participants will complete a short training scenario to learn how to interact with the system and their assigned interaction modality.
3. **VR training task:** Participants will perform the pericardiocentesis procedure using one of the interaction modalities.
4. **Data collection:** During the simulation, objective performance metrics, behavioral data, and physiological data will be recorded.
5. **Post-study evaluation:** Participants will complete standardized questionnaires.
6. **Data analysis:** The data collected will be analyzed.

The evaluation will combine subjective, objective, physiological, and behavioral measures. Subjective assessments include standardized questionnaires evaluating cognitive load, stress, presence, usability and cybersickness, and evaluation of the training outcomes. Objective measures will include task completion time, task-related events, and interaction metrics derived from controller, hand-tracking, and haptic-glove data collected during the simulation. In addition, physiological data, including eye tracking, electroencephalography, electromyography, and respiration signals, will be recorded.

The collected data will be analyzed to determine how different interaction modalities affect procedural performance, cognitive load, physiological responses, cybersickness, presence, and user experience.

Acknowledgement

The work is partially supported by the project PID2023-148922OA-I00 (EEVOCATIONS) funded by MCIN/AEI /10.13039/501100011033 of the Spanish Government and by the Celtic-Next project TIC-20240154 (DAWN6G) funded by the Spanish Administration agency CDTI.

References

Vakamudi, S., Ho, N. & Cremer, P.C. (2017). Pericardial Effusions: Causes, Diagnosis, and Management. *Progress in Cardiovascular Diseases* 59(4), 380–388. https://doi.org/10.1016/j.pcad.2016.12.009

Alerhand, S., Adrian, R.J., Long, B. & Avila, J. (2022). Pericardial tamponade: A comprehensive emergency medicine and echocardiography review. *American Journal of Emergency Medicine* 58, 159–174. https://doi.org/10.1016/j.ajem.2022.05.001

Flint, N. & Siegel, R.J. (2020). Echo-Guided Pericardiocentesis: When and How Should It Be Performed? *Current Cardiology Reports* 22(8), 71.

Lord, S., Lord, G. & Geary, S.P. (2021). A Novel, Low-cost, Low-fidelity Pericardiocentesis Teaching Model. *Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health* 22(4). https://doi.org/10.5811/westjem.2021.3.49876

Barteit, S., Lanfermann, L., Bärnighausen, T., Neuhann, F. & Beiersmann, C. (2021). Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. *JMIR Serious Games* 9(3), e29080.

Liaw, S. Y., Sutini, Chua, W. L., Tan, J. Z., Levett-Jones, T., Ashokka, B., ... & Ignacio, J. (2023). Desktop Virtual Reality Versus Face-to-Face Simulation for Team-Training on Stress Levels and Performance in Clinical Deterioration: a Randomised Controlled Trial: Liaw et al. *Journal of general internal medicine*, *38*(1), 67-73.

Rubio-López, A., García-Carmona, R., Zarandieta-Román, L., Rubio-Navas, A., González-Pinto, Á., & Cardinal-Fernández, P. (2025). Analysis of stress responses in medical students during simulated pericardiocentesis training using virtual reality and 3D-printed mannequin. *Scientific reports*, *15*(1), 7645.

Rubio-López, A., García-Carmona, R., Zarandieta-Román, L., Rubio-Navas, A., González-Pinto, Á., & Cardinal-Fernández, P. (2025). Innovative approaches to pericardiocentesis training: a comparative study of 3D-printed and virtual reality simulation models. *Advances in Simulation*, *10*(1), 19. https://doi.org/10.1186/s41077-025-00348-0

Composing RFC 8628, RFC 7636 and OpenID Connect for Standalone XR Training: Sign-In Outside the Application and Per-User xAPI Attribution

Arjun Rai Gupta

Virtual Dimension Center w. V., Fellbach, Germany

Corresponding author: arjunrai.gupta@vdc-fellbach.de

Keywords: Extended reality; authentication; OAuth 2.0; device authorization grant; xAPI; identity federation; vocational training

Problem context and motivation

XR2ACH, a project in the German InnoVET PLUS programme, develops a refrigeration-technology training application for standalone extended-reality (XR) headsets, initially targeting the Meta Quest 3/3S platforms. It trains refrigeration, air-conditioning and heat-pump technicians in collaborative scenarios, and logs learner activity as Experience API (xAPI) statements. This paper reports its sign-in flow.

Standalone XR headsets are a poor fit for the login flows the web has settled on. They have no physical keyboard, and the OAuth guidance for native applications (RFC 8252, Section 8.12; Denniss & Bradley 2017) rules out embedded web-views for the OpenID Connect (OIDC; Sakimura et al. 2014) authorization-code flow: sign-in must happen in the platform browser, and its result must find a way back into the application.

The simpler alternatives are a shared account, a credential compiled into the build, or no sign-in at all. None of them tells learners apart, so the Learning Record Store (LRS) holds records it cannot attribute to anyone: an aggregate that supports neither individual feedback nor any statement about what a named learner has achieved.

A password field drawn by the application keeps attribution but gives away the credential. A browser draws its address bar and padlock outside the page, so a user can check a login screen against them; a headset application draws the entire field of view and leaves nothing to check. A legitimate password prompt in a headset therefore looks exactly like an imitation of the identity provider's page. The credential passes through application code, drawing a virtual keyboard for instance, before it reaches any authentication endpoint. Deep-link callbacks add a third surface: nothing prevents a second installed application from registering the same custom uniform resource identifier (URI) scheme (RFC 8252, Section 8.1). The remedy is a flow in which the application never holds the credential at all.

State of the art, research gap and methodology

The device grant. RFC 8628 (Denniss et al. 2019) specifies the OAuth 2.0 Device Authorization Grant for exactly this class of device: input-constrained, or lacking a suitable browser. Its own security considerations name the residual risk, remote phishing: an attacker starts the flow on a device they control and persuades a victim to enter the code (RFC 8628, Section 5.4). **Proof Key for Code Exchange (PKCE;** RFC 7636; Sakimura et al. 2015) binds an authorization request to the client instance that made it — introduced for public mobile clients that cannot keep a secret, which is precisely an Android application package (APK) on a headset a trainee holds. The OAuth security best current practice requires it of public clients using the authorization code grant (RFC 9700, Section 2.1.1; Lodderstedt et al. 2025). RFC 8628 does not define PKCE for the device grant; we add it as defence in depth against a leaked device code, and Keycloak 26.1 enforces it on the device endpoint. **xAPI (**ADL 2016), standardised as IEEE 9274.1.1-2023 and ISO/IEC/IEEE 39274-1-1:2025 (IEEE 2023), defines learning-record statements and the store that holds them. A statement names an *actor*; under one shared credential the LRS cannot tell who wrote it, and any holder of the credential can write as anyone.

The gap. Neither practice nor research offers such a flow. Some commercial training platforms sign users in inside the application, with a server address, e-mail and password: the application-drawn password field described above. Research on authentication in virtual reality (VR) recognises who is wearing a headset, from how the head trembles (He et al. 2025) or how sound travels through it (Wang et al. 2023). Like a fingerprint on a phone, this unlocks the device for its enrolled user; it does not sign the learner in to an institution's accounts. It is also intrusive. Such signals are biometric data, and the EU General Data Protection Regulation prohibits processing them to identify a person unless an exception applies (European Union 2016, Art. 9); a biometric trait, once compromised, cannot be reissued like a password; and head and hand motion alone has singled out one user among more than 50,000 with 94% accuracy (Nair et al. 2023). Each standard is mature and each is documented in isolation. What is not documented is their composition for standalone XR, and in particular the step that makes the shared credential unnecessary: native OIDC bearer validation *at the record store*, so that no intermediary holds a credential on the learner's behalf. It carries a trap: the audience parameter is the store's own origin URL, not the OAuth client identifier, and an identity provider will not place it in the token's aud claim without an explicit audience mapper.

Methodology. We implemented the composition, then tested it in the field: a session at a vocational training centre in September 2026 with seven apprentices and five project staff on eight headsets in one workshop room, running against the production identity provider and LRS. Participants gave informed consent. We report what the LRS and the identity provider's own event log recorded, together with a post-session questionnaire with a dedicated sign-in block.

Implementation and results

On the server side, Keycloak is the identity provider and LRSQL the LRS, both behind a single reverse proxy (Figure 1); the Unity application on Quest is their client. The client's xAPI pipeline uses the Unity package OmiLAXR (Görzen et al. 2024). The authentication layer beneath it is this paper's contribution, replacing the package's shared-credential model with a per-user bearer token. The application is registered as a **public** client with PKCE S256 required and the device grant enabled; it holds no client secret, because a client secret shipped inside an APK is not a secret.

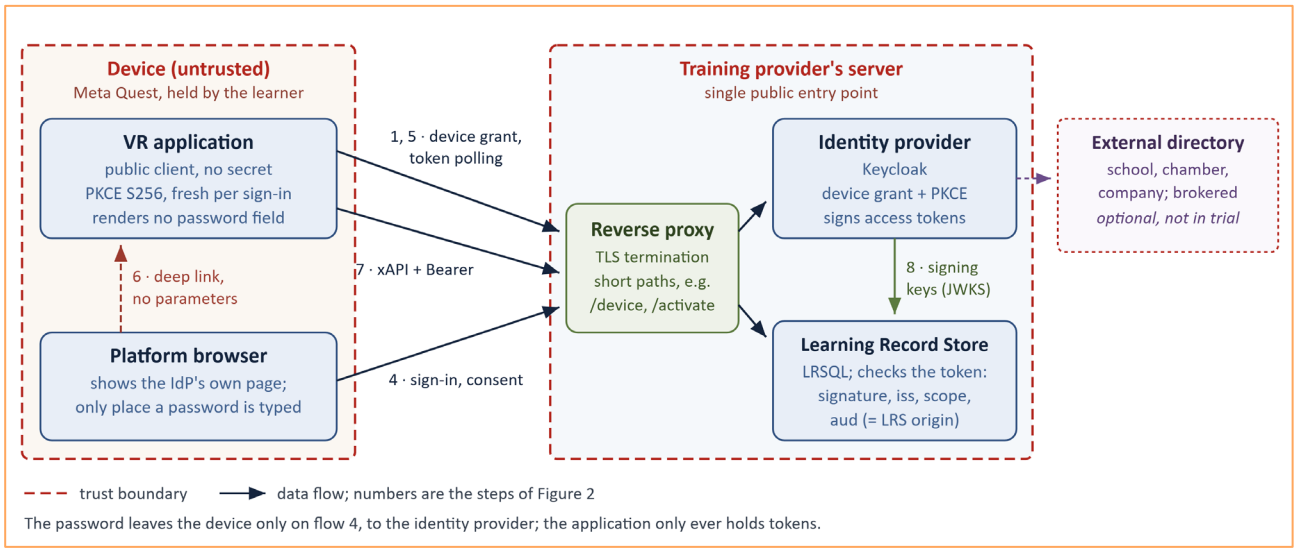


Figure 1. *Trust boundaries and data flows.*

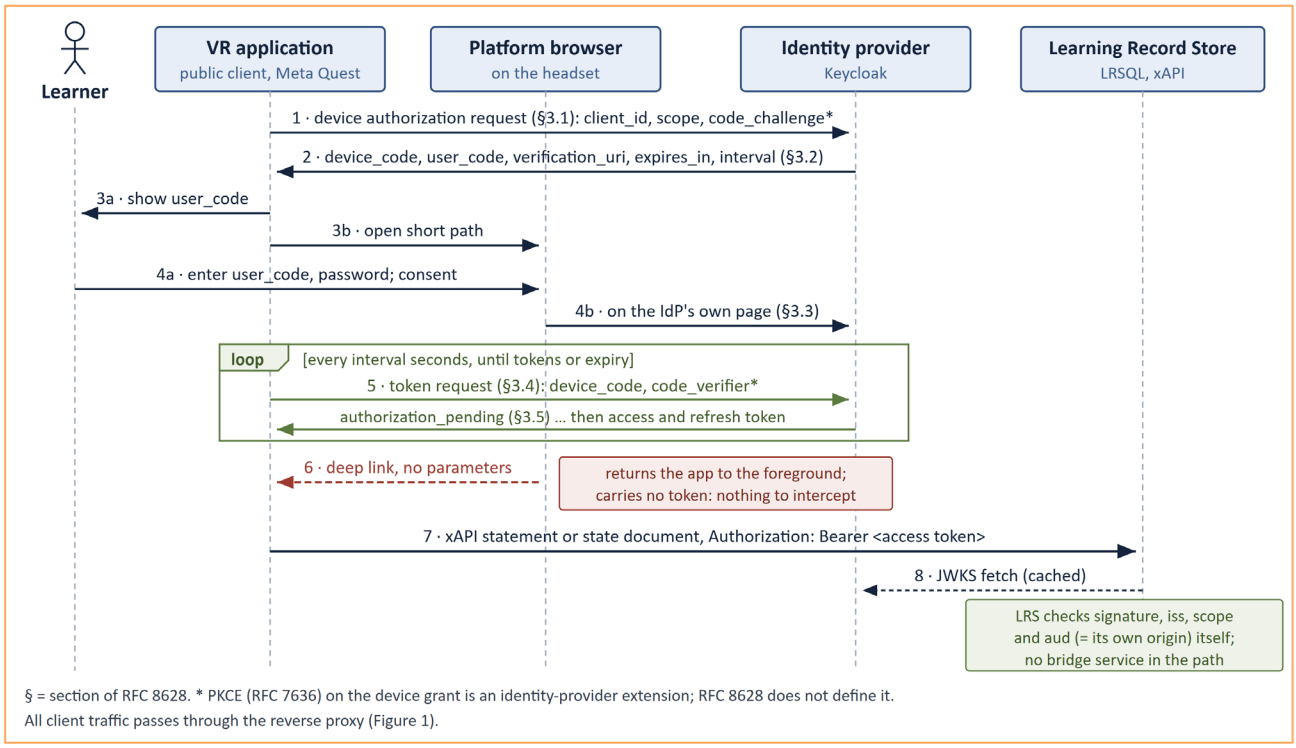


Figure 2. *The device-grant login and the reuse of its access token for xAPI writes.*

Sign-in. As Figure 2 shows, the application generates a fresh PKCE pair and requests a device code; the user code and verification URI are shown in the headset and the platform browser is opened at that URI (Figure 3). Because the realm's verification URI is long, the reverse proxy maps short paths such as /device and /activate onto it, as RFC 8628 (Section 3.2) asks for one that is short and easy to remember.

Per-user attribution. The access token is reused as Authorization: Bearer for xAPI writes. The LRS validates it itself and records the signed-in account as each statement's authority. The actor is still built by the client, with one rule for statements and the xAPI State document: the real e-mail address where the identity provider supplies one, and a pseudonymous address derived from the subject claim where it does not. The LRS does not check the actor against the token.



***Figure 3**. Sign-in happens in the Meta browser, outside the application: the learner authenticates on the identity provider's own page (credentials redacted) and grants the application the access it requests.*

What the boundary removes. The application never renders a password field, so it has no legitimate prompt to imitate, no credential passes through its code, and the deep-link callback carries no token. A malicious application could still draw a fake browser panel; what changes is that learners can be told never to type a password inside an application. Authentication is delegated to the identity provider's own page, so any directory that provider can broker works without a change in the application. The trial used local accounts only. Remote phishing (RFC 8628, Section 5.4) remains; the consent page names the requesting application, and a phished token is scoped to writing learning records.

Evaluation, impact and outlook

In the field trial the flow worked at classroom scale. Eleven loggedIn statements arrived from **eight distinct identities** (five trainee and three staff accounts) across eight headsets, and **eight per-user state documents** were written, each under its signed-in account's authority and with no shared credential. Every sign-in attempt in the identity provider's log eventually succeeded.

All six questionnaire respondents reported signing in, five at the first attempt and one at the second. Asked whether they could have signed in unaided, they gave a mean of 3.83 on a five-point scale, three of them at the midpoint. With six respondents these figures are descriptive only.

Three steps remain: testing on standalone Android headsets beyond Meta Quest; replacing the fixed custom scheme of the deep-link return, which makes Android ask which of two installed builds to open, with the build-specific schemes already implemented behind a flag, or with verified https links (RFC 8252, Section 7.2); and, for shared headsets, binding the actor to the token at the LRS and ending the identity provider's browser session at sign-out.

Standalone XR does not need a bespoke authentication design. RFC 8628, RFC 7636 and OIDC bearer validation at the record store compose into a flow that needs no text entry inside the application, ships no secret, can federate to the identity provider a training institution already runs, and ties every record to a signed-in account.

Acknowledgement

XR2ACH is funded by the German Federal Ministry of Education, Family Affairs, Senior Citizens, Women and Youth (BMBFSFJ) under the InnoVET PLUS programme (grant no. 21IVP043A-C), which is administered by the Federal Institute for Vocational Education and Training (BIBB). The author is responsible for the content of this publication. We thank the apprentices and staff of the Bundesfachschule Kälte-Klima-Technik in Leonberg who took part in the trial.

References

ADL (2016). Experience API (xAPI), Version 1.0.3. *Advanced Distributed Learning Initiative*.

Denniss, W. & Bradley, J. (2017). OAuth 2.0 for Native Apps. RFC 8252 (BCP 212). *Internet Engineering Task Force*. https://doi.org/10.17487/RFC8252

Denniss, W., Bradley, J., Jones, M. & Tschofenig, H. (2019). OAuth 2.0 Device Authorization Grant. RFC 8628. *Internet Engineering Task Force*. https://doi.org/10.17487/RFC8628

European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). *Official Journal of the European Union* L 119, 1–88.

Görzen, S., Schroeder, U. & Heinemann, B. (2024). Towards using the xAPI specification for Learning Analytics in Virtual Reality. RWTH-2024-10351. Aachen: *RWTH Aachen University*. https://publications.rwth-aachen.de/record/996055

He, Z., Ran, F., Chen, J., Gu, Y., He, K., Du, R., Jia, J. & Wu, C. (2025). HT-Auth: Secure VR Headset Authentication via Subtle Head Tremors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9(3). https://doi.org/10.1145/3749980

IEEE (2023). IEEE 9274.1.1-2023, Standard for Learning Technology – JavaScript Object Notation (JSON) Data Model Format and Representational State Transfer (RESTful) Web Service for Learner Experience Data Tracking and Access. Adopted as ISO/IEC/IEEE 39274-1-1:2025.

Lodderstedt, T., Bradley, J., Labunets, A. & Fett, D. (2025). OAuth 2.0 Security Best Current Practice. *RFC 9700. Internet Engineering Task Force*. https://doi.org/10.17487/RFC9700

Nair, V., Guo, W., Mattern, J., Wang, R., O'Brien, J. F., Rosenberg, L. & Song, D. (2023). Unique Identification of 50,000+ Virtual Reality Users from Head & Hand Motion Data. *32nd USENIX Security Symposium, 895–910.*

Sakimura, N., Bradley, J. & Agarwal, N. (2015). Proof Key for Code Exchange by OAuth Public Clients. RFC 7636. *Internet Engineering Task Force*. https://doi.org/10.17487/RFC7636

Sakimura, N., Bradley, J., Jones, M., de Medeiros, B. & Mortimore, C. (2014). OpenID Connect Core 1.0. *OpenID Foundation*.

Wang, R., Huang, L. & Wang, C. (2023). Low-effort VR Headset User Authentication Using Head-reverberated Sounds with Replay Resistance. *2023 IEEE Symposium on Security and Privacy* (SP), 3450–3465. https://doi.org/10.1109/SP46215.2023.10179367

Author Index, Tables of Figures & Lists of Tables

Author Index

Abad
Roberto, ... 160
Acampora
Ivan, ... 70
Alberto
Alberto, ... 131
Alcañiz
Mariano, ... 131
Alesani
Andrea, ... 82
Amano
Toshiyuki, ... 191
Amestoy-Alonso
Beñat, ... 258
Ampatzoglou
Aikaterini, ... 245
Arevshatyan
Sona, ... 53
Argelaguet
Ferran, ... 356
Arlati
Sara, ... 113, 117, 364
Arpaia
Pasquale, ... 117
Audrain
Arthur, ... 356
Azizi
Amir, ... 287
Bailly
Charles, ... 179, 232
Baldoni
Sara, ... 152
Balogh
Tibor, ... 275, 279
Banakou
Domna, ... 228
Banerjee
Natasha Kholgade, ... 379
Sean, ... 379
Barà
Chiara, ... 364
Barbero
David, 3 93
Battistel
Laura, ... 86
Battisti
Federica, ... 152
Bayer
Andreas, ... 41
Bellantuono
Giuseppe, ... 144
Giuseppe Maria, ... 66, 144, 173, 176
Benito
Ignacio, ... 393
Biffi
Emilia, ... 364
Biradar
Pramod, ... 169
Boardman
David, ... 66
Boile
Maria, ... 156
Bolech
Michael, ... 50
Boletsis
Costas, ... 91
Bolognini
Silvia, ... 121
Bonalumi
Samuele, ... 53
Bonino
Brigida, ... 317, 320
Borrego
Adrián, ... 291
Bortolamei
Massimo, ... 70
Bouvier
Camille, ... 58

Bouzón
Irene, ...160
Butcher
Matthew, ...173
Böhm
Sandra, ...198
Cabrera
Julián, ...352
Calvis
Isaac, ...271
Calvo
Gabriela, ...291
Campagnoli
Carlo, ...224
Campion
Simon, ...66, 144, 173
Campos
Gloria,.... 291
Caputo
Ciro, ...70
Carli
Marco, ...152, 155
Carmona
Rodrigo García, ...344
Carreres-Fluixa
Josep, ...324
Carrion-Plaza
Alicia, ...104
Castet
Julien, ...179, 232
Castiglioni
Giulia Maria, ...113
Castillo
Alejandro, ...22
Cattaneo
Leonardo, ...53
Chadjiminas
Chrysostomos, ...304
Chan
Weng, ...144
Charalambous
Panayiotis, ...287, 304
Chardonnet
Jean-Remy, ...330
Chiappini
Mattia, ...364
Chrysanthou
Yiorgos, ... 287
Civelek
Turhan, ... 135
Colombo
Vera, ... 121
Cortés,
Carlos, ... 352
Couëdel
Romane, ... 139
Cox
James, ... 379
Croce
Raffaele, ... 18, 86
Cuevas-Rodríguez
María, ... 368
Curiel
Lucía, ... 404
Cybulski
Artur, ... 383
D'Errico
Giovanni, ... 375
Dal Magro
Matteo, ... 404
Danglade
Florence, ... 330
David
Constantine, ... 208
de la Campa Crespo
Mercè Álvarez, ... 304
Dembani
Rahool, ... 283
Den Butter
Gijs, ... 335
Díaz
César, ... 352
Díaz-Marín
Carmen, ... 348
Dini
Silvia, ... 317
Dix
Martin, ... 198, 263, 267
Domingo
Julia Ajenjo, ... 218
Domingo-Ajenjo
Julia, ... 131

Donegan
Tony, ...258, 271
Dorociak
Radoslaw, ...144
dos Santos Estevães
Kátia, ...271
Du Mutel de Pierrepont Franzetti
Iris David, ...86
Egea
Álvaro, ...335
Eivers
Eemer, ...13, 17
Ellis
Harry, ...66
Esposito
Antonio, ...117
Fernández
Sergi, ...335
Fernandez Capo
Maria, ...245
Fernández-Caballero
Antonio, ...312
Ferrarotti
Anna, ...152
Fibi
Bastian, ...62
Fletcher
Simon, ...26
Foulonneau
Anthony, ...240
Franke
Jörg, ...41, 169
Fuhrmann
Arnulph, ...135
Fuller
Aston, ... 66
Fumarola
Eugenio, ...18
Fürst
Jens, ...41
Galdieri
Fortuna, ...117
Gámiz
Maria, ...245
García
Arturo S., ...312
Bruno Porras, ... 245
Rodrigo, ... 339, 404
García-Arocas
D., ... 389
Garcia-Ballesteros
Marta, ... 250
Georgiou
Christy, ... 173
Georgopoulos
Konstantinos, ... 50
Giannini
Franca, ... 317
Gilabert-Estelles
J., ... 389
Giliberti
Vincenzo, ... 18
Gimeno
J., 3..... 89
Gkanias
George, ... 156
Gkiosis
Theodoros, ... 360
Gledhill
Duke, ... 26
Gonzalez Galvez
Monica, ... 360
González-Sánchez
Noelia, ... 312
Gonzalez-Toledo
Daniel, ... 372
González-Toledo
Daniel, ... 368
Goyena
Marta, ... 404
Grandi
Fabio, ... 202
Grasso
Valentina, ... 364
Guerra
Guerra, ... 152
Guida
Mariano, ... 74
Gupta
Arjun Rai, ... 183, 398, 408

Gutiérrez
Antonio, ...37, 40, 155, 214, 249, 393
Jesús, ...393, 404
Gutierrez-Parera
Pablo, ...368, 372
Haghtalab
hanif, ...46
Haschke
Sven, ...41, 169
Hayes
Hannah, ...224
Hekiert
Daniela, ...383
Helin
Kaj, ...82
Hierck
Beerend P., ...236
Hoek
Krista, ...109
Hosseini
Ghazal, ...46
Howard
Thomas, 1 39
Iatrou
Panagiotis, ...283
Ibañez
Celia, ...352
Ibragimov
Elza, ...379
Igras-Cybulska
Magdalena, ... 383
Inches
Giacomo, ...70
Jaschinksi
Christina, ...109
Jens
Fürst, ...169
Jiménez
Javier Andión, ...344
Johnson
Shaun, ...144
Kara
Peter A., ...275, 279
Karjalainen
Jaakko,..... 82
Kárpáti
Zoltán,..... 156
Katifori
Akrivi, ... 271
Khanna
Saahil, ... 127, 183
Khomyn
Olha, ... 18
Klimant
Franziska, ... 198
Knopp
Sebastian, ... 187, 190, 263, 266, 267, 270
Koltsakidis
Savvas, ... 206, 208, 211
Kruse
Rolf, ... 398
Kulathunga
Chathuni, ... 379
Kunz
David, ... 41, 169
Kyriakidis
Ioannis Filippos, ... 321
Köhler
Robert, ... 41, 169
Lacoche
Jérémy, ... 240
Larrosa-Quesada
Jérémy, ... 245
Leech
Adam, ... 13
Leoncini
Paolo, ... 74
Levers
Andrew, ... 144, 176
Andy, ... 173
Laurence, ... 144
Litskevich
Dzianis, ... 66
Llorens
Roberto, ... 214, 291
Longstaff
Andrew, ... 26
Lorenz
Mario, ... 50, 187
Lupinetti
Katia, ... 317

Maaroufi
Fahd, ... 330
Mahmoud
Moatasim, ... 283, 296
Majdak
Magdalena, ... 383
Malerba
Giorgia, ... 364
Mariano
Alcañiz, ... 5, 74, 131, 214, 218
Marin-Morales
Javier, ... 131
Marín-Morales
Javier,..... 375
Martos
Marc, ... 335
Mastropietro
Alfonso, ... 364
Matiello
João, ... 195
Matteo Dinoi
Leonardo,..... 375
Mazeas
Damien,..... 240
McDermott
Geraldine, ... 164
Melese
Abigya Abayneh, ... 271
Michael-Grigoriou
Despina, ... 228
Millán
Luis, ... 291
Minissi
Maria Eleonora, ... 131, 218
Mirafiori
Matteo, ... 86
Molina
José P., ... 312
Montagud
Mario, ... 335
Monzón
Àlvaro, ... 22
Morales
Javier Marín, ... 218
Morán
Marcos, ... 339
Mora-Serrano
Javier, ... 300, 303
Moroni
Giacomo Giovanni, ... 113
Mossali
Elena, ... 53
Mousavi
Sayed Mahed, ... 18
Moysiadis
Georgios, ... 206
Moysiadis, ... 208
Muñoz-La Rivera[2]
Felipe, ... 300
Myers
Georg,..... 173
Najm
Ali, ... 228
Nardi
Alessandro, ... 117
Navas-Medrano
Samuel, ... 100, 245, 250, 324
Németh
Katalin, ... 156
Nocera
A., ... 389
Nossa
Roberta,..... 364
Nylund
Maria Emine, ... 91
Ors
Teodoro, ... 214
Own
Arwa, ... 50, 187
Oxley
James, ... 173
Pagano
Marie-Claire, ... 173
Page
Álvaro, ... 291
Palanca
Maria, ... 214
Palma-Ruano
B., ... 389
Panagiotopoulos
Thomas, ... 321

Papanikolaou
Athanasios, ... 296
Parin
Riccardo, ... 86
Parma
Stefano, ... 53
Passarello
Noemi, ... 236
Peinado
María Angustias, ... 404
Peña-Monferrer
Carlos, ... 348
Pepino
Alessandro, ... 375
Pérez
Abril, ... 339
Pablo, ... 393
Pérez de Arenaza Pozo
Teresa, ... 344
Perez-Caballero
Maria, ... 250
Pérez-Folgado
I., ... 389
Perret
Jerome, ... 139
Peruzzini
Margherita, ... 202
Pestelli
Matteo, ... 152, 155
Petrou
Vasilis, ... 360
Phelan
Ivan, ... 104
Pignoni
Giovanni, ... 202
Plaza Cuenca
Carlos, ... 344
Ploch
Wiktoria, ... 144
Pons
Patricia, ... 100, 245, 250, 324
Poyer
Maëlan, ... 86
Prencipe
Giuseppe, ... 121
Prillard
Ophelia, ... 91
Proboste-Martínez
Mathias, ... 300
Pupillo
Chiara, ... 121
Putz
Maximilian, ... 198
Qadeer
Camran, ... 173
Quintana
Abraham Otero, ... 344
Manuel, ... 404
Reese
David, ... 263
Revesz
Richard, ... 62
Reyes-Lecuona
Arcadio, ... 368, 372
Ribeiro de Oliveira
Afonso, ... 348
Riboni
Riccardo, ... 364
Riccardi
Giuseppe, ... 18
Rickard
Tom, ... 13
Rico
Sergio, ... 404
Riera
J.V., ... 389
Riva
Giuseppe, ... 121
Rizou
Stamatia, ... 283, 296
Rizzo
Giovanna, ... 364
Rocchi
Erica, ... 152, 155
Rodriguez
Mario Covarrubias, ... 113
Nancy, ... 57, 258, 300, 371, 372, 374
Rodriguez, ... 113
Rodríguez
David, ... 339
Isabel, ... 352

Rodriguez de la Calle
Edgar,..... 258
Rojas-Rincon
Juliana, ...245
Rothwell
Craig, ...66, 173
Roussou
Maria, ...156, 271
Rubio
Alberto, ...404
Runde
Christoph, ...127, 169, 183, 398, 401, 403
Sacco
Marco, ...113, 117, 364
Sagris
Dimitrios, ...208
Salgarelli
Tommaso, ...53
Salmela
Laura, ...82
Salvatori
Daniela, ...236
Sánchez García
Javier, ...309
Sanchez-Vives
Maria V., ...223, 258, 262, 271, 274
Sarton
Elise, ...109
Saufnay
Sarah, ...148
Schiavi
Barbara, ...5, 179, 227, 232
Schobert
Marvin, ...41, 169
Schyns
Michaël, ...58, 148
Scotson
Jim, ...66
Serres
Romain, ...232
Shigeta
Ayu, ...191
Show
Fate S.,..... 26
Siafali
Maria, ...211
Silva
Nelson J.S., ... 46
Slater
Mel, ... 271
Smagas
Konstantinos,..... 360
Soler-Dominguez
José Luis, ...100, 101, 103, 245, 246, 249, 250, 324
Solheiro
Hermenegildo, ... 96
Sotiropoulos
Iason, ... 296
Stefani
Lorenzo, ... 86
Storm
Fabio Alexander, ... 364
Stulii
Alona, ... 255
Sunny
Jackson, ... 240
Świdrak
Justyna, ... 271
Tadeja
Sławomir,..... 383
Terrile
Chiara, ... 317
Thakoer
Asmanie,..... 109
Themistokleous
Kyriacos, ... 321
Tome
Gersa, ... 62
Tommaso De Paolis
Lucio, ... 375
Tompouri
Marina, ... 287
Tottszer
Tamas K., ... 275, 279
Toyoda
Keisuke, ... 96
Trillo
Alejandro, ... 218
Tsaknaki
Electra, ... 360

Tsongas
Konstantinos, ... 321
Tsouloukidis
Zisis, ... 360
Turani
Lorenzo, ... 53
Tymchenko
Maryna, ... 255
Tzetzis
Dimitrios, ... 206, 208, 211, 321
Tzimtzimis
Emmanouil K., ... 206, 208, 211, 321
Usón
Javier, ... 352
Valdes
Marcos, ... 152
Vallefuoco
Ersilia, ... 375
van Velzen
Monique, ... 109
Viapiana
Stefania, ... 18
Vikhorev
Konstantin, ... 66, 144
Villanueva
Vicente, ... 214, 217
Viscontini
Manuela, ... 152
Wagner
Markus, ... 50
Waldow
Kristoffer, ... 135
Waller
Elbert, ... 109
Westgarth
Carri, ... 173
Whyatt
Hannah, ... 66
Winkler
Sven, ... 198
Yun
Haoran, ... 356
Zay-Hta
Aye Chan, ... 66, 144
Zhang
Yang, ... 66, 144
Zibrek
Katja, ... 356

Tables of Figures

Figures in Application Track

Young People and Youth Workers engaging remotely in a safe virtual space. 14
A training session in HOPE Metaverse Training 19
Interaction flows in HOPE Metaverse Training. 20
Photomontage of Aena's training platform, integrating the user into the project's virtual reality..... 24
Trainee using Aena's training platform in real life. 24
The experimental method. 29
User's view of the training environment from the PC. 30
User's view of the training environment from the VR headset. 30
Coloured Drawing Extract 31
A sample MRT question. 32
MRT Question Difficulty 34
EST Participant Performance. 34
EST Question Difficulty Values. 35
NASA-TLX Score Distribution 35
Platform Engagement Feedback 36
ADDIE-Model according to Sturman et al. (2026) with selected tools from the XR-Upskill project. 42
Narrative examples: video-narrator-based (a), floating-avatar-based (b), exploration-based (c) ... 43
Example of Moodle integration (left) and Learning Record Store Feedback (right) 44
System architecture: layer 1 – physiological sensing; layer 2 – data streaming; layer 3 – visualizations (2D dashboard 47
Public demonstration: a large screen (65'') positioned outside the lab showed the heart model to visitors. 48
A user wearing the PLUX Blood Volume Pulse (BVP) biosensor interacts with our application, it displays a 3D heart on a light-field display while streaming the user's biosignal heart-rate data to the model. 48
Interaction devices evaluated in the C-ECO core-sorting testbed: RealWear Navigator 520 smart glasses (left) and handheld scanning setup (right). 51
CoBotXpeRience first scenario: collaborative motorbike assembly workstation. 54
CoBotXpeRience second scenario: flexible pick-and-place workstation. 55
Highloader, Pilot check-list, Maintenance, Pushback, Snow removal training & Computer Vision DT 59
Job-description panel of the VR/360° promotion module 60
Learner-perceived performance indices for the MAN module (adapted from Droeven, 2024) 60
The operator executing the recovery task while wearing the Magic Leap 2 glasses (left) and the expanded menu view showing the dedicated menus for a specific recovery request (right). 63

Demonstration of the identified cut zone on the aircraft fuselage (left) and its associated identifier (right). 64

AI-Glass digital twin architecture, from operational furnace data and spatial assets to CFD-trained surrogate prediction and streamed XR experiences. 67

The two pixel-streamed delivery modes for the AI-GLASS digital twin. 68

Quest 3 prototype: (a) manual placement of a spatial plant reference; (b) AI-based plant detection and condition classification; and (c) contextual information with qualitative detection-to-slot matching. 71

Wearable prototypes: (a) on-device detection using Spectacles; and (b) the Meta Ray-Ban Display workflow with smartphone-based inference. 72

Collaborative sessions in the CIRA VR Lab. 75

Third person view of the developed Mars VE with two astronaut users. 76

The Weart TouchDIVER Pro gloves. 77

Correspondence between the hand wearing the Weart TouchDIVER Pro glove and the virtual avatar's hand 78

The Weart TouchDIVER Pro gloves grasping capabilities. 78

The app whilst using the Freeaim VR Shoes and the user inside the walking safety frame. 79

The OpenVR/SteamVR overlay menu and the Freeaim Mobile App. 80

Field trial set-up in Bulgaria with XR C2, several assets, cameras and drones 82

The Mobile XR-Based C2 Tool during trial 83

Left: 3D map showing tracks, detections, assets and video streaming Right: Showing additional data when icon clicked 84

Task-specific aircraft model and replicated controller interface used in the VR simulator. 87

Real and virtual Val Gardena terrain (top) and configurable clear, fog and snow conditions (bottom). 87

Standardized mission sequence: avalanche trigger, operational briefing, checkpoint search and helicopter encounter. 88

Overview of the GeoBoardXR virtual scene.. 92

Georeference inheritance and ecosystem-level integration. 97

Geometry-aware disaster-risk reasoning in the SPACE+ UI. 98

Role-scoped semantics on a shared scene-graph node:. 99

Several activities available at ITI MIND. 102

IVR-SAD prototype scenarios and coping activities. 105

Immersive virtual operating room and anesthesia training environment. 110

Example of the treadmill-based VR platform in use, showing the HMD and motion trackers. 114

The clinician control panel for the Visual Cluttering exercise, with adjustable parameters and a live view of the patient's VR environment. 115

System overview. 118

a. view of the grab action b. view of the room 119

ACTA-VR Task 1 - Adaptive savory breakfast.. 123

Task 2 response-format redesign.. 124

Six capability layers for expressive virtual humans. 132
Client–server architecture linking avatar behaviour, speech and character configuration. 133
The VR avatar displaying different expressions and gestures during conversation 134
Experimental setup with two haptic devices. 135
System Architecture for Dual-Haptic Integration in an XPBD-Based Simulation Framework. 136
Soft tissue deformation and reaction force directions. 137
User interacting with a demo built on the proposed force-feedback application architecture 140
Multi-rate architecture and data flow during runtime (simplified application example) 141
Godot editor view of a simple scene with a manipulable cube and a floor 142
(Left) Scene illustrating rigid-body contacts. (Right) Scene illustrating viscous fluid interaction… 143
MetaCity Framework architecture.. 145
First-person VR views: cycling past the Royal Liver Building (left) and viewing the skyline from a vessel on the Mersey (right). 146
Ground-level, third-person exploration at street scale. 146
Overview of the available rooms 149
Stills from the 360° video. i) The Sun; ii) Mercury and iii) Mars. 154
The VR360 Creator Canvas 157
The assembled projector is used to view a slide relevant to Malaria. 158
Asturian cider-pouring traditional gesture.. 160
Overview of the Inmersidra system architecture. 161
Left: Inventor Martin Cooper experiencing Inmersidra. Right: External display of the VR session, showing the user’s avatar, the cider-pouring interaction, and the gamified interface (time and score feedback). 162
Inmersidra booth at events, engaging both older adults (left) and children (right), highlighting its inclusive design for diverse audiences. 163
A five-scale model for building XR capacity in Irish higher education. 165
System architecture overview. 170
Instructor view on a smartphone with highlighted UI elements. 171
System architecture, structured in three layers governing configuration, persistence and delivery 174
The four DAVE interaction modes: Intervention (dog's-eye perspective), Responsive (ladder of aggression), Tutorial (explanatory content) and Interactive (free play) 175

Figures in Posters and Demos

Overview of the scenario motivating this work: asymmetric MR collaboration around a virtual factory representation. 180

a) Average task completion times per technique. b) Subjective feeling of factory confidentiality (Q1 and Q1b) and perceived quality of the collaboration (Q2) with the different techniques. 181

The Standards Register's Overall Classification System 184

User Interface of the Current XR and Virtual Worlds Standards Register 186

AR workstation for washing-machine component recovery an inspection. 190

Concept of Environmental light modulation and its potential applications. 192

Outdoor experimental setup and hue-manipulation results with and without environmental-light. 193

Spatial frequency response under varying indirect environmental illumination levels. 194

Two views from the playable S01 beta on Meta Quest. 196

Playable journey and optional planned extension. 196

Photo of the test scene. From left to right: wood, cling film, aluminum foil.. 200

Distribution of cloud-to-mesh deviation per scan scenario, as percentage share of the 1,000,000 compared points. 203

Cumulative distribution of cloud-to-mesh deviation per scan scenario in cm. 203

Scan result (green) vs. ground-truth mesh (grey). From left to right: wood, cling film,. 203

Simplified pipeline steps. 203

Conceptual overview of the digital twin development of the ancient Greek hydraulis. 207

Workflow for the digitization and immersive exploration of organ pipes. 209

Integrated development workflow of the industrial Digital Twin, from multi-modal data acquisition. 213

Screenshot of the laptop screen, displaying the virtual forest and the photodiode flashing-square. 215

Local DLPFC modulation profiles during virtual navigation. 216

Representation of the virtual environment from participant’s perspective. 220

Visual representation of the tested body movements. 220

Mean participant ratings and standard deviations by condition for (a) the perceived motion fidelity questionnaire and (b) the VEQ Agency domain. 222

Composite behavioural battery × self-report 226

Conceptual pipeline from XR contextual information to structured vibrotactile cue generation, based on the proposed **‘semantic haptics’** approach and planned evaluation using the HaptiSense platform 229

HaptiSense XR demonstration setup with the wearable device, Meta Quest 3, and Unity visualization. 230

Examples of configurable spatial cue forms in the current XR implementation. 230

OpenXR-compatible PCVR–EEG acquisition pipeline. 237

EEG signal quality and event-related potential measures. 238

Overview of the 5G network coverage a) dataset and b) 2D and 3D interfaces to control antenna parameters. 246

Overview of 3D and 2D interfaces for a) wind parameters and b) urban heat areas. 246

Overview of the simulated fire storytelling experience. a) Fire starting. b) Intervention vehicles selection users. 246

The five-stage DemoXR pipeline, from multi-camera RGB-D and telemetry capture to a schema-validated instruction document rendered in a WebXR prototype. 241

(a) Capture during the robot replay. (b) Step 1, the recovered change as before (red) and after (green) point clusters with the measured move. (c) Step 2, the instruction in the WebXR prototype, illustrating document playback; physical registration to the table frame is not yet evaluated. 242

Therapist’s perspective seeing the bioresponsive avatars (mannequins) of two users. 246

Object detection with YOLO11n running on Meta Quest 3 252

User interacting with real bottle and virtual dog 252

Draft scheme of SENSOR with connections between simulator, fleet management, AI and UAVs (HITL - on the left side and SITL - on the right side). 256

Import of satellite images (left side) and grass and forestry generation in Unreal Engine through mask (right side) to create up-to-date realistic environments for training. 257

Experimental environment. Participants observed the embodied avatar in a virtual mirror while abducting the shoulder and judging the direction of the movement discrepancy. 259

Psychometric function for avatar movement gain.. 260

Rolling 2D occupancy grid with candidate projection areas. 265

The demonstrator during a laboratory test: the remote person appears as an avatar on the forward-facing display, with the 360° camera mounted directly above it. 268

Experimental procedure across the job-interview and cafeteria sessions. 273

Embodiment scores for personalized and female avatars by condition. 273

Characteristics of two- (left) and three-domain (right) split-domain light field visualization. 276

Two-domain split-domain light field visualization with content orientation facing the center view (left) and the domain-aligned view (right). 277

Sparse point cloud video visualized on the HoloVizio 640RC. 280

Sparse point cloud video visualized on the HoloVizio 80WLT 281

Overview of the AI data assistant (WARDY) in the collaborative XR digital twin environment. 283

Examples of WARDY's assistance and execution of tools in response to different prompts by XR users. 285

Architecture linking the VR heritage environment with HERITAGEGUARD through HTTP/JSON. 288

VR interfaces: (a) mode selection; (b) Visitor Mode; (c) Heritage Manager Mode. 289

Heritage Manager workflow: (a) screening-level assessment; (b) flood-risk visualisation; (c) 3D site model. 289

Representation of the four experimental environments along the Milgram Continuum. 292

Upper limb kinematic parameters during Reaching phase. 293

Upper limb kinematic parameters during Grasping and Transportation phase. 294

Screenshots taken from the VR training sessions (user perspective)), showing examples on interactive procedural tasks, instructional module, and assessment module. 296
The DMP with logs automatically sent from the VR headset at the end of training sessions. 298
Immersive inspection workflow in the Awen social housing scenario.. 301
LoomaXR workflow. 304
LoomaXR architecture: one authoritative hub between modular services and heterogeneous clients. 306
Two participants rendering the same scene, one in Unreal Engine, and the other using a web browser. 306
Generated asset using the prompt “a bottle of wine”. 307
A user searched for “props for an antique office”. 307
Distribution of stimuli in the valence-arousal space. 315
Virtual museum environments for normal vision and adapted for low vision. 320
Conceptual overview of the proposed multiscale methodology for the structural assessment and immersive visualization of the historic column. 323
Distributed System Architecture Diagram 325
Average creation time 326
Physical workspace configured in Mixed Reality on the user's real table 327
3D Object Manipulation 328
Overall Workflow of the Proposed XR NeedStruct Approach 330
Literature-Based Construction of Keyword Association Scores 331
XR holoportation technology by Volum Tech and Nova haptic gloves by SenseGlove 335
XR holographic call through Volum technology. 336
Haptic-augmented VR holographic calls through Volum and SenseGlove technologies. 336
Methodological workflow for the technical audit of VR applications. 339
Network requests captured via mitmproxy during a Playground VR session on the Meta Quest 1, showing repeated POST traffic to Unity Cloud Diagnostics. 340
HTTP request headers for a representative POST to cdp.cloud. 340
Excerpt of intercepted telemetry payload corresponding to the POST reproduced in Figure 3. Identifiers redacted; coordinate array elided between endpoints. 341
Experimental setup for Experiment B. 344
OptiTrack ground truth (blue) and the headset-reported trajectory (red) after hand-eye calibration and temporal alignment, projected onto the three coordinate planes. 345
Collaborative digital twin architecture across heterogeneous XR interfaces. 348
Demonstrator views across tablet-based AR, desktop VE, and immersive VR interfaces 349
Capture setup used for the demo, alongside examples of the immersive experiences and of the features provided by the Immersive FVV Live Platform. 352
Architecture diagram of the Immersive FVV Live Platform. 353
Control room with four miniature social environments and a wall screen. Radial safety menu. ... 355
The four full-scale social environments: social lounge, game room, music room, and park. 356

The four harassment scenarios: spatial intrusion and object attack 357
The virtual reality platform as viewed from the headset..... 361
BeRTA user interface..... 369
Overview of the sci-fi lab environment**.** 375
High-level system architecture diagram**.** 376
We show a user building objects with Forgenie’s input methods..... 378
Using indicator arrows, we show the process of spawning “an oak tree” into the environment, the user coloring the object, the user moving the object using the gizmo, the user scaling the object using hand interactions, and the final scaled object in the passthrough environment..... 380
Proposed VR interactions.. 385
Overview of the collaborative XR environment. 389
Images captured during initial testing, showing two users collaborating over two prerecorded surgical videos. 390
Overall architecture of the immersive teleoperation system 393
Diagram of the operator monitoring system..... 394
Operator station (up) and scale circuit used in the three scenarios (down)..... 395
Global mental workload (a) and SSSQ dimensions (b) per scenario (mean ± 95% CI)..... 396
SGAM, SCIAM and RAMI 4.0, reproduced from the cited sources..... 398
The proposed VWAM in the repository's model viewer. 399
Virtual training scenario with a patient on a bed and materials required for the procedure in the trolleys..... 404
A user performing key tasks of the pericardiocentesis procedure, including patient preparation, patient monitoring, and needle insertion. 404
Trust boundaries and data flows..... 409
The device-grant login and the reuse of its access token for xAPI writes..... 409
Sign-in happens in the Meta browser, outside the application..... 410

Lists of Tables

Tables in Applilication Track

Aggregate internal evaluation metrics charting foundational training groups (..... 15
Operational roles trained on the shared VR platform. 25
Characteristics of Participants 28
Summary of the Survey Results 56
Cross-platform prototype comparison. 72
Operating modes for the CIRA Mars VR Simulator. 76
Assessment dimensions and synchronized evidence. 89
Self-reported ratings during the IVR-SAD playtest. 107
Selected outcomes from the pilot evaluation 111
Retrospective lessons from ACTA-VR deployment and implications for VR cognitive assessment. 125
Abridged example of a complete XRCM data set [VDC, 2025]. 128
Four exemplary requirement-to-verification paths. 129
Descriptive statistics of XPBD performance metrics for V1 and V2 137
Mann–Whitney U test results and effect sizes for XPBD performance metrics. 137
Comparison of the three modalities. 154

Tables in Posters and Demos

Development status and intended conference demonstration. 196
Quantitative cloud-to-mesh distance (cm) results of the six main scans. 221
English translations of the self-reported questionnaire on perceived motion fidelity 221
Perception accuracy over N = 30 controlled single-object relocations, against hand-measured ground truth. 243
Student experience with XR technologies (% of national sample) 310
Chosen colours and their representation in the Munsell and hexadecimal systems 313
Relationships among challenges, requirements, and VR design principles for low vision. 319
Correspondence between XR NeedStruct outputs and expert assessments across the extracted user-function elements 333
Proposed scenario-to-interaction specification 383

Series title and number

VTT Technology 449

Title	**EuroXR 2026** Proceedings of Application, Poster, and Demo Tracks of the 23rd EuroXR International Conference
Author(s)	Kaj Helin, Sara Arlati and Armin Grasnick (eds.)
Abstract	We are pleased to present these conference proceedings in the VTT Technology series, which contains the papers accepted for the Application and Poster & Demo Track of EuroXR 2026, the 23rd annual EuroXR conference, being hosted from November 3rd to November 6th 2026, at Polytechnic City of Innovation (CPI), Universitat Politècnica de València. This publication is thus a collection of the application papers (talks, posters and demonstrations) presented at the conference. It provides an interesting perspective into current and future applications of VR/AR/MR. EuroXR Conference 2026 is held in conjunction with the Virtual Worlds Partnership (VWP) Annual Forum 2026, creating a convergence of scientific excellence, industrial innovation, and policy leadership within Europe's immersive technologies ecosystem.In previous years, under the name EuroVR, the Conference has been held in Bremen (2014), Lecco (2015), Athens (2016), Laval (2017), London (2018), Tallinn (2019) and Valencia (2020); and as EuroXR in Milan (2021), Stuttgart (2022), Rotterdam (2023), Athens (2024), and Winterthur (2025). The focus of the EuroXR conferences is to present, each year, novel eXtended Reality (XR) technologies, including software systems, display technologies, interaction devices, and applications, to foster engagement between industry, academia, and the public sector, and to promote the development and deployment of VR/MR/AR technologies in new, emerging, and existing fields. This annual event of the EuroXR association (https://www.euroxr.org/) provides a unique platform for exchange between researchers, technology providers, and end users around commercial or research applications. We would like to warmly thank the industrial committee chairs for their great support and commitment to the conference, and special thanks go to the local organizing committee for their great effort in making this event happen.
ISBN, ISSN, URN	ISBN 978-951-38-8814-5 ISSN-L 2242-1211 ISSN 2242-122X (Online) DOI: 10.32040/2242-122X.2026.T449
Date	November 2026
Language	English
Pages	429 p.
Name of the project	
Commissioned by	
Keywords	XR, virtual words, EuroXR association, applications, posters, demos
Publisher	VTT Technical Research Centre of Finland Ltd P.O. Box 1000, FI-02044 VTT, Finland, Tel. 020 722 111, https://www.vttresearch.com

EuroXR 2026

Proceedings of Application, Poster, and Demo Tracks of the 23rd EuroXR International Conference

The 23rd EuroXR International Conference – EuroXR 2026 will be held on 3 to 6 November 2026 at the Polytechnic City of Innovation (CPI), Universitat Politècnica de València, Spain.

This conference continues a series of successful European VR/AR conferences that have been held since 2004, previously known as INTUITION, JVRC, and more recently EuroVR. EuroXR 2026 will bring together researchers, technology providers, industry representatives, end users, and public-sector stakeholders interested in eXtended Reality (XR). It aims to facilitate the exchange of knowledge and experiences, showcase new results and applications, and promote the development and deployment of XR technologies in new, emerging, and existing fields.

EuroXR 2026 is held in conjunction with the Virtual Worlds Partnership (VWP) Annual Forum 2026, bringing together scientific excellence, industrial innovation, and policy leadership within Europe's immersive technologies ecosystem. The conference provides a platform for exchange between researchers, technology providers, end users, industry, and public-sector stakeholders and supports collaboration on future XR and Virtual Worlds research, innovation, and applications.

ISBN 978-951-38-8814-5
ISSN-L 2242-1211
ISSN 2242-122X (Online)
DOI: 10.32040/2242-122X.2026.T449

VTT

beyond the obvious