

EuroXR 2025

Proceedings of the Application, Poster, and Demo Tracks of the 22nd EuroXR International Conference

Kaj Helin (ed.)

VTT Technical Research Centre of Finland Ltd, Finland

Barbara Schiavi (ed.)

XR Research & Development at Immersion, France

Electra Tsaknaki (ed.)

Game Developers Association of Greece

Päivi Vahala-Partanen

Technical Editing & Graphical Design

VTT Technical Research Centre of Finland Ltd, Finland

ISBN 978-951-38-8804-6

VTT Technology 440

ISSN-L 2242-1211

ISSN 2242-122X (Online)

DOI: 10.32040/2242-122X.2025.T440

Copyright © VTT 2025

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>

EuroXR 2025: Winterthur, Switzerland

The EuroXR 2025 conference will focus on cutting-edge advancements in Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) 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 5 September 2025

Winterthur, Switzerland

Conference organizers

Zurich University
of Applied Sciences



EUROXR

VTT

Preface

We are pleased to present these conference proceedings in the VTT Technology series, which contains the papers accepted for the Application Track and Poster & Demo Track of EuroXR 2025, the 22nd annual EuroXR conference, being hosted by ZHAW (School of Management and Law) from September 3rd to September 5th 2025, in Winterthur, Switzerland. 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.

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) and Athens (2024). The focus of the EuroXR conferences is to present, each year, novel Virtual Reality (VR), Mixed Reality (MR) and Augmented Reality (AR) 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 association¹ 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 😊.

¹ <https://www.euroxr.org/>

On behalf of the organising committee,



Kaj Helin



Kaj Helin

*EuroXR Application track,
Principal Scientist at VTT Technical Research
Centre of Finland Ltd, Finland*



Barbara Schiavi



Barbara Schiavi

*EuroXR Application Track, XR Research &
Development at Immersion, France*



Electra Tsaknaki



Electra Tsaknaki

*Euro XR Poster & Demo Track,
XR Researcher, Game Developers Association of
Greece*

Application Program Co-chairs

- ▶ **Barbara Schiavi**, Immersion, France
- ▶ **Jérôme Perret**, Haption, Germany
- ▶ **Kaj Helin**, VTT, Finland
- ▶ **Frédéric Noël**, INP Grenoble, France
- ▶ **Thomas Hulin**, German Aerospace Center (DLR), Germany
- ▶ **Bhuvanewari Sarupuri**, TOPdesk, Hungary
- ▶ **Alina Kadlubsky**, Hochschule Mittweida, Germany
- ▶ **Fabio Mosca**, Another Reality, Italy

Posters and Demos Co-chairs

- ▶ **Electra Tsaknaki**, University of Cyprus, Cyprus
- ▶ **Mario Lorenz**, TU Chemnitz, Germany
- ▶ **Armin Grasnick**, Int'l University, Germany
- ▶ **Kayvan Mirza**, Optinvent, France
- ▶ **Vladimir Kuts**, Taltech, Estonia

General Co-chairs

- ▶ **Wolfgang Schaefer**, ZHAW, Winterthur, Switzerland
- ▶ **Wolfgang Stuerzlinger**, Simon Fraser University, Canada
- ▶ **Luciana Nedel**, Federal University of Rio Grande do Sul, Porto Alegre, Brazil

Scientific Program Co-chairs

- ▶ **Despina Michael-Grigoriou**, Cyprus University of Technology, Cyprus
- ▶ **Gabriel Zachmann**, University of Bremen, Germany
- ▶ **Regis Kopper**, Iowa State University, USA
- ▶ **Sang Ho Yoon**, Korea Advanced Institute of Science & Technology (KAIST), South Korea
- ▶ **Stefanie Zollmann**, Aarhus University, Denmark
- ▶ **Patrick Bourdot**, CNRS & University Paris-Saclay, France

Organizing Team

- ▶ **EuroXR** –Patrick Bourdot, Mónica González, Jerome Perret, and Gabriel Zachmann
- ▶ **ZHAW** (Zurich University of Applied Sciences) – Wolfgang Schäfer.

Table of contents

EuroXR 2025: Winterthur, Switzerland	1
Preface	3
APPLICATION TRACK	
Empowering Civic Education Through Metaverse-Enriched Teaching Scenarios.....	13
Immersive Storytelling for Peacebuilding: CoXist Plus' XR Approach to Empathy and Intercultural Dialogue.....	17
Making the Invisible Visible - AR supported situation awareness in the snow groomer operations	22
GAIA project: Enhancing awareness of gender harassment in higher education settings through immersive training.....	27
Haptic surgical training based on recorded data	31
A Virtual Reality Stroop Art Exergame for Cognitive-Motor Rehabilitation in Chronic Stroke Survivors	37
ExplAining the Future of Work: WIRKsam XR Demonstrators.....	41
SurgiScope: An Immersive and Interactive Platform for Surgical Training.....	45
Interactive 3D Visualization and Configuration of Solar Panel Layouts in XR.....	50
Immersive Exploration and Data Acquisition: XR Virtual Field Trips for 3D Photogrammetry and Remote Sensing (SURE-VFT).....	54
XR Helicopter Simulator for LLM-based Virtual Assistant in Maintenance Support	58
An Extended Reality Interface for Ergonomics Training	63
SRswift: A Spatial Reasoning Library in Swift.....	68
V CARE: A Virtual Reality (VR) Artificial Intelligence (AI) Assisted Training Tool for Alzheimer's Caregivers	72
INFINIBRICK: A Tool for the Generation of Endless LEGO Assembly Tasks for Studies on Prolonged Immersive Virtual Reality Exposure	76
Fostering Inclusive Work Environments Through Immersive Training Solutions for Visual Impairment Awareness	80
Globe Viewer for Location-Based Augmented Reality.....	84
SITE-VR: A Framework for Improving User Experiences in Immersive Training	89
Validating an Immersive Virtual Environment for Learning 3D Transformations: From Design to Student Impact to Expert Evaluation.....	94
Digital Characters Augmenting Cultural Spaces through Handheld and Head-Worn Devices	98

POSTERS

HOLOVER-5G: Industrial Trainings Integrating Spatial Gaussian Splats with Real-Time 3D Holograms over 5G.....	104
BAZKARIA: Eating Together Through Immersive Telepresence.....	109
Virtual Reality Evacuations for Resilient Mobility Systems: Users' Perspectives.....	113
SELFEX ² : Real-time Telepresence Dexterity Training for Manufacturing Using Wearable Technologies and Augmented Reality.....	117
ROTIE (Robot Teleoperation Through Immersive Environments).....	121
Robust AR Anchoring Using Ultra-Wideband (UWB) Beacons for Spatial Consistency in XR Environments.....	125
VISOR: Virtualization Service for Object Reconstruction.....	128
SAME-XR: A VCS-Integrated Asset Management Pipeline to Bridge the Gap Between 3D Artists and XR Developers.....	132
MANTRA: Mixed-reality Augmented Networks for Teleoperated Robotics Applications.....	135
FLYTEX: Merging IoT data into Videoconferences for a Better Decision-Making Process in Agriculture.....	139
Quality of Experience Evaluations for Multi-Cloud Streaming.....	144
HOPE: Holographic Optimized Processing Engine.....	149
On the Skinning of Gaussian Avatars.....	152
Promoting Pro-environmental Behaviour through Immersive Animal Perspective-Taking.....	157
XRRehab: Extended Reality for Neurological Rehabilitation.....	160
Effects of Musical Pleasantness on Intrinsic Motivation During Exercise in Non-Immersive Virtual Reality: A Pilot Study.....	165
VR Journeys for Older Adults: Storytelling-Enhanced Virtual City Tours and Their Impact on Senior Customer Experience.....	170
Dynamic, Room-Adapted Spatial Audio for Augmented Reality Telepresence.....	174
Illustrating AR Patterns: A Visual Language for Event-Driven Augmented Reality.....	179
Telepresence-Enhanced Network Music Performance.....	184
AgriVision: Extended Reality for Efficient and Sustainable Farming.....	189
Immersive Fundraising: Motivational Mechanisms Converting VR Call-to-Action into Actual Donations.....	193
Fitting the Future: How AI-Driven Virtual Try-Ons Re-shape Brand Positioning Across Fashion Segments.....	197
Auditory Presentation of Stimuli for Cognitive Load Dual tasks in Redirected Walking.....	201
CoVA: LLM Data Augmentation to Enhance User-Defined Intents Recognition for Voice-Based Assistance in Collaborative XR Environments.....	206
RHYTHM (Real-time HYbrid Test for Harmonic Music).....	211

Predicting Success Rate of Target Hitting in VR Ball Throwing for Physical and Psychological Support	213
Demonstrating Immersive Teleoperation of Robots via Adaptive Point Cloud Streaming	217
VR CHARLIE: Chat Assistant for Digital OSCE Examinations	221
Individual Influences in Highly Immersive VR: Two Exploratory Studies on Personality, Perception, and Presence	225
CORE-MHC: CO-facilitated and REmote Mixed-Reality Mental Health Care interventions	229
3D Selection through Hand Tracking in XR for Point Clouds Using Signed Distance Fields.....	234
Toward Efficient Storage and Real-time Rendering of Dense 3D Gaussian Splatting Reconstruction in Virtual Reality.....	238
AI-Powered Scene Generation in VR	243
Genba SaVR: A Secure and Scalable Virtual Reality Platform for Industrial Training.....	246
Advancing Collaboration Paradigms in Extended Reality.....	251
 DEMOS	
Z-3PO: Seamless Switching Between 3 Representation Perspectives for Real-Time Immersive Teleconferencing	257
XR Meets AI and Biofeedback: Technology-Enhanced Stress Management Training for Healthcare Professionals	262
Control Your XR: A Dashboard for XR Applications.....	266
Transmitting Volumetric Video via Media over QUIC to Subscribers with Contradicting Spatial and Temporal Preferences	270
Drivable Group Tele-presence Demo	274
Beyond Classical Cooperation: Reimagining Business Meetings through VR Conferencing	278
Demonstrating V-PCC Bitstream Transcoding	283
EMPATHI: Emotion-Responsive Avatars for Personalized Cultural Heritage Interaction	287
Real-Time Avatar-Guided VR Tours: Integrating Live Telepresence in Cultural Heritage Applications ..	291
SURE-VFT Remote Sensing: An XR based educational application for enhancing skills in Land Use Land Cover (LULC) mapping through satellite imagery	295
Virtual Reality-based exercises for the rehabilitation of vestibular disorders.....	299
Immersive Exploration of Fluid Dynamics: Developing an XR/VR Platform for CFD Virtual Testing in Education and Social Inclusion	303
Author Index	309
Tables of Figures.....	313
Lists of Tables.....	318

Application track

Empowering Civic Education Through Metaverse-Enriched Teaching Scenarios

Konstantinos Tsaramirsis¹, Ilona Lasica², George Lampropoulos², Avgoustos Tsinakos³, Ioannis Kazanidis³, George Terzopoulos², Konstantinos Bafes¹, Shana Hazboun⁴, and Eleni Mangina¹

¹University College Dublin, Ireland

²University of Piraeus, Greece

³Democritus University of Thrace, Greece

⁴KU Leuven, The Netherlands

Corresponding author: eleni.mangina@ucd.ie

Keywords: Metaverse in Education, Civic Learning, XR Pedagogy

Abstract

This paper presents the MetaCivicEdu project, which integrates XR and Metaverse technologies into primary school civic education across Ireland, Greece, and Cyprus. By co-creating immersive learning scenarios with educators, the project promotes student engagement, civic knowledge, and digital skills. The Zenodo-based project repository hosts over 100 open-access educational scenarios. Teacher training workshops have improved confidence and pedagogical innovation. The System Usability Scale (SUS) evaluation provided excellent results, affirming the repository's effectiveness. MetaCivicEdu advances civic education through technology-driven and learner-centred approaches.

Introduction

On average, one in five people report not being at all interested in politics across OECD countries (OECD, 2018). The role of the European Commission (EC) is to become a leader on promoting and raising awareness of the importance of Participatory Citizenship. According to the IEA international study to include measurement of civic knowledge, which had an international mean of 500, Ireland ranked seventh out of 36 countries with a mean score (534). A total of 14 countries including Cyprus and Greece have a mean score significantly below the international average. Offering innovative learning strategies for civic education can help prepare students for active citizenship in the 21st century by promoting civic skills, knowledge, and values. It is essential to ensure that civic education remains relevant and engaging for students to promote lifelong engagement in civic life. Modern citizenship education should create innovative pedagogical models, engaging and interactive learning environments (European Commission, 2017) and immersive technologies such as Virtual Reality can support this goal by enabling experiential, student-centred learning that strengthens civic understanding and participation (Qin et al., 2025). In this project we show how civic education can help promote active citizenship and encourage young students to participate in achieving these goals. Civic education and the United Nations Sustainable Development Goals (SDGs) are closely related. The SDGs

aim to promote sustainable development and address global challenges, such as poverty, inequality, and climate change. Technology can be used to enhance civic education by providing interactive and immersive experiences that promote civic skills and knowledge.

The MetaCivicEdu project addresses the challenge of disengagement in civic education by introducing metaverse-based learning tools. These tools offer immersive, interactive environments that make civic concepts more relatable and experiential. The motivation lies in enhancing participatory citizenship and digital skills through innovative pedagogy. By linking civic education to real-world digital contexts, the metaverse supports relevance, inclusivity, and lifelong civic engagement. The project supports primary school educators in promoting active citizenship using Virtual Reality (VR) and Metaverse-based learning environments. This paper highlights the creation of an open-source scenario repository and early project activities included the collection of XR-enriched learning scenarios (ZENODO MetaCivicEdu Community, 2024) along with teacher training seminars across Cyprus, Greece, and Ireland. XR technologies are central to MetaCivicEdu project's goal of fostering active citizenship in primary education. They enable experiential learning through immersive scenarios that promote critical thinking, collaboration, and civic awareness. XR also supports teacher training and scenario co-creation, increasing educators' confidence in digital pedagogy. Ultimately, XR bridges abstract civic topics with tangible, engaging experiences for young learners.

Metaverse Learning Scenarios

MetaCivicEdu consortium identified, designed, implemented, and disseminated metaverse - enriched civic education scenarios. Initially 12 traditional civic education scenarios (4 from each country: IE, GR, CY) were identified and addressed themes such as SDGs, democratic participation, and cultural awareness. Multiple workshops were held (LIS2024 F2F, UNIPI, UCD, CYENS), engaging 142 teachers, with 93 evaluation survey responses. The professional development program, structured around Bloom's Taxonomy, the SAMR model, and Triple E Framework, successfully trained teachers on creating metaverse-enriched civic scenarios. Participants expressed increased confidence in the use of metaverse and lesson design innovation, with evaluation forms confirming high satisfaction. The Zenodo-powered repository was developed and the MetaCivicEdu Moodle platform was set up for the teachers' training. MetaCivicEdu Zenodo Community hosts 67 learning scenarios, indexed with educational metadata and released under Creative Commons licenses. The System Usability Scale (SUS) score of 82 has indicated excellent user experience. While the digital repository is still in progress, more than 100 metaverse enriched scenarios have been uploaded. The Quality Assurance and Evaluation of the learning scenarios included:

Scenario Validation: A robust QA framework was applied using a structured rubric to evaluate pedagogical wholeness, effectiveness, and creativity. The results of 12 rubric-based evaluations (via Google Forms) demonstrated consistent quality, with strong feedback on alignment with goals, learner engagement, and innovative use of metaverse enriched scenarios.

Teacher Training Impact: Post-workshop surveys (UCD, CYENS, UNIPI) showed that teachers gained valuable practical experience and confidence in applying XR technologies. Feedback specifically highlighted enhanced creativity, relevance of materials, and practical scenario application.

Repository Usability: The SUS assessment validates the repository as a high-performing digital tool, with SUS score average 82, Grade A in ease of use, learning curve and system stability. Currently it is reported 521 views and 317 downloads of the repository contents.

Title	Distribution of the World's Resources	
Targeted audience	Grade 6 (11-12 years old)	
Short Description	This interdisciplinary lesson plan introduces students to the concept of resource distribution, focusing on global inequalities. Through engaging activities such as a musical mats game and a resource-sharing board game, students experience the challenges and decisions involved in equitable resource allocation. Reflection and discussion activities guide students to explore solutions and foster empathy.	
Duration & sequence	Teaching time WITHIN Metaverse: 90 minutes (adjustable) The virtual educational activities that can be implemented within Metaverse are (can be implemented in any sequence or in row, while the duration is adjustable):	
	Musical Mats Game	Activity 1 30 minutes
	Resource Sharing Board Game	Activity 2 45 minutes
	Reflection and Discussion	Activity 3 15 minutes
	Free spaces/ collaboration and brainstorming activities	Out of sequence Free time



Figure 1. Example: "Distribution of resources [ZENODO Learning Scenario](#) & [Spatio.io](#) environment activities.

Figure 1 presents a Metaverse-based civic learning scenario on "Distribution of Resources" set in a virtual schoolyard. Students engage with interactive mats and then participate in a collaborative board game simulating resource sharing. They take turns as game pawns, using a virtual dice to advance. The activity concludes with class discussion to reflect on the experience.

Metaverse Learning Scenarios Repository Evaluation

The open-access digital repository was designed to serve as a central hub for storing, sharing, and retrieving metaverse-enriched learning scenarios, catering to the needs of educators and trainers in civic education. By integrating advanced technologies and leveraging the robust infrastructure of Zenodo, the repository aligns with the project's commitment to fostering innovation and accessibility. Its development was guided by the principles of usability, inclusivity, and sustainability, ensuring that it meets the diverse needs of its users while supporting the broader goals of the MetaCivicEdu project. The SUS evaluation was based on the rating of 10 usability statements on a 5-point Likert scale. Positive and negative items were alternately scored, and the aggregate scores were standardized to a 0–100 scale. The results of each respondent's SUS scores are calculated as per the standardized SUS calculation:

$$SUS = \Sigma ((Q_o - 1) + (5 - Q_e)) \times 2.5$$

- For each of the odd numbered questions (Q_o), subtract 1 from the user response
- For each of the even numbered questions (Q_e), subtract the use response from 5
- This scales all values from 0 to 4 (with four being the most positive response)
- Add up the converted responses for each user and multiply that total by 2.5 - this converts the range of possible values from 0 to 100 instead of from 0 to 40

The open-access digital repository demonstrates high usability and effectively supports its intended civic education objectives. The SUS assessment underscores its strengths while highlighting opportunities for improvement. Future iterations will benefit from enhanced user guidance, simplified interfaces, and stronger integration with complementary tools. The SUS evaluation took place November 4–18, 2024 with 13 participants and provided SUS Score 82, Grade A. The main strengths highlighted by users included: Simple and intuitive navigation; Users quickly became comfortable with the repository; Reliable and consistent performance.

Discussion

Although the project is still in its early-to-mid implementation phase, MetaCivicEdu has already generated tangible impact on the participants, partner organisations, and target groups across the three participating countries, with signs of growing influence at the European level. Primary school teachers who participated in the project's training and scenario co-creation activities have significantly enhanced their digital pedagogy and confidence in using VR/Metaverse tools in the classroom. Many of them reported that this was their first experience integrating immersive technologies into civic education. Students involved in the pilot studies have demonstrated increased engagement and interest in civic topics. Initial qualitative feedback indicates that the immersive environments have made abstract civic concepts more tangible and personally relevant. Participating schools have strengthened their capacity for innovation, particularly through the hands-on exposure to scenario-based learning and immersive technology. These institutions are now seen as early adopters of VR/XR education approaches. Partner organisations have deepened their collaboration, expanded their expertise in immersive learning design, and enhanced their visibility at the national and European levels. Several have begun exploring new partnerships or extensions based on the project outcomes. The educational content and open-access learning scenarios have already been accessed and used by educators outside the immediate partner institutions, thanks to the Zenodo repository and dissemination efforts. The project has contributed to raising awareness among school leaders on the importance of embedding civic education into digital transformation strategies in primary education.

References

- European Commission. 2017. Citizenship education at school in Europe, 2017. Publications Office of the European Union. <https://data.europa.eu/doi/10.2797/536166>
- Qin, H., Zhang, Y., & Wei, J. 2025. Virtual Reality (VR) Technology in Civics Practice Teaching Evaluating the Effect of Immersive Experience. [International Journal of Advanced Computer Science & Applications](https://doi.org/10.1007/978-98-99-10-000-0_16), 16(3).504-516
- OECD 2018. Citizens with a saying. [Trends Shaping Education 2016, OECD Publishing, https://www.oecd.org/content/dam/oecd/en/publications/reports/2017/12/citizens-with-a-say_8f7c306b/3092726b-en.pdf](https://www.oecd.org/content/dam/oecd/en/publications/reports/2017/12/citizens-with-a-say_8f7c306b/3092726b-en.pdf)
- MetaCivicEdu Erasmus+ project, Project Reference 2023-1-EL01-KA220-SCH-000159834, <https://metacivic.eu/>
- ZENODO MetaCivicEdu Community 2024. MetaCivicEdu - XR Education for Civic Impact. <https://zenodo.org/communities/metaversecivicedu/records?q=&l=list&p=1&s=10&sort=newest>

Immersive Storytelling for Peacebuilding: CoXist Plus' XR Approach to Empathy and Intercultural Dialogue

Mahder Mulusew Getaneh

U.S. Embassy Addis Ababa, Co(X)ist, Vista Technologies, Ethiopia

Corresponding author: mahlermulusew@gmail.com or madiloveskore@gmail.com

Keywords: XR for social impact, immersive storytelling, peacebuilding, behavioral change, empathy, intercultural communication, youth engagement

Introduction: Building Empathy Through Immersion

Global tensions, cultural misunderstandings, and polarized identities continue to threaten social cohesion, especially among youth in marginalized communities. People often fear what they do not know, leading to stereotypes and the belief that their own cultural values and norms are superior. This egocentric perspective fosters division and exclusion.

CoXist Plus leverages immersive storytelling through XR (Extended Reality) to place users in someone else's shoes, enabling them to see the world from a different cultural perspective and feel genuine empathy. Our XR scenarios recreate the lived experiences of different ethnic groups, their cultures, values, and norms, helping users understand and appreciate diversity. By embodying these experiences in VR, users move beyond fear and stereotypes toward acceptance and inclusion, fostering intercultural dialogue and peacebuilding.

Application Description: The CoXist Plus Model

CoXist Plus creates immersive VR experiences that simulate real-life cultural practices and daily scenarios of diverse African ethnic groups. Users actively engage by:

- Wearing and choosing traditional cultural clothing, and you can check your outfit from the mirror reflection of your avatar wearing it, that is hanging in the virtual room
- Experiencing ethnic greetings in local languages
- Learning about traditional foods and their preparation
- Participating in ethnic dance styles, physically mimicking movements
- Understanding traditional means of transport and community rituals



Figure 2. CoXist Plus VR game Level 1 and 3 scene.

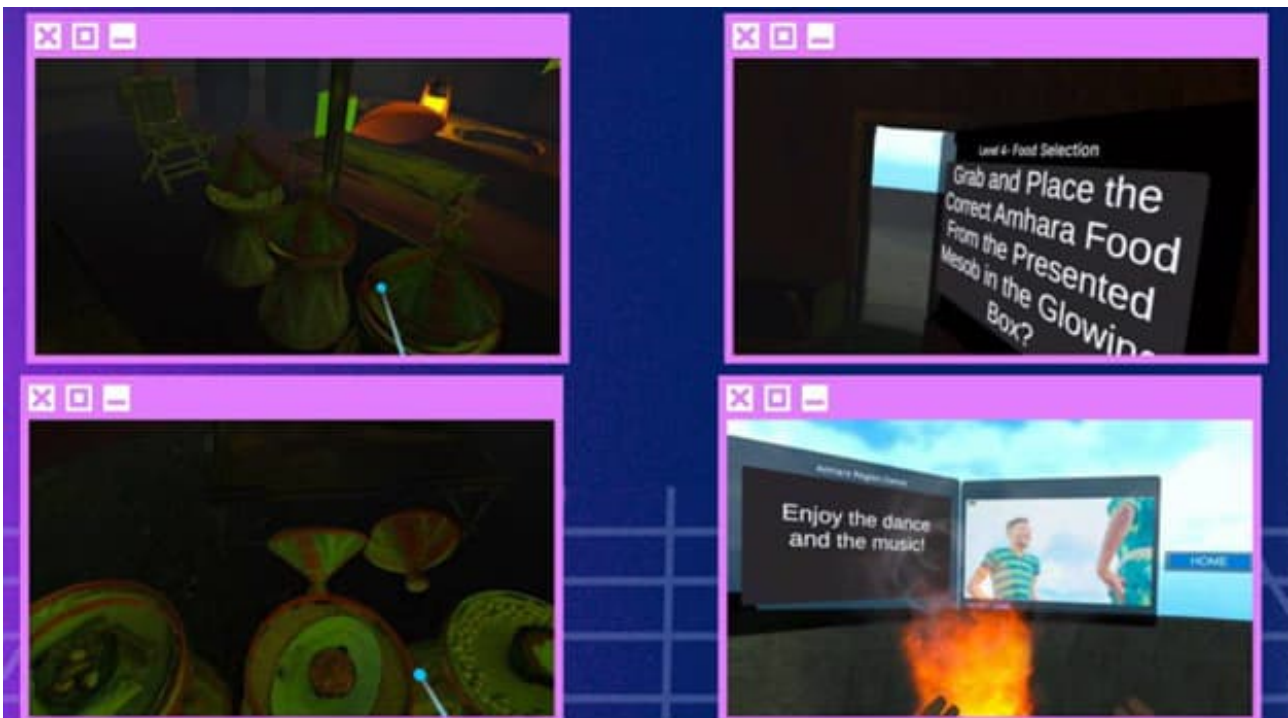


Figure 3. CoXist Plus VR game Level 4 and 5 scene

These cultural immersion activities are interwoven with conflict and reconciliation narratives, often based on authentic stories collected from African communities. Through branching narratives, users' choices impact outcomes, enhancing their understanding of prejudice, injustice, and reconciliation.

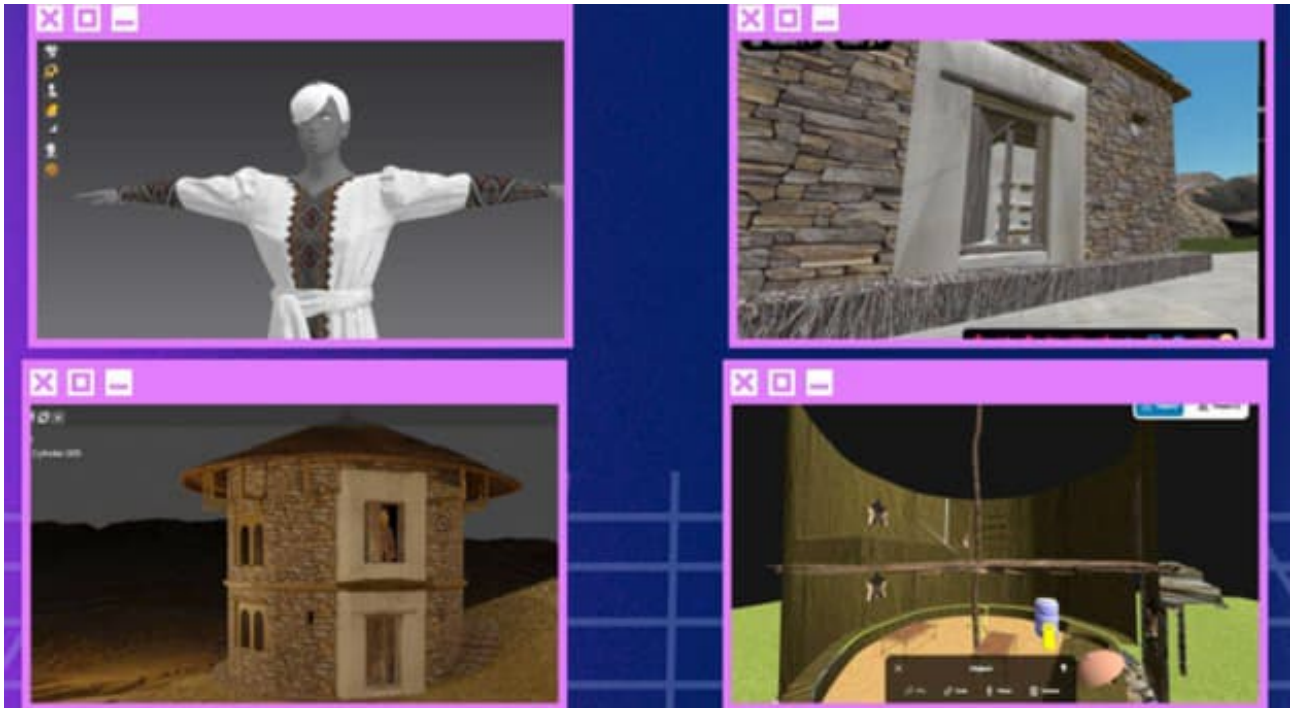


Figure 4. CoXist Plus VR game Vernacular architecture and designed modeled assets

Our platform is developed in Unity and deployed on Meta Quest 2 and 3 VR headsets and through labs located in schools, youth centers, and community spaces to maximize accessibility and impact.



Figure 5. CoXist Plus VR game Vernacular architecture interior of rural houses

Impact and Evaluation

We piloted CoXist Plus in Ethiopia with over 1,000 students in collaboration with local schools and NGOs. Key findings include:

1. 78% of participants reported heightened empathy and reduced ethnic bias based on pre/post-experience surveys
2. Teachers and community leaders observed an increase in open civic dialogue among youth
3. Participants expressed greater willingness to engage respectfully with peers from different ethnic backgrounds

Table 1. CoXist Plus - Impact and Evaluation testing data

Institutions and events	Females	Males
Addis Ababa Science and technology university	80	120
St. Mary University	45	62
Ice Addis Hackathon	37	76
Tariku and Desta Education Through Tennis	198	242
Creative hub project close up	41	59
Addis Ababa Science Museum, Aviation expo	67	133

These outcomes demonstrate immersive storytelling as an effective, scalable tool for social transformation, peace education, and behavioral change.



Figure 6. CoXist Plus VR game testing, hackathon and workshop

Technical and Design Implementation

Our storytelling platform leverages:

- Unity3D for interactive narrative design
- Meta Quest 2 and 3 hardware for accessibility
- Local language dubbing and subtitles to ensure inclusivity
- Future integration of Emotion AI and interaction tracking to capture real-time behavioral responses

Ethical storytelling is central: content is co-created with local artists, cultural experts, and conflict resolution specialists to ensure authenticity and respect.

Relevance to EuroXR

This work aligns with EuroXR's objectives by:

- Demonstrating a tested, functional XR application with strong social impact
- Tackling urgent real-world challenges through immersive technology
- Emphasizing inclusive design, ethical considerations, and intercultural dialogue
- Offering a replicable model for education, civic engagement, and peace advocacy

We propose presenting this as an oral presentation or demo, accompanied by a short VR experience booth for attendees.



Figure 7. CoXist Plus VR game demo at science museum for government officials

References

- Serrano, B., Baños, R.M., & Botella, C. 2016. Virtual reality and stimulation of touch and smell for inducing empathy: An experimental study. *IEEE Transactions on Visualization and Computer Graphics*, 23(4), 1105–1112.
- Slater, M., & Wilbur, S. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 6(6), 603–616.

Making the Invisible Visible - AR supported Situation Awareness in the Snow Groomer Operations

Kaj Helin, Andrea Alesani, and Timo Kuula

VTT Technical Research Centre of Finland Ltd

Corresponding author: kaj.helin@vtt.fi

Keywords: Making the Invisible Visible, Augmented Reality, Situation Awareness

Introduction

This work was carried out as part of the EU-funded project Theia-XR (Making the Invisible Visible for Off-Highway Machinery by Conveying Extended Reality Technologies). This abstract introduces an augmented reality (AR)-supported situational awareness application designed to enable a seamless transition between the real-world view and an AR-enhanced interface. Users can intuitively select the level of information transparency, tailoring the experience to their needs. A test use case involving a snow groomer demonstrated how operators can preview and execute grooming tasks using real-time terrain overlays provided by AR, leading to enhanced efficiency, safety, and situational awareness.

Human centric approach

The development process has been iterative and cyclical, and it followed as much as possible the best practices of human-centred design as described in ISO standard see (Helin et al, 2024). The basic design cycle consists of the following phases: Understanding the context of use, specifying user requirements, producing design solutions, and evaluating the design (ISO 9241-210).

The First Prototype

The first implementation of the AR supported situation awareness application was conducted using a snow groomer as the primary operational use case, selected for its demanding terrain interaction and the precision required in slope management. This implementation integrates mobile AR interfaces—specifically on the iPad Pro M4 (2024)—with immersive mixed reality experiences via high-end headsets like the Apple Vision Pro. The system is designed to enable operators to fluidly transition between real-world machine operation and virtual environments, creating a continuous, context-aware workflow that enhances both situational awareness and operational efficiency.

The iPad-based version of the AR supported situation awareness application was tested in a real-world deployment using an actual snow groomer. This implementation utilized the iPad Pro M4 (2024),

leveraging its powerful M4 chip, high-resolution display, and onboard sensors—including GPS, IMU, and camera—to deliver a responsive and spatially accurate AR experience. The AR interface featured a 3D grid overlay aligned with the terrain, helping operators visualize slope geometry and grooming coverage in real time (see Figure 8). A key component of the interface was the interactive 3D minimap, which allowed users to freely rotate, zoom, and pan around the grooming area from any perspective. This minimap was dynamically updated using real-time GPS data from the snow groomer, ensuring that the virtual representation of the environment remained synchronized with the machine's actual position and movement. To enhance user control and adaptability, the interface included a slider in the tablet UI that allowed the operator to adjust the density and visual prominence of the grid overlay, effectively controlling the level of immersion and visual guidance provided during operation. This feature enabled operators to tailor the AR experience to their preferences and environmental conditions, improving both usability and situational awareness.

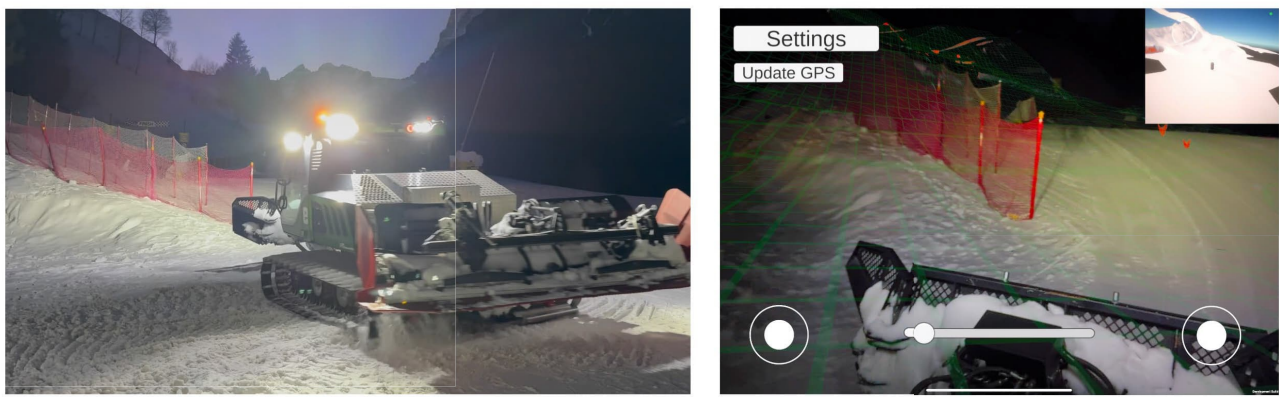


Figure 8 Tablet (iPad Pro M4 (2024)) interface with snow groomer.

The Apple Vision Pro version of the MR supported situation awareness application was tested on a ski slope in a dynamic, real-world environment while skiing. This implementation fully utilized the Vision Pro's advanced spatial computing capabilities, including its dual-chip architecture (M2 and R1), ultra-high-resolution micro-OLED displays, and precise environmental tracking. The immersive interface featured a fully navigable 3D minimap, which served as a real-time spatial representation of the slope (see Figure 9). Users could freely move within this virtual space, rotate their viewpoint, and reposition themselves to explore terrain features and skier movement from any angle. The minimap was constructed using spatial mapping and environmental sensing, allowing for a highly realistic and interactive representation of the slope. Additionally, the interface included a slider-based control that allowed users to adjust the density and visual intensity of the overlaid grid, tailoring the level of visual guidance to their preferences or environmental conditions. This level of immersion, combined with customizable visual aids, enabled users to analyze terrain, anticipate movement patterns, and make informed decisions in a way that was both intuitive and highly engaging. The Vision Pro's capabilities demonstrated how mixed reality can enhance situational awareness and planning in outdoor, dynamic environments such as ski slopes.

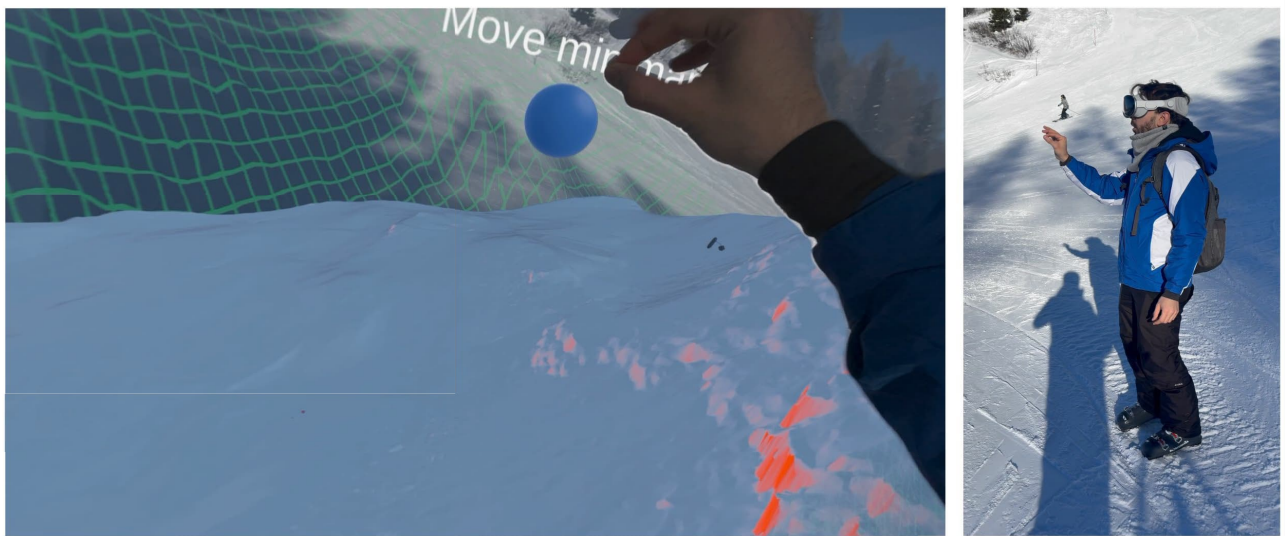


Figure 9. Head mounted (Apple Vision Pro) interface in slope.

User Feedback

The first user evaluation of the AR supported situation awareness application was conducted on January 14, 2025, at Prinoth's premises in Sterzing. The VTT development team demonstrated the tablet-based system to three expert users from Prinoth—two test engineers and one demo operator, all male – who then tested the system hands-on. Feedback was collected during and after the session. Overall, the concept was positively received, particularly the potential to visualize snow-related data such as depth and distribution, which users found more valuable than the terrain model shown in the test. Users appreciated the potential benefits in poor visibility conditions and highlighted the importance of adjustable visual elements, such as the grid visibility slider. Additional suggestions included color-coded snow depth visualization, zoom functionality in the minimap, and expanding the system to support multi-user collaboration and real-time awareness of other machines and personnel on the slope – features that could significantly enhance safety and coordination in large ski resorts.

The Final Prototype

Based on the user evaluation, several meaningful updates were implemented in the iPad version. One of the most significant enhancements was the introduction of a color-coded snow depth visualization (see Figure 10), which allows users to quickly interpret snow distribution across the terrain using intuitive color scheme – from deep (red) to orange (less deep). This visual aid improves decision-making by making snow conditions immediately understandable at a glance. In addition to this, the iPad interface underwent several usability improvements, including refined e.g. touch interactions and clearer UI layout. These updates were driven directly by user feedback and aim to make the system more practical and efficient in real-world slope environments. Given the immersive capabilities of the Vision Pro, these features – particularly the snow depth visualization – are well-suited for adaptation into its 3D interface, where spatial awareness and intuitive feedback are even more critical for user engagement and operational effectiveness.

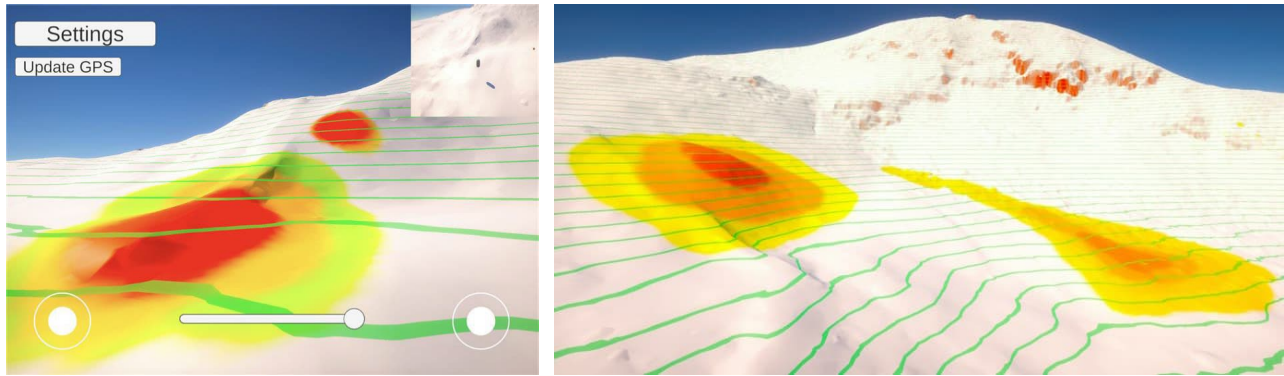


Figure 10. Updated iPad features e.g. a color-coded snow depth visualization

User feedback

An expert evaluation of the final prototype was conducted on June 3, 2025, at VTT's XR lab in Tampere with three participants, aiming to complement earlier end-user testing. Experts found the system intuitive and well-suited for snow grooming, appreciating its clear purpose and snow depth visualization. However, concerns were raised about tablet ergonomics, potential interface complexity, and the red color in the snow depth feature being misinterpreted due to undefined ranges. While the main and settings screens were praised for clarity, extensive customization and joystick-like controls were seen as less user-friendly. Calibration was noted as essential for trust, and suggestions included adding annotation tools and multi-user sharing to enhance value.

Conclusion

The AR-based situation awareness approach proved instrumental in designing a flexible, dual-platform system that supports both mobile AR and immersive VR. The iPad Pro M4 enabled real-time AR overlays and spatial data visualization in the field, while the Apple Vision Pro offered a fully immersive environment for simulation, planning, and spatial analysis. Seamless transitioning between these modes—rather than treating them as separate systems—created a more cohesive and context-aware user experience, aligning with the continuum principle that mixed reality is a fluid spectrum adaptable to user needs and environments.

The implementation demonstrated how real-time sensor data, spatial mapping, and interface design can enhance situational awareness, decision-making, and training. User testing at Prinoth's premises directly informed improvements such as color-coded snow depth visualization and UI refinements for gloved use and variable lighting, boosting both usability and the system's ability to present machine data intuitively.

Expert evaluation of the final version at VTT's XR lab confirmed the system's strong usability and relevance for snow grooming professionals. While the intuitive design and data clarity were praised, the snow depth visualization raised interpretational issues due to ambiguous color meanings. Experts recommended adding annotation tools and multi-user sharing features to further enhance the system's value.

Acknowledgements



**Funded by
the European Union**

This project has received funding from the European Union Horizon Europe research and innovation programme under grant agreement No 101092861. 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 Commission. Neither the European Union nor the granting authority can be held responsible for them.

References

- Helin, K., Alesani, A., Kuula, T., & Goriachev, V. 2024. Early-stage user experience design of the remote operation concept of the harbour's reachstacker by exploiting eXtended Reality. In Proceedings of 15th International Conference on Applied Human Factors and Ergonomics (AHFE 2024) and the Affiliated Conferences: Usability and User Experience (Vol. 156) <https://doi.org/10.54941/ahfe1005411>
- ISO 9241-210. 2019. Ergonomics of human-system interaction. Part 210: Human-centred design for interactive systems. European Committee for Standardization.

GAIA Project: Enhancing Awareness of Gender Harassment in Higher Education Settings Through Immersive Training

Kalliopi-Evangelia Stavroulia¹, Konstantinos Theodoridis², Dimitris Vagianos²,
Damianos Gavalas³, Vlasis Kasapakis³, and Angeliki Gazi²

¹*CYENS Centre of Excellence, Cyprus*

²*Panteion University of Social and Political Sciences, Greece*

³*University of the Aegean, Greece*

Corresponding author: k.stavroulia@cyens.org.cy

Keywords: GAIA, gender-based harassment, virtual reality

Introduction

This paper presents a novel Virtual Reality (VR) based training approach to address women's harassment within universities. Gender-based harassment in the higher education sector is a critical challenge that affects students, faculty, and staff. These forms of harassment undermine the goals of education, which should foster safe, inclusive, and respectful environments for learning, personal development, and academic achievement. Sexual harassment of women professors, researchers and staff by male colleagues is a common practice, with significant consequences for those affected (Bondestam & Lundqvist, 2020; Wood et al., 2021). For example, inappropriate comments or restricting women's independence are considered forms of harassment and violations of professional boundaries. These behaviors are correlated with depression, posttraumatic stress disorder (PTSD), isolation, and other negative effects undermining women's sense of safety and professionalism, impacting their success (Wood et al., 2021).

The GAIA project (Gender equality and anti-discrimination practices via virtual reality approaches) focuses on the academic environment, promoting systematic training and gender perspective for the next generation of active citizens, along with personal and professional empowerment of the target audience. GAIA aims to build on this change by leveraging VR-based training, as VR is considered a powerful tool for fostering empathy (Venture et al., 2020). GAIA proposes immersive experiences that allow users to experience firsthand women's sexual harassment within the academic sphere. The added value of the project is followed by its significant impact on the scientific and social levels. GAIA addresses one of today's major societal challenges, namely, the systemic gender discrimination and inequalities in higher education. The project addresses this gender gap and inequalities in higher education through a novel VR-based approach that aims to eliminate gender bias within the university through experiential training and cultivation of skills such as empathy. Such a VR-based approach offers new research perspectives in the field and could form the basis for future initiatives towards inclusion and equality in

higher education. Equally important is that GAIA aims to train both students and faculty, promoting awareness and diversity within higher education.

The use of VR for sexual harassment awareness and training is still emerging. While VR has shown promising results in fostering empathy-building and perspective-taking in contexts like racial bias (Salmanowitz, 2018), bullying (Gu, 2022), and violent incidents (Rovira et al., 2009), its application to sexual harassment training remains underexplored, particularly in academic contexts. Some studies have begun to address this gap. Ventura et al. (2021, 2022) used a 360° VR approach to explore the impact of perspective change and entering the position of a woman victim of harassment, with the results indicating an impact of the immersive experience on participants' empathy. Sadeh-Sharvit et al. (2021) found that women participants felt stress and fear during a VR job interview harassment scenario. In another study, Steinbauer (2024) reported that VR training impacted students' emotional awareness of workplace harassment. Research results indicate the growing efforts to explore the use of VR training to address sexual harassment because of the immersive nature of this technology, which can make scenarios feel more realistic and emotionally impactful. The GAIA project, by taking advantage of VR technology, developed a novel VR-based training application tailored to the needs of academic contexts, reflecting real-life and in-person harassment scenarios. By placing participants in realistic harassment scenarios, they can "feel" the situation, promoting empathy, perspective-taking, and a deeper understanding of the impact of harassment.

Scenario Design and Technical Specifications

The GAIA application was developed to offer immersive, first-person perspective experience designed to place users in realistic harassment scenarios for empathy-building and awareness training. The scenario begins in the office of a senior male professor, called Giannis, with a junior female professor, called Maria, entering to discuss her potential promotion. Initially, Giannis is warm and casual, but he quickly shifts to making personal comments about Maria's high heels that soon become increasingly inappropriate. He becomes bolder, constantly proposing that he and Maria should go on a professional trip together to Paris, disregarding her discomfort. He ultimately creates a hostile and uncomfortable environment as he locks the door of the office and physically approaches her closely (Figure 11a). The VR training experience is designed as a passive scenario, with no interactive elements or user-driven influence. The narrative unfolds through a scripted dialogue between the avatars of Giannis and Maria. As the conversation evolves, what begins as a seemingly friendly interaction gradually reveals itself as a case of sexual harassment, allowing participants to observe the progression of inappropriate behavior. The VR experience takes place within an approximately five-minute timeframe.

The GAIA scenario unfolds in two stages. In the first stage, participants experience the scenario from Maria's perspective, using a first-person point of view. In the second stage, participants switch to third third-person perspective to observe the same scenario from the perspective of an external observer of the incident. This dual-perspective approach helps end-users experience harassment from different perspectives, enhancing their empathy, awareness, and understanding. In the position of the female professor Maria, users experience what it is like to be in a vulnerable position, witnessing how quickly discomfort can escalate as they try to keep the conversation professional, while Giannis increasingly behaves like a harasser, even daring to lock the door. On the other hand, viewing the scenario from a third-person perspective makes harassment behaviors easier to spot. Participants can more critically analyze Giannis' body language and verbal cues, like tone shifts, physical proximity, or disregard for personal boundaries. Moreover, experiencing the situation as an observer highlights the role of a third person in recognising and responding to potential harassment.

To develop the GAIA application, Unity was used to build the virtual office environment, ensuring accurate spatial alignment by modelling it with real-world dimensions. Character Creator by Reallusion was employed to design detailed 3D character models. Full-body animations were recorded using the OptiTrack Motion Capture system, while facial expressions were captured with the Reallusion iPhone Facial Mocap app. High-quality voice audio was recorded using Rode wireless microphones. All motion, facial, and voice data were integrated into Reallusion iClone Studio using the OptiTrack and Facial Mocap Motion Live Plugins, along with the software's built-in voice recording tools (Figure 11b). The completed animation assets were then exported to Unity, where Animator Controllers managed animation playback and AudioSource components handled voice playback. To support the Oculus Quest 2 HMD, the XRI Toolkit for Unity was utilized, enabling consistent and recentered user perspectives across sessions.

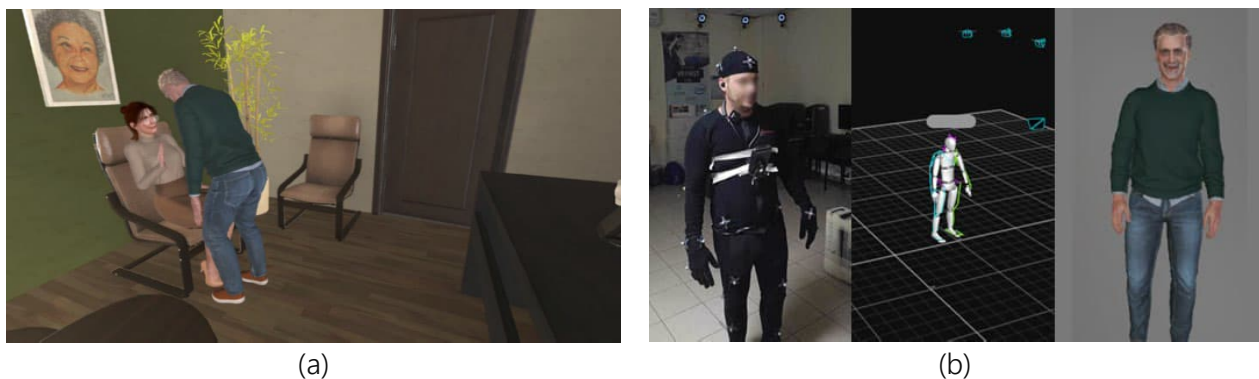


Figure 11. (a) The characters of Maria and Giannis (from the observers' perspective); (b) Animation process for the 3D model of Giannis.

The VR application is currently being implemented to evaluate its impact, with more than 70 higher education students and faculty participating so far. Preliminary observations reveal that the scenario elicits strong emotional responses from participants, regardless of gender. Several participants, both male and female, showed avoidance behaviors when Giannis' avatar approached Maria's avatar closely. These behaviors included physical gestures and attempts to move their bodies away. In some cases, female participants were unable to complete the scenario, experiencing emotional distress, panic, and crying, which led to the termination of their participation. The initial data collected from the pre- and post-surveys appear promising. A comprehensive analysis will be conducted upon the completion of the implementation phase.

Conclusions

GAIA aims to address gender-based harassment within higher education and contribute to women's empowerment and change of attitudes towards harassment. To achieve this goal, the project proposes the use of a VR-based training approach that reflects real-life scenarios and allows end-users to experience firsthand the harassment of a female academic by a male colleague. The VR implementation is still ongoing, with observations indicating high emotional responses from the participants. The results will provide insights into the potential impact of the VR intervention, as well as its integration into higher education settings.

Acknowledgements

This project has received funding by the Hellenic Foundation for Research and Innovation, under the framework Greece 2.0 National Recovery and Resilience Plan "Basic Research Financing" (Horizontal support for all Sciences) (Project ID: 16297).

References

- Bondestam, F., & Lundqvist, M. 2020. Sexual harassment in higher education—a systematic review. *European Journal of Higher Education*, 10(4), 397-419. <https://doi.org/10.1080/21568235.2020.1729833>
- Gall, D., Roth, D., Stauffert, J. P., Zarges, J., & Latoschik, M. E. 2021. Embodiment in virtual reality intensifies emotional responses to virtual stimuli. *Frontiers in Psychology*, 12, 674179. <https://doi.org/10.3389/fpsyg.2021.674179>
- Gu, X., Li, S., Yi, K., Yang, X., Liu, H., & Wang, G. 2022. Role-exchange playing: An exploration of role-playing effects for anti-bullying in immersive virtual environments. *IEEE transactions on visualization and computer graphics*, 29(10), 4215-4228. <https://doi.org/10.1109/TVCG.2022.3184986>
- Rovira, A., Swapp, D., Spanlang, B., & Slater, M. 2009. The use of virtual reality in the study of people's responses to violent incidents. *Frontiers in Behavioural Neuroscience*, 3, 917. <https://doi.org/10.3389/neuro.08.059.2009>
- Sadeh-Sharvit, S., Giron, J., Fridman, S., Hanrieder, M., Goldstein, S., Friedman, D., & Brokman, S. 2021. Virtual reality in sexual harassment prevention: proof-of-concept study. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (pp. 87-89).
- Salmanowitz, N. 2018. The impact of virtual reality on implicit racial bias and mock legal decisions. *Journal of Law and the Biosciences*, 5(1), 174-203. <https://doi.org/10.1093/jlb/lisy005>
- Ventura, S., Badenes-Ribera, L., Herrero, R., Cebolla, A., Galiana, L., & Baños, R. 2020. Virtual reality as a medium to elicit empathy: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, 23(10), 667-676. <https://doi.org/10.1089/cyber.2019.0681>
- Ventura, S., Cardenas, G., Miragall, M., Riva, G., & Baños, R. 2021. How does it feel to be a woman victim of sexual harassment? The effect of 360-video-based virtual reality on empathy and related variables. *Cyberpsychology, Behavior, and Social Networking*, 24(4), 258-266. <https://doi.org/10.1089/cyber.2020.0209>
- Ventura, S., Miragall, M., Cardenas, G., & Baños, R. M. 2022. Predictors of the sense of embodiment of a female victim of sexual harassment in a male sample through 360-degree video-based virtual reality. *Frontiers in human neuroscience*, 16, 845508. <https://doi.org/10.3389/fnhum.2022.845508>
- Wood, L., Hofer, S., Kammer-Kerwick, M., Parra-Cardona, J. R., & Busch-Armendariz, N. 2021. Sexual harassment at institutions of higher education: Prevalence, risk, and extent. *Journal of interpersonal violence*, 36(9-10), 4520-4544. <https://doi.org/10.1177/0886260518791228>

Haptic Surgical Training Based on Recorded Data

Jerome Perret, and Anna Perret

Haption GmbH, Germany

Corresponding author: jerome.perret@haption.com

Keywords: Force-feedback, surgical training, virtual fixtures

Abstract

Haptic feedback is an important factor for the effectivity of surgical training in a virtual environment. Especially, the integration of realistic force-feedback based on physics simulation is a complex technical challenge. In this paper, we propose an alternative based on virtual fixtures and recorded data. By partially locking a tool on an axis or restraining it inside a cone, it is possible to place the user in a similar context as during the data recording. Then, the data replay gives a realistic illusion of the action. We demonstrate our approach for two different actions: bone drilling and screw fastening.

Introduction

Force-feedback is a very important sensory modality for the effectivity of surgical training in a Virtual Environment (VE) (Overtoom et al., 2019). However, the integration of force-feedback based on realistic human tissue models is a very complex technical challenge. In addition, the simulation realism, often referred to as "face validity", is not a good indicator in itself for the usefulness of a training system (Vander Poorten et al., 2014). The design of the training system should thus focus on optimizing the dimension of "transfer of training", focusing the development effort on the key learning cues.

One alternative to complex physics simulation is to render force-feedback based on recorded data. For example, the nature of a material can be conveyed by playing a recorded signal upon impact, provided the recording was done in similar impact conditions (Kuchenbecker et al., 2006). However, recording event-dependent force information is a real challenge in most practical cases: depending on the type of activity, the number of possible events could be almost infinite, and the real material is usually scarce (e.g. cadavers).

In this paper, we present the approach adopted for two surgical procedures addressed as part of the collaborative project VIRTOSHA (MindPort, 2025), combining virtual fixtures and recorded data.

Problem statement

The purpose of the project VIRTOSHA is to develop a VE in which students can train surgical procedures in a safe and controlled environment. The chosen operation is an orthopedic surgery, more

specifically the treatment of an ankle fracture by open reduction and internal fixation (ORIF). Ankle fractures are very common, and if not treated properly they can lead to serious reduction of mobility or even to forced amputation. In this paper, we focus on the step called retention, which consists in holding the bones in position and creating a defined pressure between fragments by means of a lag screw. More specifically, the surgeon must train to:

1. Drill a hole through both bone fragments at 90 degrees of the fracture, without pushing through the second fragment. Should the drill exit the second fragment, then it might damage tissue behind and jeopardize the healing process. A skilled surgeon is able to know when to stop drilling by monitoring the resistance of the material.
2. Insert and fasten the screw in the drilled hole without exceeding the mechanical limit of the bone structure. The screw cuts its own thread in the bone structure, and if it is fastened more than the bone structure can bear, then the thread breaks and the screw does not hold properly. Again, a skilled surgeon is able to know when to stop fastening by monitoring the resistance torque on the screwdriver.

While it might be possible to simulate the behavior of the tool (drill or screw) when penetrating the bone based on a volumetric model, it would be both extremely computing-intensive and probably irrelevant. Indeed, the behavior depends a lot on the sharpness of the tool, how the saw dust accumulates, how humid it is, etc. According to expert surgeons, it is much more important to learn the shape of the behavior curve and how it may vary between individuals, depending on gender, age, and clinical condition.

On the contrary, it is relatively straightforward to record the behavior of the tool when penetrating the bone, for example on cadavers. Figure 13 shows the penetration speed of the drill in a beef rib, recorded by fixing the drill at the tip of our force-feedback device and monitoring its position while applying a constant force of 10 N (Figure 12).



Figure 12. Recording of the drilling progression speed with a Virtuose 6D High-Force device.

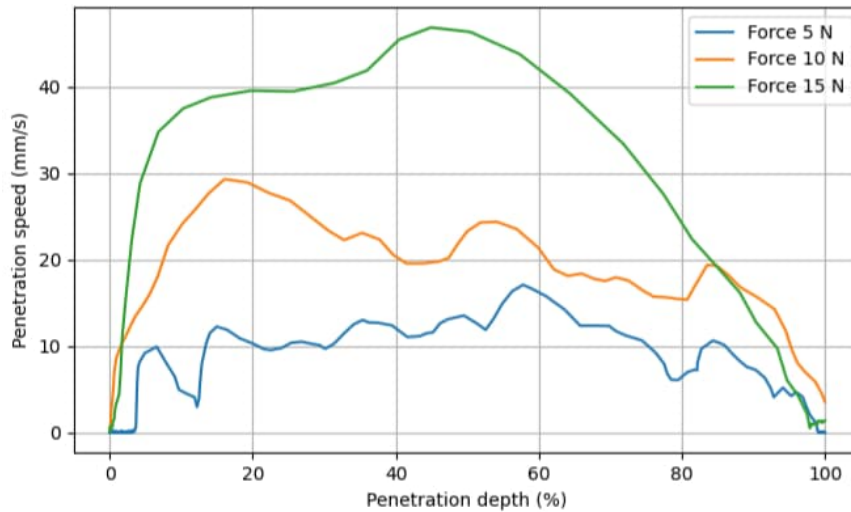


Figure 13. Records of drill penetration speed for the drilling depth with constant force.

Figure 14 shows the resisting torque of the screw when fastening it in a beef rib, recorded by fixing a ratchet screw driver to our force-feedback device and applying an alternating rotation at constant speed. In both cases, the figures show a single profile for a single dimension and a constant parameter (applied force or speed), so it is only valid if the trainee follows the exact same scenario (perfect alignment and constant force). Therefore, the problems to be addressed are: i) how to make sure that the user follows this exact scenario ii) how to handle variations of the parameters (e.g. bone radius, incident angle).

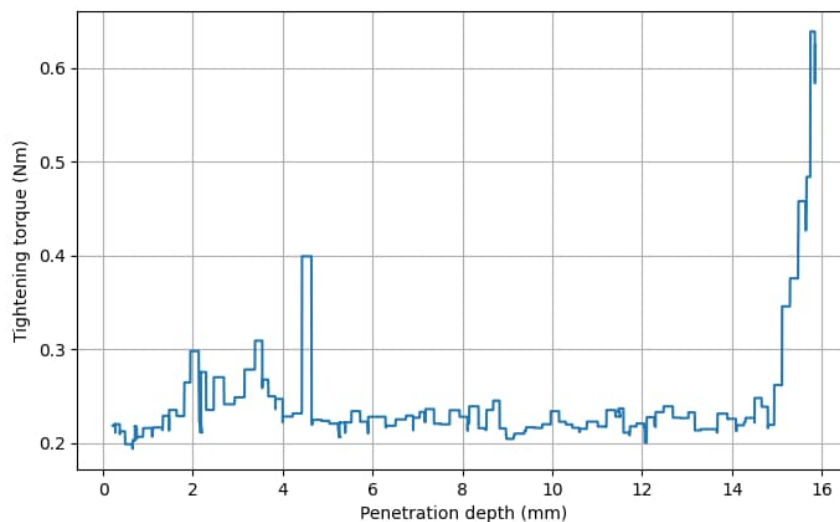


Figure 14. Record of resistance torque while fastening the screw with rotation speed 60 deg/s.

Methodology

In order to offer a correct drilling and screw fastening experience, we use virtual fixtures inspired by the approach of (Maier et al., 2019), combined with custom physical props attached to the end-effector of the haptic device. The physical props are identical to the handles of the real surgical instruments, so as to produce the right tactile feedback. Hereafter, we focus on the description of the virtual fixtures and on the replay of recorded data.

In the case of drilling: after the initial contact with the bone, the drill can only rotate around the contact point; then after penetrating into the bone, it can only rotate along its own axis. Actually, during the whole drilling process, we can consider that the drill is attached to a cylindrical joint with a limit on the sliding part, that being the drilling depth given by integrating the speed curve shown in Figure 1. In addition, the drill action creates a small reaction torque, counter-clockwise, which only needs to be applied as an overlay to the cylindrical joint rotation while the drill depth is progressing. Overall, the succession of virtual fixtures is almost identical to the state machine shown in (Maier et al., 2019).

In the case of screw fastening: after the screw has started cutting the thread in the bone, it is constrained within a cone, which angle is diminishing with the fastening progress. After a couple of turns, we can consider that the screw is attached to a helical joint with a constant helix angle equal to the screw thread angle (Figure 15). The only input from recorded data is the resisting torque, which varies with the screw progress in the bone.

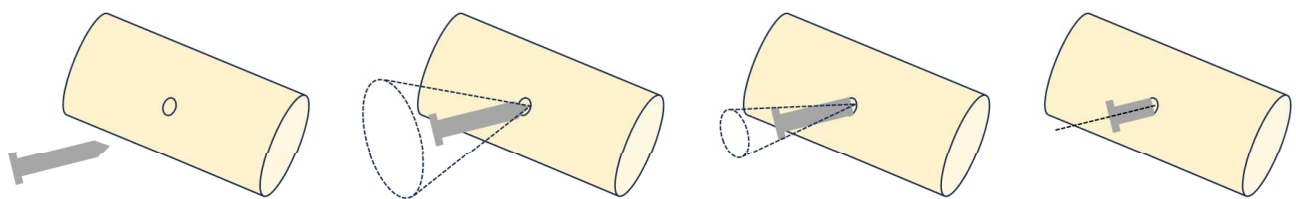


Figure 15. Succession of virtual fixtures (left to right): free space, cone with diminishing radius, helical joint.

In both cases, the variation of the parameters is addressed by linear interpolation between curves recorded with different forces (drilling) or speeds (screw fastening), after normalizing for the bone radius or the screw length.

Results

This system has been tested by an expert surgeon who is part of the project team, and has been deemed realistic. In the case of drilling, the surgeon stated that it was possible to feel when the drill is inside each part of the bone structure, enabling him to know when to stop the drilling procedure (Figure 16). In the case of fastening, the surgeon was able to tell when the screw was passing each part of the bone structure, when the final tightening was happening, thus enabling him to stop before the screw threads would break.

By using this simulation, medical staff will be able to try surgical procedures, and experience changes in sensation that are similar to the ones in a real operation. We have strong indications that the system will make transfer of training possible, and that medical staff can acquire operational skills more effectively than with synthetic bone.



Figure 16. Second author testing the haptic rendering (without physical prop).

Conclusion and Future Work

The key takeaway of our research on haptic surgical training for this project is that it is possible to meet training requirements without a huge investment in high-fidelity physics simulation. Instead, we demonstrate that we can record the behavior of the tool in controlled conditions, and reproduce that behavior in Virtual Reality by forcing the trainee to follow a controlled scenario by use of virtual fixtures.

However, we want to stress that the input of an expert surgeon is invaluable in this approach, in order to identify the critical cues for the transfer of training, and supervise the progress throughout the development.

As it is not possible to drill through a bone at the same position multiple times, the recorded curves in Figure 13 have differences in shapes due to the difference of the bone structure. Because of this, interpolating between them will not be accurate. While one approach would be to drill at the same position in multiple bones of the same type, this is not feasible in practice, and bones from different individuals would also have different properties. A better approach for future work is to use an ideal curve that has the typical shape an expert surgeon would expect in that scenario, and to use real recordings solely for calibrating the curve.

During the future course of the project, we are planning to refine the recording process with more animal bones, before applying it to a small number of human cadavers. Then, our haptic training module will be integrated in a complete XR training system, which will be subject to a user study.

Acknowledgement

This work was carried out within the collaborative project VIRTOSHA. VIRTOSHA is part of the Gesünder.in.NRW EFRE program and funded by the IN.NRW Innovationsförderagentur, the MKW NRW, and the European Union.

References

- Overtoom, E., Horeman, T., Jansen, F.-W., Dankelman, J. & Schreuder, H. 2019. Haptic Feedback, Force Feedback, and Force-Sensing in Simulation Training for Laparoscopy: A Systematic Overview, *Journal of Surgical Education*, Volume 76, Issue 1, 2019, Pages 242-261, ISSN 1931-7204, <https://doi.org/10.1016/j.jsurg.2018.06.008>
- Vander Poorten, E., Perret, J., Muyle, R., Reynaerts, D., Vander Sloten, J. & Pintelon, L. 2014. To Feedback or not to Feedback – the Value of Haptics in Virtual Reality Surgical Training. Conference and Exhibition of the European Association of Virtual and Augmented Reality (EuroVR 2014), Bremen, Germany.
- Kuchenbecker, K. J., Fiene, J. & Niemeyer, G. 2006. Improving contact realism through event-based haptic feedback. *IEEE Trans Vis Comput Graph*. 2006 Mar-Apr;12(2):219-30. <https://doi.org/10.1109/TVCG.2006.32>
- MindPort 2005. <https://www.mindport.co/research/virtosha> accessed on May 29, 2025.
- Maier, J., Perret, J., Huber, M. & Simon, M. 2019. Force-feedback assisted and virtual fixtures based K-wire drilling simulation. *Computers in Biology and Medicine*, September 2019, <https://doi.org/10.1016/j.combiomed.2019.103473>

A Virtual Reality Stroop Art Exergame for Cognitive-Motor Rehabilitation in Chronic Stroke Survivors

Christos Hadjipanayi¹, Domna Banakou^{1,2}, Abraiz Azhar³, Dimitris Sokratous^{1,4},
and Despina Michael-Grigoriou^{1*}

¹*GET Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Cyprus*

²*Interactive Media, Arts and Humanities Division, New York University Abu Dhabi, UAE*

³*Computer Science, Division of Science, New York University Abu Dhabi, UAE*

⁴*Physiotherapy Unit, Neurology Clinics, The Cyprus Institute of Neurology and Genetics, Cyprus*

*Corresponding author: despina.grigoriou@cut.ac.cy

Keywords: virtual reality, rehabilitation, exergames, artmaking, social facilitation

Introduction

This work introduces a prototype VR exergame (Hadjipanayi, et al., 2024) that blends creative artmaking with social facilitation elements and integrates a VR adaptation of the Stroop test; already a staple pen-and-paper counterpart exercise for promoting cognitive flexibility in stroke therapy. Recent reviews on VR artmaking for well-being revealed the applicability of VR artmaking to rehabilitation (Hadjipanayi, Banakou, et al., 2025; Hadjipanayi, et al., 2023). Yet most interventions rely on bespoke software and hardware (Baron, et al., 2021; Iosa, et al., 2021; Zhang et al., 2021). Among the few commercial tools mentioned, "Google Tilt Brush" (used with Oculus or VIVE headsets) appeared in (Alex, et al., 2021), where it served primarily as a community-building tool among stroke survivors rather than direct rehabilitation. In surveying the current marketplace, we found no readily deployable VR artmaking application that adequately meets the ergonomic and cognitive needs of chronic-stroke survivors—underscoring the need for a purpose-built solution.

A VR Rehabilitation Exergame integrating the Stroop Test and Artmaking

Following a systematic evaluation of four commercial VR platforms – VRChat, Mozilla Hubs, Multibrush, and Hand Brush with Friends—conducted alongside an experienced physiotherapist, none met the minimum clinical and usability thresholds required for an artmaking rehabilitation intervention (**Error! Reference source not found.**). VRChat and Mozilla Hubs achieved strong social co-presence yet lacked flexible gesture mapping and fine-grained environmental control. Multibrush VR delivered the richest drawing toolkit but was restricted to Quest hardware and imposed a very high cognitive load. Hand Brush offers a low-complexity interface but sacrificed online collaboration and full avatar embodiment. Across all candidates, the absence of intuitive gesture-based interaction – a prerequisite for motor-impaired users – and limited affordances for manipulating virtual agents or the scene

emerged as critical shortcomings. These gaps motivated the creation of the solution we present here: a VR exergame that gamifies the Stroop test (Jensen and Rohwer, 1966) and integrates collaborative artmaking features expressly geared toward the rehabilitation of chronic stroke.

Table 2 Affordance comparison of four commercial VR artmaking applications with the potential of being used as part of a chronic stroke rehabilitation regimen.

Affordances	VRChat	Hand Brush with Friends	Mozilla Hubs	Multibrush VR Multiplayer version of 'Google Tilt Brush'
Device compatibility	Most VR devices	Oculus Quest	Most VR devices	Oculus Quest
Embodiment	Allowed	Not found	Allowed	Waist-up only
Avatar options	Custom avatar from photo	Not found	Custom avatar from upload	Meta's avatars only
Virtual Agent options	Not Available	Not Available	Not Available	Not Available
Drawing options	Gesture-based	Not found	Not available	Controller-based
Drawing tool options	Not found	Not found	Not found	Allowed
Pressure sensitivity	Not found	Not found	Not found	Allowed
Drawing space	3D	3D	3D	3D
Color palette	Not found	Allowed	Not found	Allowed
Eraser options	Gesture-based	Not found	Not available	Controller-based
Audio	Allowed	Not found	Allowed	Allowed
Open source	Not available	Not available	Available	Not available
Collaborate online	Yes	Not found	Yes	Yes
Level of Cognitive Load	High (but can be regulated)	Low	High (but can be regulated)	Very High

While motor impairments due to stroke usually tend to improve with time, the accompanying cognitive decline tends to linger and can markedly undermine stroke survivors' well-being long-term (Mijajlović et al., 2017). Targeted cognitive training can stall the occurrence of adverse neurodegenerative effects (Rost et al., 2022), making VR an appealing medium to explore – especially for dementia prevention in high-risk populations. Gradl et al. (2019) propose that a gamified Stroop test within multiplayer virtual environments, where users feel socially observed, could reveal new insights about how VR stressors influence executive function performance. The Stroop test measures interference, the condition in which two separate cognitive functions, namely the automatic “word reading” and non-automatic “color naming”, conflict due to incongruent stimuli (Belletier et al., 2019). It is hypothesized that the added cognitive load of co-acting peers may divert attention from the incongruity of these cues, thus reducing interference (Muller and Butera, 2007). Recent evidence links Stroop interference to fluctuations in attentional focus, affective state, and neural circuits involved in social cognition (Arioli et al., 2021), underscoring the value of investigating this paradigm within socially immersive VR environments while focusing on the rehabilitation of chronic stroke survivors (Hadjipanayi, Sokratous, et al., 2025).

The Design of the Exergame

The prototype is a single-player, first-person exergame that embeds a Stroop-style color-naming task within a creative artmaking scenario. The player enters a pleasant virtual environment, designed to inspire, while embodying a humanoid avatar (Figure 1A). Positioned in front of an easel-mounted canvas,

the user can freely draw using the index finger of the impaired upper limb (Figure 17B) while the contralateral hand acts as an eraser. Across the avatar stands a virtual human - remotely controlled by the session administrator through a screen menu - who can deliver guidance or encouragement as needed (Figure 17A). The art studio can be populated with other virtual humans occupying neighboring easels to provide a sense of social co-presence. Gameplay begins when the user presses a start button beside the canvas (Figure 17C). The virtual human welcomes the user to the session and provides gameplay instructions. Each prompt points towards either a color word-naming (automatic response) or color ink-naming (non-automatic response) request in a fashion reminiscent of the Stroop test (Figure 17D). For each prompt, the player has a configurable time window to choose the appropriate hue from a palette and render a drawing. The drawing should ideally portray the object-word of the displayed phrase (Figure 17D), but the user is free to draw whatever comes to mind, as the fidelity of the painted object to its real-life counterpart is irrelevant to the objectives of this game and game points accrue solely from brushstrokes executed with the correct color. Rounds repeat until the session ends.

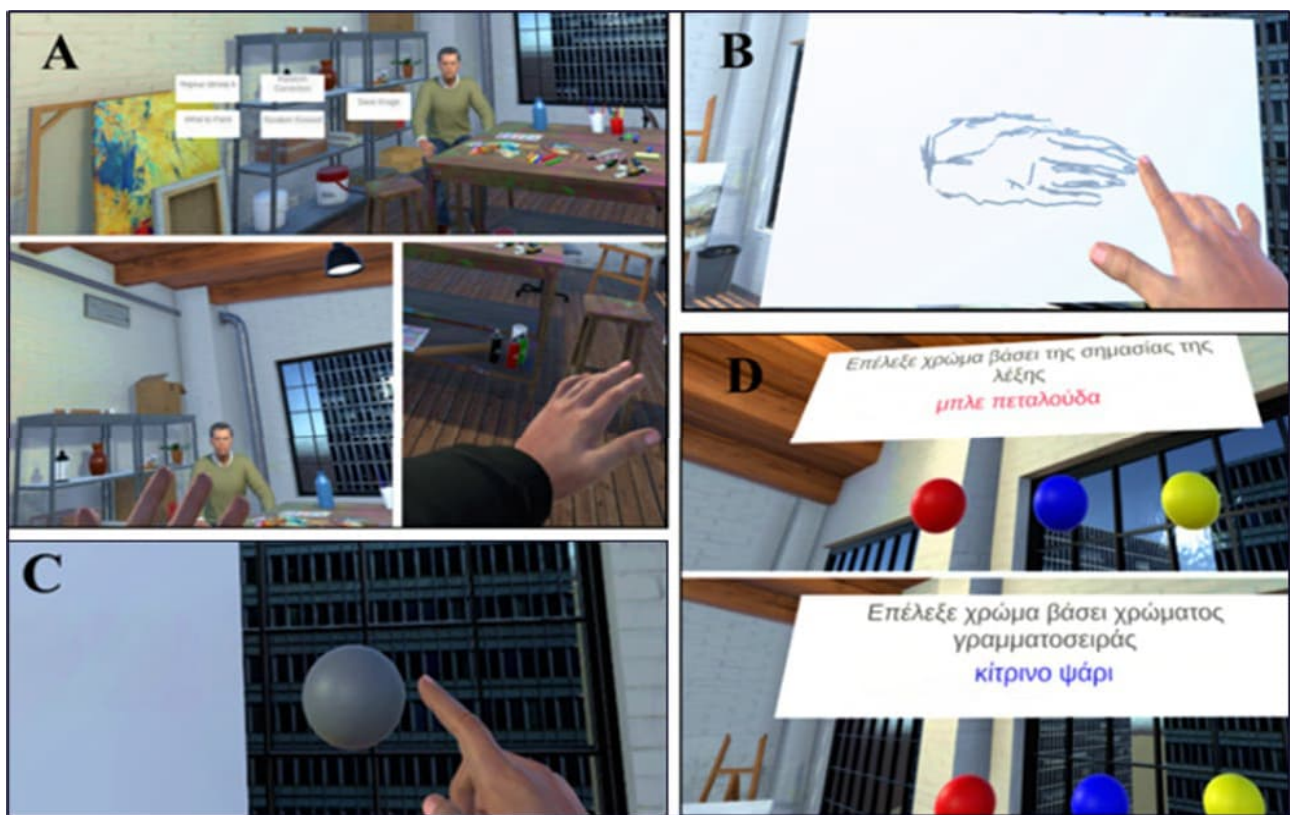


Figure 17. The environment and gameplay of the featured VR application. **A.** Interactions between the player and the virtual human. **B.** Free drawing on the canvas. **C.** Starting the exergame. **D.** Instances of the Stroop task. **D (top).** "Choose color based on the word meaning: blue butterfly" (translated from Greek), where 'blue butterfly' appears with red font color. **D (bottom)** "Choose color based on the font color: yellow fish" (translated from Greek), where 'yellow fish' appears with blue font color.

The prototype currently runs on an Oculus Quest 3 headset, employing its upper-body tracking for avatar control and its native finger-tracking to drive the drawing interaction.

Conclusions

The prototype is now in pilot trials with stroke survivors to gauge its therapeutic impact and acceptability for both patients and physiotherapists. Results from this phase will inform a larger study

that will compare the gamified application's effects on rehabilitation and well-being and assess its viability for future at-home use.

References

- Alex, M., Wünsche, B. C., & Lottridge, D. 2021. Virtual reality art-making for stroke rehabilitation: Field study and technology probe. *Int. J. Human-Computer Stud.* 145, 102481. <https://doi.org/10.1016/j.ijhcs.2020.102481>
- Arioli, M., Basso, G., Poggi, P., & Canessa, N. 2021. Fronto-temporal brain activity and connectivity track implicit attention to positive and negative social words in a novel socio-emotional Stroop task. *Neuroimage*, 226, 117580. <https://doi.org/10.1016/j.neuroimage.2020.117580>
- Baron, L., Wang, Q., Segear, S., Cohn, B. A., Kim, K., & Barmaki, R. 2021. Enjoyable physical therapy experience with interactive drawing games in immersive virtual reality. In *Symposium on Spatial User Interaction*, 1–8.
- Belletier, C., Normand, A., & Huguet, P. 2019. Social-facilitation-and-impairment effects: From motivation to cognition and the social brain. *Current Directions in Psychological Science*, 28(3), 260-265. <https://doi.org/10.1177/0963721419829699>
- Gradl, S., Wirth, M., Mächtlinger, N., Poguntke, R., Wonner, A., Rohleder, N., & Eskofier, B. M. 2019. The stroop room: A virtual reality-enhanced stroop test. In *Proceedings of the 25th acm symposium on virtual reality software and technology* (pp. 1-12)
- Hadjipanayi, C., Banakou, D., & Michael-Grigoriou, D. 2023. Art as therapy in virtual reality: A scoping review. *Frontiers in Virtual Reality*, 4, 1065863. <https://doi.org/10.3389/frvir.2023.1065863>
- Hadjipanayi, C., Banakou, D., & Michael-Grigoriou, D. 2024. Virtual Reality Exergames for Enhancing Engagement in Stroke Rehabilitation: A Narrative Review. *Heliyon*, 10(18), e37581.
- Hadjipanayi, C., Banakou, D., & Michael-Grigoriou, D. 2025. Virtual Reality in Stroke Rehabilitation and Treatment: Focusing on the Patient's Experience and Needs. In: Charalambous, A. (eds) *Critical Perspectives on Technological Innovations in Healthcare*. Springer, Cham.
- Hadjipanayi, C., Sokratous, D., Kyriltsias, C., Banakou, D., & Michael-Grigoriou D. 2025. Social facilitation within immersive virtual reality enhances perseverance in stroke rehabilitation training. *Front. Virtual Real.* 6:1581240.
- Iosa, M., Aydin, M., Candelise, C., Coda, N., Morone, G., Antonucci, G., ... & Tieri, G. 2021. The michelangelo effect: art improves the performance in a virtual reality task developed for upper limb neurorehabilitation. *Frontiers in psychology*, 11, 611956.
- Jensen, A. R., & Rohwer Jr, W. D. 1966. The Stroop color-word test: a review. *Acta psychologica*, 25, 36-93.
- Mijajlović, M. D., Pavlović, A., Brainin, M., Heiss, W. D., Quinn, T. J., Ihle-Hansen, H. B., ... & Bornstein, N. M. 2017. Post-stroke dementia—a comprehensive review. *BMC medicine*, 15, 1-12.
- Muller, D., & Butera, F. 2007. The focusing effect of self-evaluation threat in coaction and social comparison. *Journal of Personality and Social Psychology*, 93(2), 194.
- Rost, N. S., Brodtmann, A., Pase, M. P., van Veluw, S. J., Biffi, A., Duering, M., ... & Dichgans, M. 2022. Post-stroke cognitive impairment and dementia. *Circulation research*, 130(8), 1252-1271. <https://doi.org/10.1161/CIRCRESAHA.122.319951>
- Zhang, Y., Wang, H., & Shi, B. E. 2021. Gaze-controlled robot-assisted painting in virtual reality for upper-limb rehabilitation, in 2021 43rd Ann. Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE), 4513–4517.

Explaining the Future of Work: WIRKsam XR Demonstrators

Yücel Uzun, Urs Riedlinger, and Leif Oppermann

Fraunhofer FIT, Sankt Augustin, Germany

Corresponding author: yuecel.uzun@fit.fraunhofer.de

Keywords: XR, AI, future of work, demonstrator

Introduction

WIRKsam Competence Center supports regional companies in the Rhenish region's textile and related industries in leveraging artificial intelligence (AI) to shape the future of work and strengthen their competitiveness. As part of this project, we are currently developing a series of XR demonstrators to explain the project to the general public, potential new partners, and current employees of existing partners. While the use of XR in outreach and marketing is not novel (Oppermann et. al. 2024, Atiker et. al. 2023), our approach must carefully address the sensitive and often controversial topic of AI, particularly concerning potential job displacement and convince users of its added value.

Given our goals, storytelling plays a crucial role in these demonstrators. Our first two demos are designed to address both the general public and potential new partners, who need to understand the value and impact of AI in the industry, a message most effectively conveyed by showing real-world tasks both with and without AI support. However, simply replicating mundane tasks in VR would not result in an engaging user experience. To address this, we incorporated gamification elements and provided users with feedback on their performance in different scenarios, making the experience both informative and motivating.

The XR demonstrators are developed iteratively using the Unity game engine and the OpenXR framework to ensure cross-device compatibility and long-term maintainability. They are designed as single-user virtual reality experiences, currently displayed on Meta Quest 3 VR headsets. Interaction is primarily based on hand tracking to promote intuitive use, while the controllers are supported as an accessible alternative. Additionally, user scores, statistics, and other interaction data from the gamification elements are sent to a custom Node.js server and can be displayed through an Angular-based web interface. This setup allows us to monitor user activity in real time and showcase high scores, making public events more appealing.

Recycling Demo

To initiate the development of XR demonstrators, a use case was designed independently of direct industry partner input due to scheduling constraints and limited early availability of external stakeholders. While not based on a specific industrial scenario, it reflects the overarching themes of the

project and provides a foundation for evaluating technical feasibility, user interaction strategies, and communication approaches.

In this initial demonstrator, users participate in a sorting task within an imagined recycling facility during a sorting process. The objective is to separate metallic objects from plastic bottles. The experience begins with a short training phase, during which instructions are presented via a virtual display, read aloud by a voice assistant, and reinforced through spatial visual cues. Users then perform the sorting task until they sort 10 metals or make 10 mistakes without AI support, followed by a second round in which an AI system assists by highlighting metallic objects on the conveyor belt. Throughout both phases, performance metrics – such as correctly sorted metals, incorrectly removed plastic items, and missed metallic objects – are tracked and displayed to the user. This approach is intended to illustrate the potential benefits of AI assistance, while also emphasizing the complementary role of human and machine collaboration, rather than automation-driven replacement.

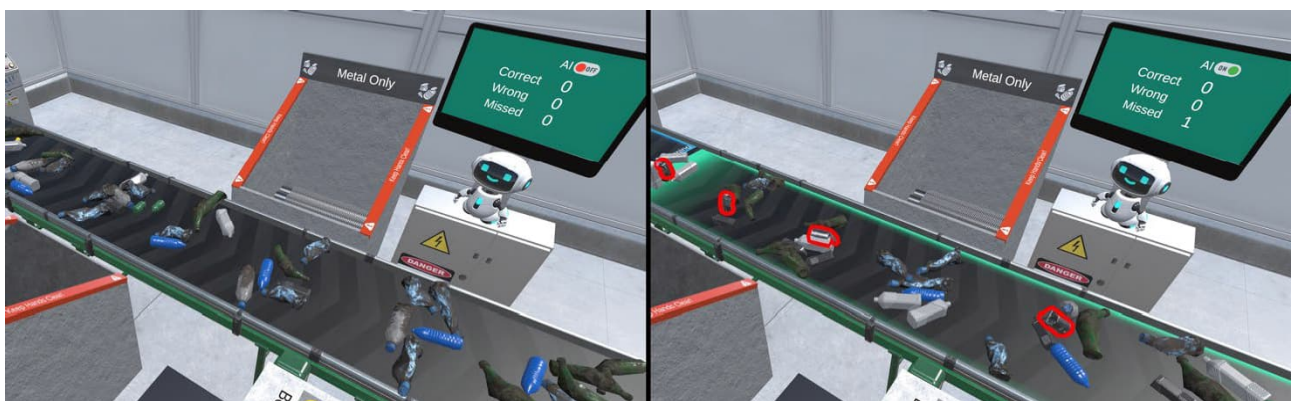


Figure 18: Left: Sorting task performed without AI assistance, illustrating the challenge of distinguishing plastics from metals. Right: Sorting task with AI assistance, where metallic objects are highlighted in red for easier identification.

Wing Inspection Demo

For the second iteration, part of a real-world use case from within the WIRKsam project was implemented: defect detection in the production process of automotive wings. In the actual industrial scenario, project partners use cameras to identify various types of defects, classify them, and initiate appropriate responses. Building on this, the XR demonstrator replicates a simplified version of the inspection process using a real CAD model of the produced car wings.

Following the structure established in the recycling demo, users begin with a brief training session before performing the inspection task twice, first without and then with AI assistance. The AI system supports the user by visually highlighting potential defect areas, though it is intentionally designed to be imperfect, occasionally producing false positives or missing some errors. This setup encourages users to remain attentive and not rely entirely on the AI, emphasizing the importance of human oversight.

To ensure accessibility and maintain user engagement, the demonstrator avoids complex classification tasks from the real use case, which would require domain-specific knowledge and significantly extend the interaction time. Instead, users identify errors by simply pointing at them with their index fingers and confirming their input via button presses. To accommodate inspection of the wing's underside, users can trigger a robotic arm to rotate the model by pressing a designated button, which also does not exist in the real case. Unlike the recycling demo, this version introduces a time constraint: users have 90 seconds to complete the inspection task, adding a layer of urgency to simulate

real-world pressure and decision-making, while ensuring that each demo session remains brief – an important consideration for large public exhibitions where maximizing reach and throughput is essential. User performance is evaluated by comparing correct detections, false detections, and identifying AI mistakes.



Figure 19: Left: Overview of the environment. Wings are transported by driverless systems, and a robotic arm is used to rotate the wings; the user controls these functions using buttons on the left. The robot and information display are visible on the right side of the image. Right: Illustration of how users mark defects. The yellow circles indicate defects detected by the AI system.

Deployment and Refinements

The recycling demonstrator has been showcased at several public events, most notably at Science Night Aachen, with over 100 users participating. Although no formal usability study was conducted, observational feedback revealed that the default pinching gesture came with Unity XR Interaction Toolkit (XRIT) for object manipulation was found to be unintuitive for many users. In response, we replaced XRIT with the AutoHand package, enabling users to grasp objects using their whole hand, which proved to be more natural and accessible.

Another notable observation was that users tended to ignore the on-screen instructions and instead sought direct interaction with event staff. To encourage more independent engagement, we introduced a robot character in both demonstrators, serving as an in-experience guide and narrator.

The wing inspection demonstrator has also been publicly exhibited, including Hannover Messe 2025, with over 200 users participating. While the concept was generally well understood, user performance was paradoxically better without AI assistance. At the time of drafting this paper, out of 281 users who fully completed the demonstration, only 132 exhibited improved performance, which was assessed based on the number of wings inspected, errors identified, and errors overlooked. This may have been due to issues with the AI's visualization or users spending additional time with the AI-enabled version in an attempt to locate all errors. Additionally, some users thought that no further input was necessary after the AI's detection. To address this, we simplified the scenario to present only false positives, ensuring that our intended message about the role of AI assistance was communicated more accurately.

Discussion

Through our XR demonstrators, we are exploring how immersive experiences can help users understand the potential and limitations of AI-assisted industrial tasks. Early results suggest that XR can

effectively convey complex topics, such as human-AI collaboration, to a broad audience. However, it is essential to address potential user distrust toward AI, especially concerns about the job displacement. From an exhibitor perspective, it is also important to ensure that the demonstrators are independently usable and appropriately balanced in terms of content and duration for large-scale exhibitions.

With these goals in mind, we deliberately selected our use cases: the recycling scenario was chosen for its accessibility and ease of understanding for the general public, even though it is not part of the project's real use cases. In contrast, the wing inspection demo was selected for its direct relevance to the project, allowing us to convey more specific and technical content beyond general ergonomic considerations.

Storytelling presented a significant challenge, as we needed to create experiences that were accessible and engaging for both XR novices and experienced users, ensuring newcomers were not overwhelmed while keeping seasoned users interested. Additionally, some users assumed they were interacting with a fully developed application for real-world use, rather than a demonstrator intended to communicate project concepts. This misunderstanding highlighted the importance of clearly framing the XR experience as a demonstration rather than an operational tool. Balancing these aspects was particularly important when integrating gamification elements, since features like high-score rankings could otherwise favor experienced users. We also observed that user performance often improved in the second round with AI assistance, likely due to increased familiarity, though in some cases, monotony or inattention might have led to lower scores. Coincidentally, this monotony helps to illustrate the value of AI in automating repetitive tasks and freeing users to focus on more meaningful work. In future work, we will conduct structured user studies and surveys to evaluate this approach further and gain deeper scientific insights.

References

- Atiker, B. 2023. The conceptual design framework for XR marketing. *Communication Design and Branding*, Springer Series in Design and Innovation, 317–341 https://doi.org/10.1007/978-3-031-35385-7_19
- Oppermann, L., Riedlinger, U., Schmitz, M., Uzun, Y., Brosda, C., Syrek, C. & Fühles-Ubach, S. 2024. Enhancing customer shopping experience through AR mini-games. *Springer Proceedings in Business and Economics XR and Metaverse*, 16–28. https://doi.org/10.1007/978-3-031-50559-1_2

SurgiScope: An Immersive and Interactive Platform for Surgical Training

Tatiana Gerth¹, Roni Lekar¹, Sergey Prokudin^{2,3}, Matthias Seibold³,
Armando Hoch³, Siyu Tang², Philipp Fürnstahl³, and Helmut Grabner¹

¹*School of Engineering, Zurich University of Applied Sciences, Switzerland*

²*Vision and Learning Group, ETH Zurich, Switzerland*

³*Research in Orthopedic Computer Science, Balgrist University Hospital, Switzerland*

Corresponding author: tatiana.gerth@zhaw.ch

Keywords: 3D Reconstruction, Immersive Training, Education, Surgery, Total Hip Arthroplasty

Introduction and Motivation

Watching live surgery is rarely feasible for most trainees, as it requires coordination, scheduling, and access to operating rooms, making consistent exposure to real procedures challenging. As a result, surgical education still relies heavily on passive learning materials such as textbooks and intraoperative videos (Orthobullets, 2025; Youssef et al., 2023). While these resources convey procedural knowledge, they lack interactivity, spatial depth, and learner control, factors that have been shown to improve learning efficiency (Talebi et al., 2022).

Extended reality (XR) technologies are increasingly being explored in surgical education (Silva et al., 2017, Farshad-Amacker et al., 2023, Wu et al., 2025), including applications in Total Hip Arthroplasty (THA). These approaches show promise for improving procedural accuracy and supporting skills assessment (Kenanidis et al., 2024, Logishetty et al., 2020). However, the need for specialized hardware often limits their accessibility and widespread adoption.

Recent advances in 3D reconstruction, particularly methods like 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023), now enable photorealistic, spatially accurate models that can be explored interactively on standard devices. This creates new opportunities for immersive and accessible learning experiences without the barriers of dedicated XR setups.

SurgiScope was developed as an interactive platform that combines video instruction with photorealistic 3D reconstructions of the operating field using 3DGS. Those 3D scenes can be explored from various angles, allowing users to actively engage with the content and test their understanding through interactive tasks like placing tools or implants in the correct position. While the current version focuses on THA, the approach is very adaptable and can be easily extended to other surgical domains, including non-orthopedic procedures.

The platform was shaped through interviews with senior surgeons, whose insights highlighted the strong educational value of high-fidelity 3D reconstructions. Furthermore, studies have shown that medical students and residents benefit significantly more from these 3D reconstructions than from traditional video-based instruction (Lekar et al., 2025).

SurgiScope Platform

SurgiScope is a web- and desktop-based platform that allows learners to explore key steps of THA through interactive 3D scenes of the surgical site. Users can freely navigate the wound region — zooming, rotating, and moving the camera — and toggle layers such as bones and implants to improve spatial understanding. These scenes are generated using high-fidelity 3DGS, resulting in realistic, immersive visualizations.

Beyond passive exploration, SurgiScope includes interactive quizzes and evaluation tasks that support deeper engagement and knowledge retention.

User Interface and Interaction

The application begins with a video of the surgery. A timeline and playback controls allow users to pause or jump to specific points. Key steps are marked on the timeline and listed in a menu in the top right corner. (Figure 20) Each step can be viewed in three modes: **Video**, **Exploration**, and **Evaluation**.

In **Exploration Mode**, users view a 3D reconstruction of the surgical scene at that step. They can explore the scene from multiple angles and toggle anatomical and procedural layers (Figure 21). Senior surgeons particularly appreciated the layer system during the interview sessions, noting its novelty and educational value.

In **Evaluation Mode**, users actively apply their knowledge by positioning surgical tools or implants within the 3D scene (Figure 22, left). Using familiar gizmo-based controls, commonly seen in professional 3D software, they can precisely move and rotate objects, with optional green highlights indicating the correct placement as a hint. What sets this mode apart is its seamless integration within the same environment where learning occurs, allowing users to switch instantly between exploration and evaluation. This direct transition from passive learning to active skill application is a novel feature in surgical education and offers a more cohesive, interactive experience for medical students and residents. Currently, Evaluation Mode is available for three key surgical steps: Femoral Neck Osteotomy, Acetabular Cup Liner Insertion, and Femoral Stem Insertion. Those were defined by a senior surgeon.

Additionally, a quiz module is available (Figure 22, right), showcasing the potential for knowledge checks and theoretical reinforcement throughout the procedure.

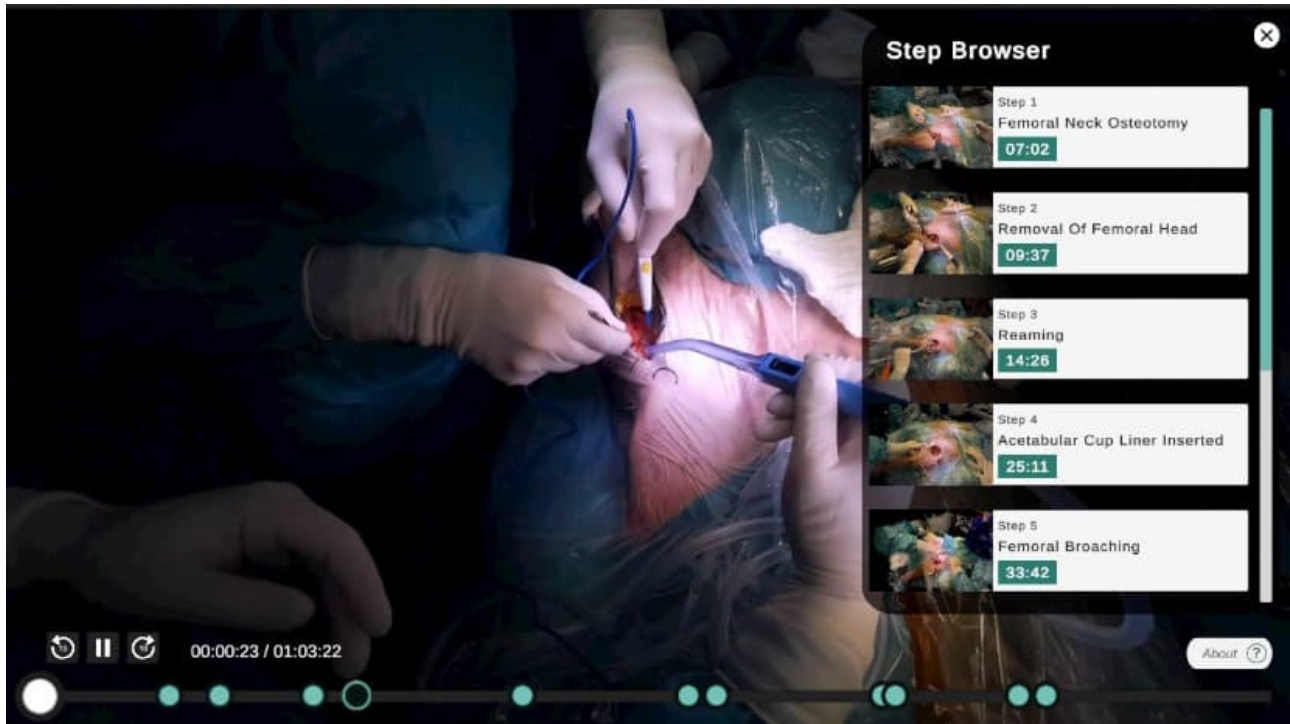


Figure 20. SurgiScope video interface with integrated Step Browser. Key surgical steps are listed with thumbnails and timestamps for quick access. The green buttons in the video timeline correspond to these key steps, enabling seamless navigation between video and interactive content. The black button with a green outline opens the quiz module.

Learning and Assessment Features

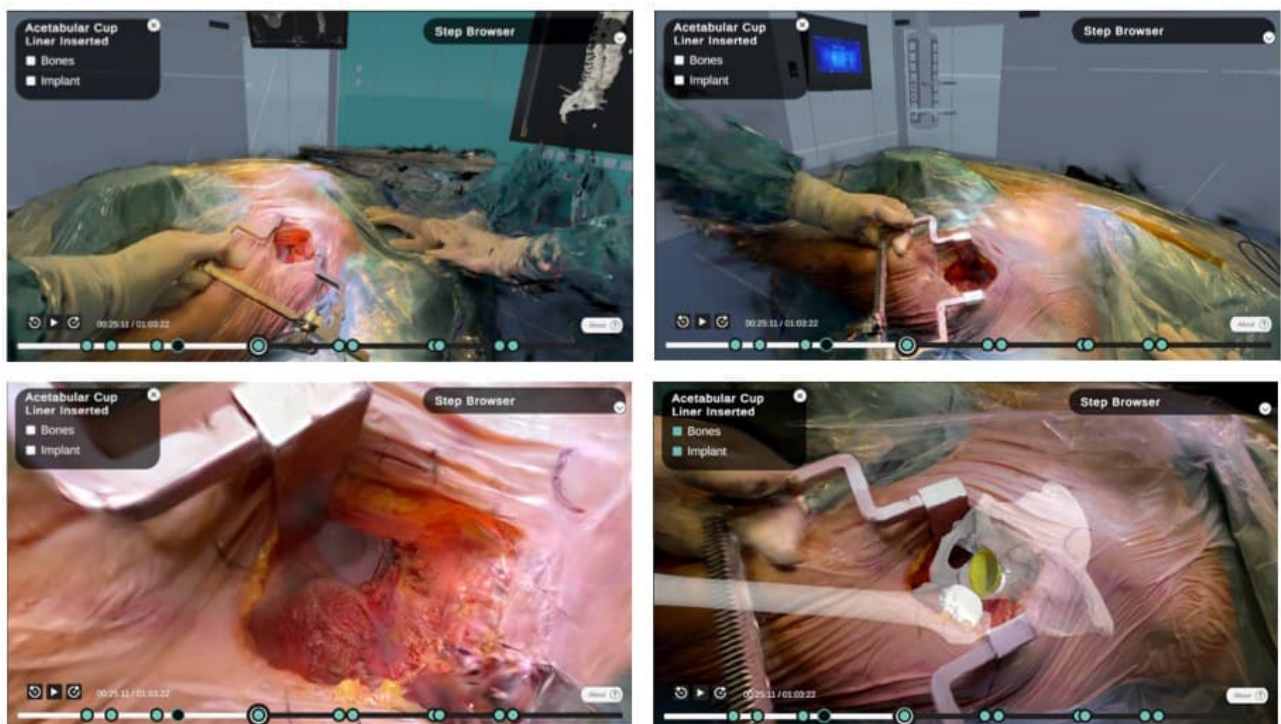


Figure 21. Same 3D reconstructed step from different viewpoints, including anatomical layers on bottom right.

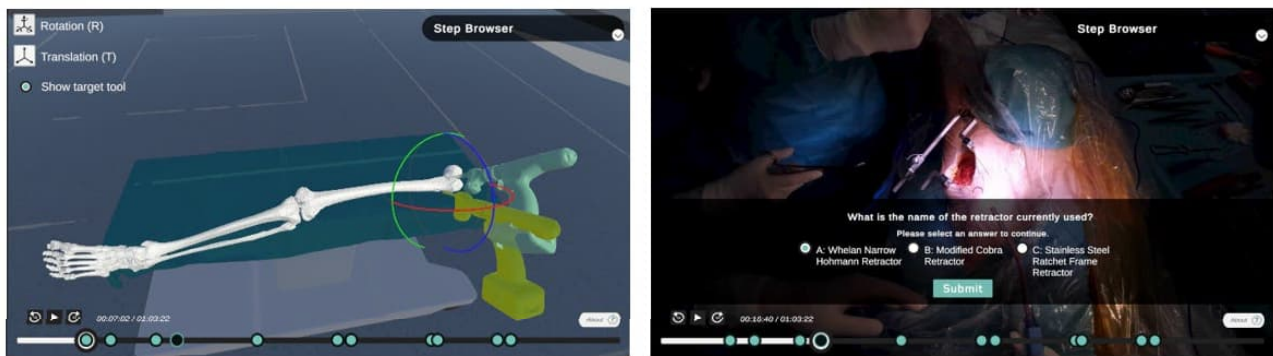


Figure 22. Left: Evaluation mode showing rotation gizmos and target saw. Right: Quiz interface.

Versions and Accessibility

The platform is built using the Unity game engine allowing deployment across desktop and web environments. A web demo version as well as the full version for Windows, macOS, and Linux are available at: <https://panteroni.github.io/3dsurgerytraining>.

Conclusion and Future Work

SurgiScope addresses current gaps in surgical education by offering a more interactive, learner-driven experience. It allows users to observe procedures in detail, engage with spatially rich reconstructions, and test their understanding in a low-pressure environment. Feedback from experienced surgeons as well as medical students and residents confirm the educational and engaging value of this approach. In particular, medical students emphasized the value of the layer system in improving their spatial understanding of anatomical structures.

Future developments will expand evaluation features across all surgical steps and integrate additional video perspectives. Planned UI improvements will better distinguish between video, exploration, and evaluation modes. Additional quizzes and assessment tools are also underway. Moreover, SurgiScope will extend beyond THA to cover other surgical disciplines, progressively building a comprehensive library of interactive surgical procedures.

Acknowledgements

This research is funded by the Innosuisse Flagship project PROFICIENCY No. PFFS-21-19. We thank the senior surgeons whose valuable input helped shape this platform.

References

- Farshad-Amacker, N. A., Kubik-Huch, R. A., Kolling, C., Leo, C. & Goldhahn, J. 2023. Learning how to perform ultrasound-guided interventions with and without augmented reality visualization: a randomized study. *European radiology*, 33(4), 2927–2934.
- Kenanidis, E., Boutos, P., Voulgaris, G., Zgouridou, A., Gkoura, E., Gamie, Z., Papagiannakis, G. & Tsiroidis, E. 2024. Effectiveness of virtual reality compared to video training on acetabular cup and femoral stem implantation accuracy in total hip arthroplasty among medical students: a randomised controlled trial. *Int. orthopaedics*, 48(3), 625–633.

- Kerbl, B., Kopanas, G., Leimkühler, T. & Drettakis, G. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4).
- Lekar, R., Gerth, T., Prokudin, S., Seibold, M., Bürgin, R., Vella, B., Hoch, A., Tang, S., Fürnstahl, P. & Grabner, H. 2025. Enhancing Orthopedic Surgical Training With Interactive Photorealistic 3D Visualization.
- Logishetty, K., Gofton, W. T., Rudran, B., Beulé, P. E. & Cobb, J. P. 2020. Fully Immersive Virtual Reality for Total Hip Arthroplasty: Objective Measurement of Skills and Transfer of Visuospatial Performance After a Competency-Based Simulation Curriculum. *J. bone joint surgery. American volume*, 102(6). <https://doi.org/10.2106/JBJS.19.00629>
- Orthobullets. 2025. An educational resource for orthopedic surgeons. <https://www.orthobullets.com>
- Silva, I., Klein, G. & Brandão, D. 2017. Segmented and Detailed Visualization of Anatomical Structures based on Augmented Reality for Health Education and Knowledge Discovery. *Advances in Science, Technology and Engineering Systems Journal*, 2(3), 469-478.
- Talebi, M. & Safaei, M. 2022. Measuring User Engagement in Educational Platforms: A Study on Interactive vs. Passive Learning. *International Journal of Advanced Human Computer Interaction*, 1(1), 53-62.
- Wu, L., Seibold, M., Cavalcanti, N. A., Hein, J., Gerth, T., Lekar, R., Hoch, A., ... & Fürnstahl, P. 2025. A novel augmented reality-based simulator for enhancing orthopedic surgical training. *Computers in biology and medicine*, 185, 109536. <https://doi.org/10.1016/j.combiomed.2024.109536>
- Youssef, S. C., Aydin, A., Canning, A., Khan, N., Ahmed, K. & Dasgupta, P. 2023. Learning Surgical Skills Through Video-Based Education: A Systematic Review. *Surgical innovation*, 30(2), 220-238. <https://doi.org/10.1177/1553350622112014>

Interactive 3D Visualization and Configuration of Solar Panel Layouts in XR

Cloe Hüsser, Luca Fluri, and Hilko Cords

*Institute of Interactive Technologies, School of Computer Science,
FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland*

Corresponding author: hilko.cords@fhnw.ch

Keywords: XR, Interaction, Collaboration, Solar Energy

Introduction

The transition to renewable energy demands accessible planning tools that effectively communicate solar panel designs to diverse stakeholders. Conventional approaches rely on static 2D renderings or limited 3D visualizations, which fail to convey critical environmental factors such as shading patterns, sun paths, building surroundings and spatial relationships, creating barriers to effective decision-making and collaborative design.

Recent XR technologies offer promising solutions for environmental visualization and collaborative design in the field of solar energy. AR has already been successfully applied to evaluate solar energy setups on site and enables the interactive analysis of photovoltaic setups (Meireles et al., 2023) and the detection of damaged PV modules (Oulefki et al., 2024). Complementary, VR applications demonstrate value in immersive environmental representations, enabling visualization of complex factors like dust accumulation effects (Ni et al., 2017) and supporting collaborative design through multi-user virtual environments (Ha et al., 2022).

However, there is still a lack of dual-platform AR/VR tools specifically designed for solar panel layout visualization that integrate seamlessly with existing design workflows. This work addresses this limitation. We present a comprehensive and fast system that bridges collaborative, digital planning in VR with immersive, collaborative experiences in AR.

System Architecture

Our solution features a unified architecture that supports both AR and VR applications through a shared codebase, significantly reducing development time and ensuring consistent functionality across platforms. Built on Unity with strong cross-platform support, our architecture provides modules for geometric modelling, visualization, interaction logic, project management, networking, and sun simulation.

Thereby, our system interfaces with an existing web-based 3D solar design tool, accessing 3D setups via RESTful APIs. The fetched scene data includes geospatial coordinates, mesh data for buildings and panels, and environmental elements for shadow simulation. We use lightweight deserialization of the gathered data to recreate 3D models, which are then rendered using custom mesh builders and lighting logic.

Platform-specific AR and VR modules interface with our shared core code basis to enable different interaction paradigms and rendering optimizations. This approach ensures scalability, maintainability, and seamless feature consistency across different platforms.

Augmented Reality

Our AR component transforms mobile devices into spatial viewers for interactive solar layout exploration. Built using Unity 2022.3 LTS and Vuforia Engine, the system provides markerless tracking, geographic data integration, and real-time sun simulation.

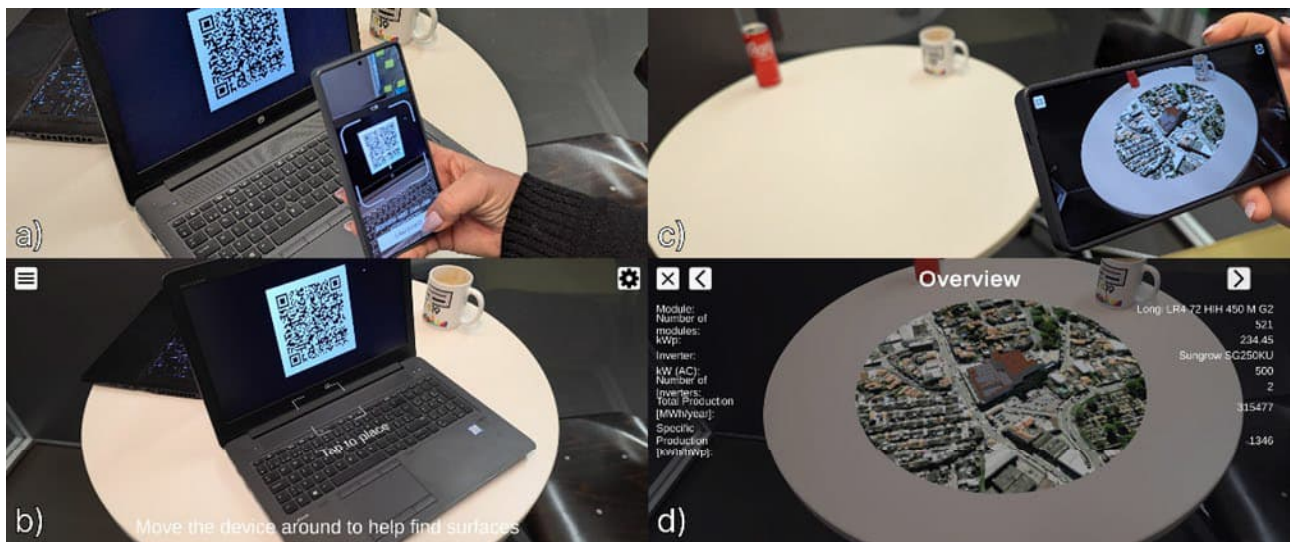


Figure 23 Our AR workflow: a) Scan the project QR code, b) detect the projection surface including visual guidance, c) augment the 3D scene on the surface, and d) view the details page within the app.

The user workflow begins with QR code scanning from the web-based design platform. These QR codes contain AR flags and backend URLs that trigger project data retrieval and spatial calibration. Afterwards, visual placement guides assist users in anchoring 3D scenes in their physical environment (Figure 23).

The rendered AR scene includes all buildings and panel placements from the web platform, enhanced with terrain data dynamically fetched from ArcGIS services. This provides crucial spatial context through geographic base layers beneath installation sites. Users can explore scenes through device movement and gesture interactions, including pinch-to-zoom, drag, and tap controls.

Performance optimization includes static batching and custom mobile-optimized shaders that preserve visual quality. Our application is publicly available on Google Play Store's open-testing channel and demonstrates consistent performance across midrange Android devices.

Virtual Reality

Our VR component complements the AR application by providing immersive collaborative environments for spatial assessment. Designed for standalone VR headsets (Meta Quest series), the system shares core data pipelines and mesh generation features with the AR counterpart, ensuring consistency across platforms. Additionally, our VR interaction system utilizes also Unity's XR Interaction Toolkit and OpenXR runtime for broad device compatibility.

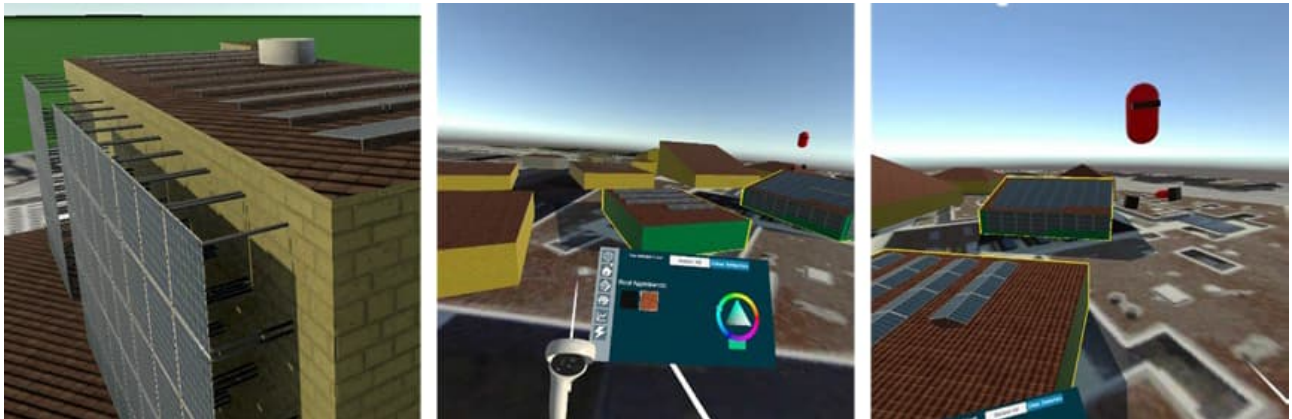


Figure 24. Example screenshots from an active VR session. The VR controller menu and ray-based building selection is shown, as well as the visualization of another user inside the collaborative VR space (red capsule). Notably, the first image showcases visualized

In a session, users manipulate environments through standard VR controllers, with the left controller providing hand-attached UI menus for material and panel customization. Ray-cast pointer interactions enable configuration of roof materials, facade options, and solar panel styles. Navigation employs thumbstick-controlled teleportation to prevent motion sickness while enabling exploration of large-scale scenes.

To achieve real-time rendering of thousands of poles of solar panels with minimal overhead, we apply custom GPU instancing shaders to visualize solar panel support structures (Figure 24). This significantly improves spatial understanding and shadow estimation capabilities while maintaining high frame rates.

Our collaborative functionalities are based on Photon PUN2 for networking facilities. A companion mobile application manages session creation and user authentication, avoiding the challenges of text input in VR environments. Once connected, users spawn into a shared environment with synchronized avatars and full visibility of participant locations and interactions (Figure 24). All customization events synchronize via Photon Remote Procedural Calls (RPCs), ensuring coherent multi-user experiences.

Our VR system achieves stable 90-120fps frame rates on Meta Quest devices through efficient mesh instancing, occlusion culling, and Unity's URP configuration with single-pass stereo rendering.

Discussion & Conclusion

Our system demonstrates the technical feasibility and practical utility of immersive XR tools for solar panel layout planning and communication. Thereby, our dual-platform approach successfully addresses different stakeholder needs: AR provides accessible, context-aware visualization for live assessment, while VR enables detailed collaborative analysis in controlled environments.

Early user feedback from internal testing sessions with researchers, developers, and industry stakeholders has been positive. Participants appreciate the AR experience's accessibility and visual clarity, particularly the QR onboarding process and energy yield visualizations. The ability to explore virtual layouts and evaluate shadow behavior over time and seasons was highlighted as particularly valuable compared to traditional 2D planning tools.

In VR, users value the interactive scaling and teleportation features that enable rapid perspective shifts and detailed configuration inspection. The controller-attached UI is considered intuitive, with scene scaling emerging as the most frequently used feature. Collaborative sessions enhance spatial awareness through visible participant positions, proving especially useful for sales meetings and remote consultations.

Our unified codebase approach proved beneficial, ensuring that improvements in scene generation, material loading, and simulation logic benefited both platforms simultaneously. This strategy reduced development and maintenance effort and improved consistency, positioning the system for scalable deployment.

We presented a scalable, immersive platform for visualizing and evaluating solar panel installations across AR and VR modalities. By bridging digital planning and real-world implementation, our system enhances communication and understanding while laying the groundwork for future innovation in renewable energy planning through immersive XR technology.

References

- Alqallaf, N., & Ghannam, R. 2024 Immersive Learning in Photovoltaic Energy Education: A Comprehensive Review of Virtual Reality Applications, *Solar*, 4(1), pp. 136–161. <https://doi.org/10.3390/solar4010006>
- Ha, E., Byeon, G., & Yu, S. 2022. Full-Body Motion Capture-Based Virtual Reality Multi-Remote Collaboration System, *Applied Sciences*, 12(12), p. 5862. <https://doi.org/10.3390/app12125862>
- Mehta, K., Sakhapara, A., Pawade, D., & Surve, V. 2020. Novel Approach of Computing Optimal Placement of Solar Panel Using Augmented Reality, in Vasudevan, H., Michalas, A., Shekokar, N. and Narvekar, M. (eds.) *Advanced Computing Technologies and Applications*. Singapore: Springer, pp. 533–542.
- Meireles, C., Carmo, M.B., Cláudio, A.P., Ferreira, A., Afonso, A.P., Redweik, P., Catita, C., Brito, M.C., & Soares, D. 2023. Mobile Augmented Reality for Analysis of Solar Radiation on Facades, in *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. Lisbon: SCITEPRESS, pp. 131–142.
- Ni, J., Yang, M., & Jiang, Y. 2017. Virtual Reality Simulation of Dust Accumulation on the Surface of Solar Panel, in *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*. IEEE, pp. 425–430.
- Oulefki, A., Himeur, Y., Trongtirakul, T., Amara, K., Agaian, S., Benbelkacem, S., ..., & Mansoor, W. 2024. Detection and analysis of deteriorated areas in solar PV modules using unsupervised sensing algorithms and 3D augmented reality, *Heliyon*, 10(6).
- Rzeszewski, M., & Orylski, M. 2021. Usability of WebXR Visualizations in Urban Planning, *ISPRS International Journal of Geo-Information*, 10(11), p. 721. <https://doi.org/10.3390/ijgi10110721>

Immersive Exploration and Data Acquisition: XR Virtual Field Trips for 3D Photogrammetry and Remote Sensing (SURE-VFT)

Petros Patias¹, Charalampos Georgiadis¹, Georgios Mallinis¹,
Anastasios Stamnas¹, Vasiliki Fragkoulidou¹, Themistoklis Roustanis²,
George Patias², Alexandra Makridou², Alain-Pierre Bourdot^{3,4},
and Maurizio Sacco^{3,5}

¹*School of Rural and Surveying Engineering, Aristotle University of Thessaloniki, Greece*

²*KIKLO - Geospatial Information Technologies, Greece*

³*EuroXR - European Association for eXtended Reality, Belgium*

⁴*LISN/CNRS, University Paris-Saclay, France*

⁵*CNR-STIIMA -National Research Council of Italy, Italy*

Corresponding author: patias@auth.gr

Keywords: Extended Reality, Education, Virtual Field Trips, 3D Photogrammetry, Remote Sensing

Abstract

The project Immersive Exploration and Data Acquisition: XR Virtual Field Trips for 3D Photogrammetry and Remote Sensing (SURE-VFT), is an extended reality (XR)-based educational initiative designed to enhance training in the field of RS and photogrammetry through immersive virtual field trip (VFT) experiences. The project develops and pilots three VFT focused on close-range photogrammetry, land use/land cover (LULC) mapping, and in-situ spectral signature collection. By using VR environments, students engage in realistic data acquisition tasks, such as capturing images for 3D modeling, interpreting RS imagery, and identifying spectral responses of land cover types, thereby bridging theoretical concepts with hands-on practice. SURE-VFT integrates feedback mechanisms, user-guided exploration, and exportable datasets to simulate professional workflows and align with industry standards. Early results suggest that the platform increases learner engagement, supports skill development, and offers a scalable model for XR implementation in geospatial education.

Introduction

Virtual technology has been linked in recent research to improvements in students' academic performance and motivation, as well as in their psychomotor and cognitive skill sets, social and collaborative skills, and more (Martín-Gutiérrez et al., 2017). The benefits of employing virtual technology can be divided into a number of categories. Students' attention and engagement are raised by providing immersive experiences that allow them to learn 3D representations as active participants. A learning method made possible by virtual technology enables students to interact freely with one

another, promoting experimentation, investigation, and feedback that enhances the assimilation of knowledge. Additionally, recent advancements in technology have reduced the cost of VR, AR, and MR while increasing their availability and accessibility on desktop and mobile devices (Pedram et al., 2024).

The project "Immersive Exploration and Data Acquisition: XR Virtual Field Trips for 3D Photogrammetry and Remote Sensing (SURE-VFT)", is a pilot funded by XR4ED under Horizon Europe. It aims to deliver a robust educational solution that enhances learning outcomes for Surveying Engineering students, establishes a scalable model for replication in extended reality, and fosters commercial acceptance of the XR4ED platform. The integration of industry needs, validation through pilot launches in XR environments, and focus on accessibility, user control, and learning effectiveness will ensure the program's relevance and effectiveness across diverse educational contexts.

Photogrammetry Virtual Field Trip

This pilot scenario helps students to optimize data acquisition for close range photogrammetry applications both for movable and immovable objects. The students are called to create 3D models of buildings or small artefacts. At the first scenario (Figure 25) they visit the area where the building is situated (the Aristotle University of Thessaloniki observatory), and they are asked to collect all the data needed for successful processing (the photogrammetric documentation of the building). They are led to predefined points to capture images and select control and check points on the building surface. At the end of the trip the collected data (5 ground images and 5 Ground Control Points (GCPs) can be exported and used to produce the building's 3D model.



Figure 25. The 3D representation of the observatory, the student (as an avatar), the predefined points (big 3D red pointers) to capture images and the Ground Control Points (GCPs) on the surface of the observatory.

LULC Mapping for Remote Sensing

The LULC VFT aims to train students to landmark identification in the field, as well as image interpretation. The experience is set in a virtual scene where a large satellite map from Sentinel 2 satellite is displayed on another wall with four field photos, taken from the cardinal directions (North, South, East, and West), to simulate real-world ground-truth conditions (Figure 26). The field images are from the Land Use/Cover Area Frame Survey (LUCAS). A dedicated wall showcases high-resolution satellite

imagery related to the field photo in zoom, accompanied by an orientation board that allows users to track their progress throughout the activity and a list of classification types is placed on another wall with a message board to inform users of feedback.

The user navigates into the scene by walking toward the four 3D red directional markers (North, South, East, West). Each field photo must be correctly aligned with the correct red orientation pin. When correctly matched, the field image is placed next to the corresponding pin, allowing progression to the next marker.

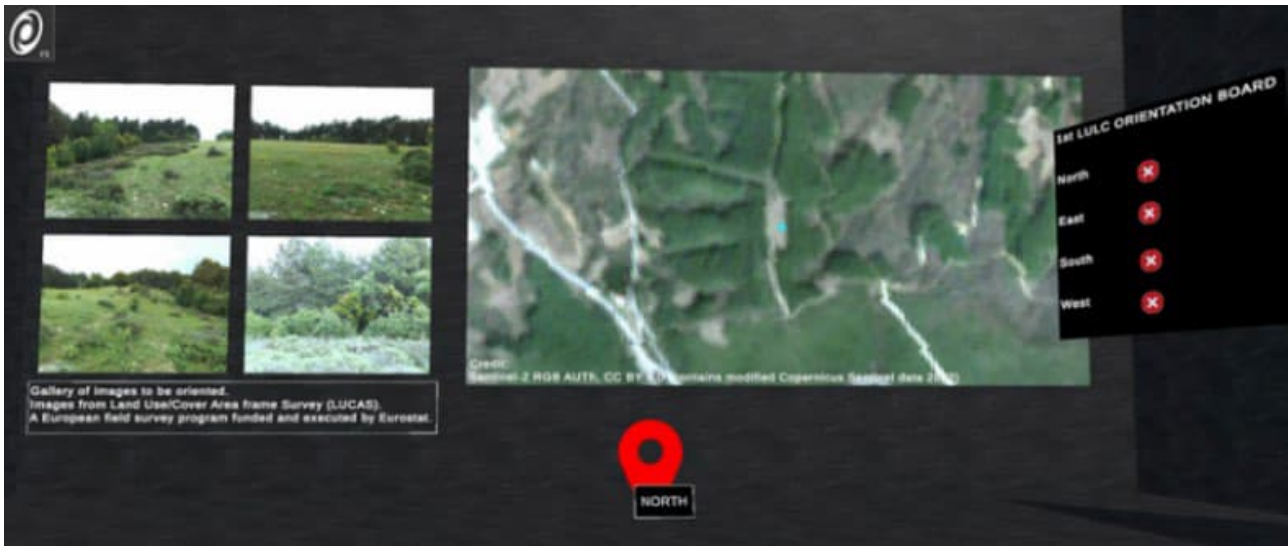


Figure 26. Ground level photos and Sentinel-2 image.

Spectroradiometer for Spectral Signature Collection

The Spectroradiometer Usage for Spectral Signature VFT aims to introduce students to the concept and application of in-situ spectral signatures collection. This hands-on training develops skills in remote sensing interpretation, enhances understanding of how different surfaces reflect light across the spectrum, and supports accurate land cover classification in environmental and geospatial studies. The VR experience is set in a virtual scene where a large satellite map from Sentinel 2 is displayed on a dedicated wall with two screenshots of a spectroradiometer, taken from the yellow point to simulate real-world ground-truth conditions (**Figure 27**). The user navigates into the scene, examines the satellite photo and the screenshots of spectral usage and corresponds to which spectral screenshot is correctly matched. The correct answer presents a new button to continue to the next POI (Point of Interest). When the set of 6 POIs is answered correctly, the user finalizes the session.

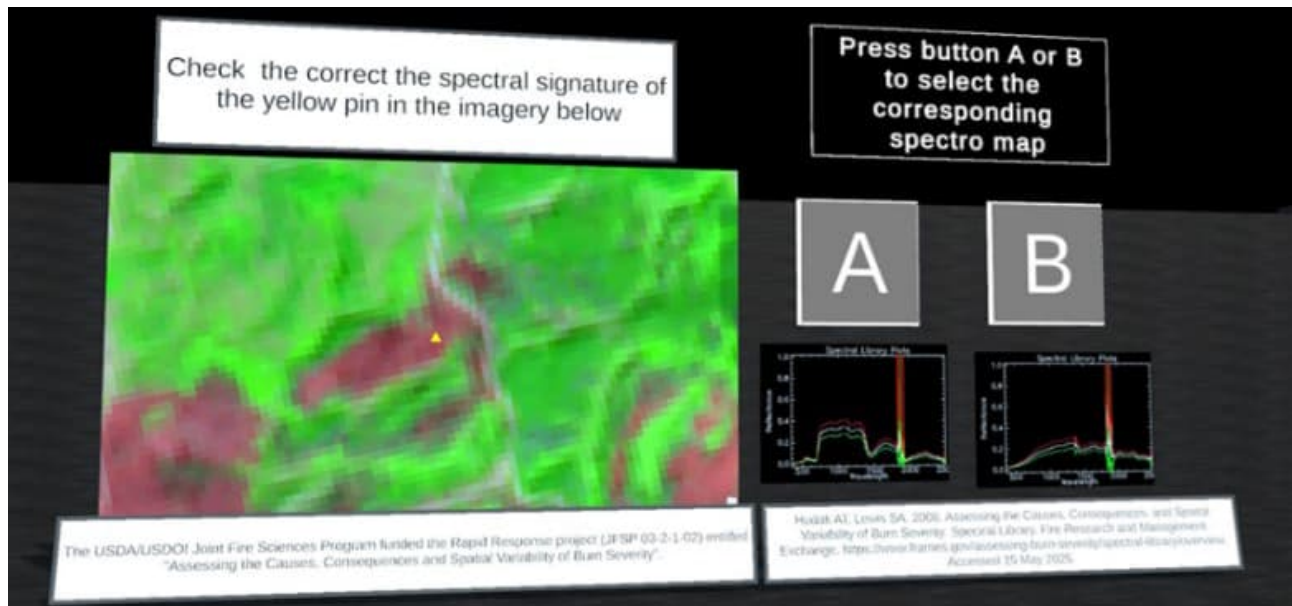


Figure 27. Satellite Imagery Wall and spectroradiometer screens (Spectroradiometer Usage for Spectral Signature Lesson).

Conclusions

The SURE-VFT project is based on the idea that extended reality (XR) can improve the learning process in Surveying Engineering. Compared to conventional teaching techniques, this XR-based approach can lead to higher motivation, engagement, and long lasting knowledge. Key design considerations in the SURE-VFT project include the development of user-friendly interfaces, modular learning units according to specific training subject need, realistic but simple 3D modelling, and the existence of feedback mechanisms. Over the following semesters the XR based training approach will be evaluated against lecturing approaches involving only traditional multimedia material (i.e. slides, videos).

Additionally, the XR-based Surveying Engineering training integrates usability, structured processes, and relevant data export capabilities. To develop Virtual Field Trips that are efficient, technically accessible, and scalable in a variety of institutional and commercial contexts, the SURE-VFT project intends to combine the best practices found existing research advances with the needs of users, especially from the undergraduate students. The XR4ED platform will advance and XR techniques for Surveying Engineering education will be widely used across Europe and beyond.

Acknowledgements

The SURE-VFT project has indirectly received funding from the European Union's Horizon Europe research and innovation action programme, via the XR4ED – Open Call issued and executed under the XR4ED project (Grant Agreement no. 101093159).

References

- Martín-Gutiérrez, Jorge, et al. "Virtual technologies trends in education." *Eurasia journal of mathematics, science and technology education* 13.2, 469–86. <https://doi.org/10.12973/eurasia.2017.00626a>.
- Pedram, S., Kennedy, G., & Sanzone, S.. 2024. Assessing the Validity of VR as a Training Tool for Medical Students. *Virtual Reality* 28 (1): 15. <https://doi.org/10.1007/s10055-023-00912-x>.

XR Helicopter Simulator for LLM-based Virtual Assistant in Maintenance Support

Andrea Alesani, Vladimir Goriachev, Jaakko Karjalainen, and Petri Tikka

VTT Technical Research Centre of Finland

Corresponding author: andrea.alesani@vtt.fi

Keywords: extended reality, wall-display, LLM, maintenance

Introduction

This paper presents the design and implementation of an Extended Reality (XR) simulator for helicopter maintenance training, leveraging the Unity engine and a large-scale “powerwall” visualization setup.

The system simulates a simplified helicopter model in a photorealistic scenario, enabling users to interact with operational and maintenance procedures in a controlled, yet realistic virtual setting. A virtual assistant is embodied as an intelligent avatar and harnesses the power of a large language model (LLM) to allow for verbal conversation. The assistant is designed to guide users through common maintenance workflows by cross-referencing the official PDF manuals and real-time telemetry data streamed directly from the simulated aircraft systems.

User Interface and Virtual Assistant

The simulator is rendered across a three-wall display configuration, forming a 120-degree immersive environment that enhances immersion and realism for users engaged with technical tasks and procedures.

The interface splits the powerwall: two-thirds for helicopter visualization, one-third for the assistant. It features a semi-transparent look with gradient panels, orange for user actions, blue for the assistant. Panels are draggable, linked to toggle buttons by Bezier lines, and can be hidden with a hotkey. UI settings are saved and restored on restart. Help and input panels support multi-line text, and conversation history can be exported. Users can pick different assistants, each visually styled with blue holographic effects.

The AI assistant uses a state machine—idle, listening, thinking, and speaking—with varied animations for each state. It responds to both text and push-to-talk, displays thinking status in conversation history, and speaking can be paused or cancelled. Voice and ambient sound can be muted for flexibility. Including several other “emotional” animations controlled by the LLM assistant could have been beneficial to the overall user experience, as demonstrated in previous research (Helin, et al., 2024),

but our focus on local, less performant LLM models led us to choose a simplified approach for it in order to avoid hallucinations.



Figure 28. Overview of the simulator setup. It includes a wall-sized display, a push-to-talk microphone, and an “IoT box” with 3 rotating dials and a red LED light. Keyboard and mouse are used to customize the UI.



Figure 29. The assistant and the chat panel (on the left). The virtual model features 4 different animated poses and embodies the RAG-enhanced LLM assistant.

The helicopter model features several visual cues that mimic real-life scenarios and indicate different problems that the user is required to address. Our contribution includes a custom-written, simplified

maintenance manual that the assistant can access through *RAGFlow* and that describes both the cues, the problems they indicate and the recommended actions to solve them. The main goal presented to the user is to calibrate the fuel and air flow into the engine mixture, in order to keep the Air-Fuel Ratio (AFR) within the thresholds specified in the documentation.



Figure 30. The visual effects coming from the helicopter model. Black smoke (on the left) indicates a “rich” condition, where too much fuel is present on the mixture, while white/grey smoke indicate a “lean” condition (on the center) with excessive air flow. Different engine sounds are also played for different conditions.

Physical IoT interactions were simulated using rotary encoders connected to a *BeagleBone* board, offering a tangible interface for users to engage with virtual aircraft systems as they would in actual maintenance scenarios. The *BeagleBone* board was then encapsulated in a plastic case. A companion iPad application was also developed to rapidly iterate on the prototype before building the actual physical component.



Figure 31. The input devices for the MQTT broker. The physical “IoT box” includes three rotary-encoded dials to handle fuel flow, air flow, and throttle. A red LED turns on or blinks when AFR goes slightly out of range or it reaches critical levels. The iPad application emulated the box, with the three sliders on the upper part of the UI.

Conversation architecture

At the heart of the backend, our system utilizes *RAGFlow* as the principal retrieval-augmented generation (RAG) tool for extracting relevant information from the manual documentation. The *RAGFlow* server is connected to the *Qwen2.5 32B* model (quantized to 4 bits), which represents the best trade-off between inference speed and performance for our *NVIDIA 4090* GPU. To provide flexibility and comparative insights, users can also select from two additional “assistants” within *RAGFlow*: one running *Llama3.3 70B* and another based on *GPT-4o*. This multi-assistant approach allowed to showcase the

validity of the presented architecture regardless of the current performance bottlenecks with local models when using consumer hardware, while still demonstrating the feasibility of such a system when built locally.

For voice interaction, we incorporated *faster_whisper* to transcribe spoken queries into text rapidly and accurately, while the system responds through synthesized speech generated by *kokoro* for natural text-to-speech (TTS) output.

A “MqttParser” script has been developed with the role of both caching data from our custom-made MQTT broker and preparing it to be fed to the LLM assistant. Three techniques have been tested to feed the data from MQTT into the LLM assistant: (i) a “StaticAnalysis” function that performs some predefined calculations on the data, such as extracting the mean, the variance, and the last available value for each of the monitored features, and then creates as an output a structured report; (ii) a “DynamicAnalysis” function that creates a Markdown-formatted table with the last recorded data and lets the LLM assistant analyze it for itself; (iii) a “DoubleDynamicAnalysis” function that creates the same Markdown table as “DynamicAnalysis”, but then feeds it to a different LLM session to analyze the data and its output is then fed to the actual LLM conversational assistant. In the final version of our system, the StaticAnalysis technique has been used, as it seemed to provide better results with local assistants.

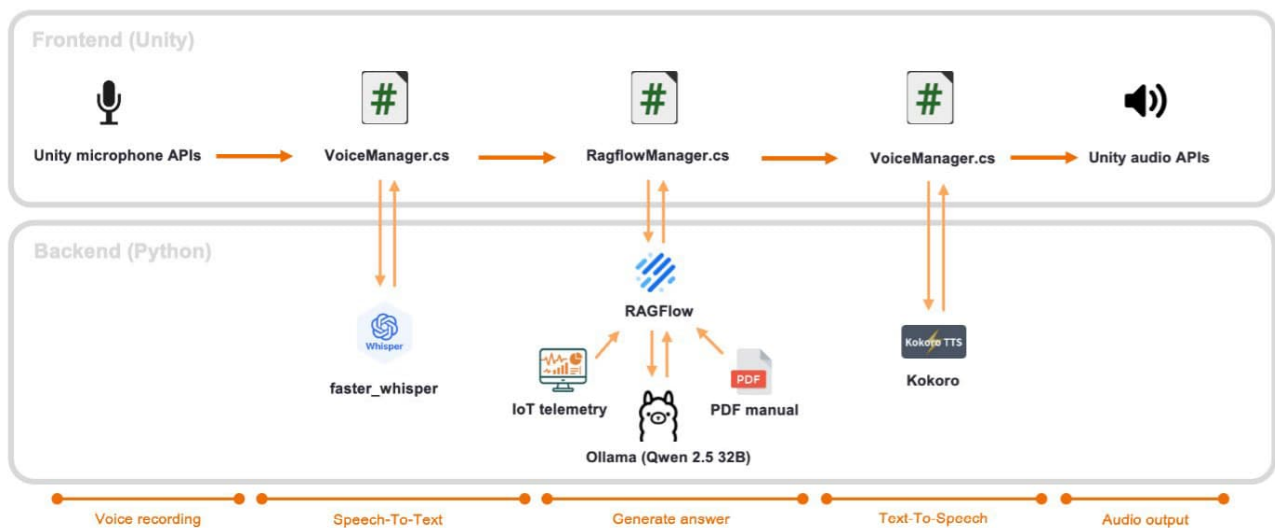


Figure 32. Simplified conversation flow. Text-to-speech (TTS) and speech-to-text (STT) have been modularized so that the application can work even without them. Even though marked as “backend”, the “IoT telemetry” MQTT component resides into the Unity application. RAGFlow represents the core of the system, and it orchestrates information flow between documents, IoT data and LLMs to provide the user with the most relevant answer.

Conclusions

This project showcases how the combination of immersive XR visualization, intelligent language model-based assistance, and real-time system integration can improve maintenance training and operational support—especially in high-stakes domains like aerospace. The assistant interprets natural language queries and live telemetry data, delivering step-by-step guidance, troubleshooting insights, and procedural support within a visually realistic XR environment. Integrated with physical IoT elements and customizable interfaces, it bridges the gap between simulation and real-world maintenance tasks.

Thanks to RAG, the assistant offers accurate, context-aware interactions while remaining flexible enough to support multiple conversational roles. Its embodied presence encourages intuitive engagement, both through voice and text, making it accessible for all kinds of users. Future directions

will focus on defining metrics to evaluate the assistant's performance with different models and/or architectures, porting the assistant to HMDs, and extending the framework to real industrial use cases.

References

- Helin, K., Goriachev, V., Tikka, P., Alesani, A., Karjalainen, J. 2024. LLM-Driven Human-Robot Interaction in eXtended Reality. Proceedings of the 21st EuroXR International Conference, 35-40.
- SYSTRAN. faster_whisper [Software]. Available at <https://github.com/SYSTRAN/faster-whisper>.
- Unity Technologies. Unity (Version 6000.0.38f1) [Software]. Available at <https://unity.com>.
- Infiniflow. RAGFlow (Version 0.19.0) [Software]. Available at <https://github.com/infiniflow/ragflow>.
- Ollama Inc. Ollama (Version 0.69) [Software]. Available at <https://ollama.com>.
- BeagleBoard.org Foundation. BeagleBone Black [Hardware]. Available at <https://www.beagleboard.org/boards/beaglebone-black>.
- Hexgrad. Kokoro-82M: An Open-Weight TTS Model. Available at <https://pypi.org/project/kokoro/>.
- Qwen Team. Qwen2.5: A Party of Foundation Models. Available through Ollama (or <https://qwenlm.github.io/blog/qwen2.5/>).
- Meta Platforms, Inc. Llama 3.3: Multilingual Large Language Model. Available through Ollama (or <https://ai.meta.com/blog/meta-llama-3/>).
- OpenAI. GPT-4o: Omni-Modal AI Model. Available through <https://platform.openai.com/>.

An Extended Reality Interface for Ergonomics Training

Spyridon Nektarios Bolierakis¹, Kostas Naskou^{1*}, Ioannis Karagiannis¹,
Aimilios Seimenis¹, Konstantinos Stathoulas², Aris Dadoukis², Athanasia Tsertou²,
Tina Katika¹, Georgios Tsimiklis¹, Ioannis Karaseitanidis¹, and Angelos Amditis¹

¹ICCS - Institute of Communication and Computer Systems, Greece

²CIBOS, Greece

Corresponding author: kostas.naskou@iccs.gr

Keywords: XR, REBA, ergonomics

Introduction

Workers in automated manufacturing participate in repetitive and physically demanding activities. Appropriate training is essential but working environments are mostly unavailable and often inappropriate for training and evaluation purposes. The aim of this work is to provide the tools to assist both trainers and trainees in the industrial automation domain to assess the actual working experience in a realistic and safe environment. The main objective regarding trainers and educators is to provide a tool that allows them to assess the exposure of workers to ergonomic risk factors, raise awareness, and help minimize them by evaluating various parameters corresponding to both the workflow and the working environment setup. As for the trainees, the tools provide hands-on experience of the required working tasks in a virtual yet realistic and safe environment. In the context of this work, an XR application, called Ergon-XR, is developed for the VIROO platform to allow a trainer to assess a trainee's ergonomic risk, via a Mixed Reality interface, while the trainee is interacting with a virtual scenario. The first implemented scenario involves a conveyor picking line as the virtual work environment case study but the Ergon-XR app can be easily applied to any other simulated working environment. The concept includes a trainee interacting with the virtual scene through their Head-Mounted Display (HMD) while being monitored by an external off-the-shelf web camera. The trainer, wearing their own HMD, places the camera at different points of view, as required, so that the trainee's physical ergonomics, i.e. posture, can be assessed not only online, with near-real time results to provide the necessary feedback for corrective actions, but also offline with all the gathered data being logged for further analysis. At this point in the development phase of the app, Google's MediaPipe pose landmark detector is being assessed to ascertain its performance for the Rapid Entire Body Assessment (REBA) score estimation. Subsequently, the Ergon-XR app will undergo a formal validation process by testers to ascertain it is fit for purpose.

Literature Review

This work addresses two critical fundamental constraints of ergonomics risk assessment methods, namely accessibility and usability (Lowe, 2019), and provides a portable solution deployed on XR-HMD devices which are compatible with the VIRTUALWARE VIROO platform and its requirements and constraints. Several methods, and corresponding tools, have been developed to provide practical means for evaluating physical ergonomics risks (e.g. RULA, REBA, EAWS, OWAS, OCRA). These tools are aids to ergonomics experts and require the collection of information through systematic observations and questionnaires or direct measurements. At the early stages of the project a similar study was published by Chu, (2025) which indicates a growing interest in this subject. Both Manghisi (2022), who offer a VR experience to a trainee tracked by a Microsoft Kinect sensor, and Chu (2025) also use the Rapid Upper Limb Assessment (RULA) tools, which is a subset of REBA (Hignett, 2000), but the latter also use the Microsoft HoloLens 2 device sensors for the trainer to track the trainee. There are approaches in the literature that use 3D landmarks to train a deep learning model to output a REBA score directly from the pose landmark coordinates (Hossain, 2023) but this approach is limited by the availability of data representative of all the possible body poses and offloads straightforward geometric computations (from pose-landmark coordinates to angles at joints to score) to a black box algorithm that lacks interpretability (Yang, 2024). The characteristics of the conveyor, e.g. height, in the virtual scene, which is being used as a case study, were selected according to findings in Emmatty (2021) and Aberger (2025).

Design

There are three user roles involved in each implemented case study, namely the Content Creator preparing the training scenario offline and, during runtime, the Trainer and the Trainee. In practice the Content Creator and the Trainer can be the same person. The Trainer is an educator (college teacher or university professor) or a supervisor at a workplace who needs a digital tool to assess a Trainee's ergonomics risk while they are performing a task within a virtual industrial automation environment.

High-level Architecture

The system comprises two modules, the VIROO Single Player app, which drives each HMD, and the Server app; these two modules are connected by bidirectional communication on a local network with an API that is responsible for handling data exchange of single actions and (de-)activating the UDP connection sockets. The UDP connection is responsible for the live communication and data exchange between the modules prioritizing low latency at the expense of the possibility to lose packets which is not crucial for this application (Figure 33). Each one of the two HMDs, the Trainer's and the Trainee's, are connected to a different PC or laptop each running a session of the Ergon-XR app on the VIROO Single Player app utilizing VIROO platform's networking capabilities to provide a shared experience of the running virtual scenario. A shared experience between trainer and trainee is crucial to achieve both proper monitoring of the work task execution and remote evaluation if necessary, broadening the application's range of use cases. There is a built-in option to select the PC (either the Trainer's or Trainee's) that will provide the external camera feed. By default, the Trainer's PC will handle the external camera feed, but if required the camera feed could be provided by the Trainee's PC. The camera feed is shared among both parties through VIROO networking but only the Trainer's PC is responsible for sending and receiving data from the server. To achieve real-time results and to minimize latency, the setup uses the Edge Computing concept since the server application must be deployed to either the trainer's PC or to a PC in the same local network. To ensure the minimum latency possible, only the

Trainee's PC is responsible for streaming the virtual scenario. Finally, all output data generated are saved and exported for review and post processing.

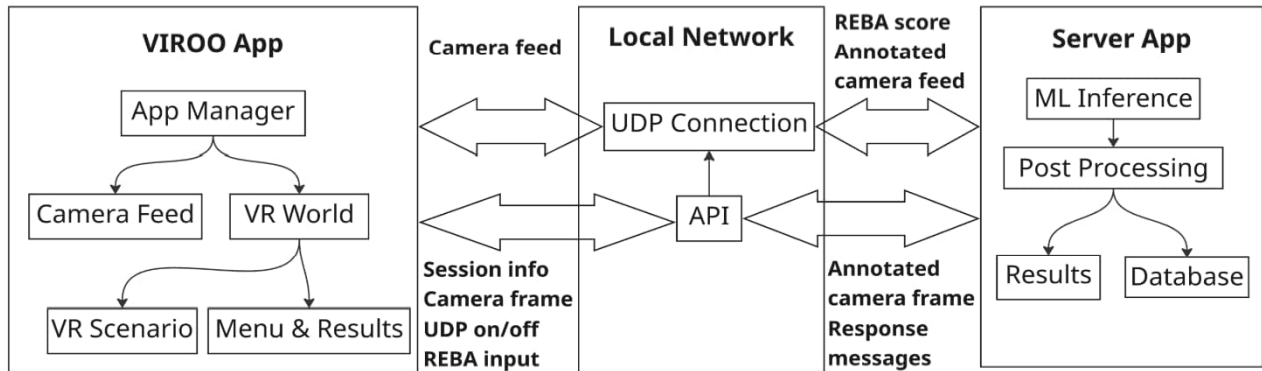


Figure 33. Overview of the Ergon-XR system architecture

Main Components

On the Unity editor, the Content Creator is provided with the tools to prepare a training scenario with different parameters and varying complexity so that different setups can be assessed for their physical ergonomics impact on the Trainee. During runtime, the backbone of the system is the ergonomics assessment module which is responsible for processing the input from the Trainer's UI and the camera feed.

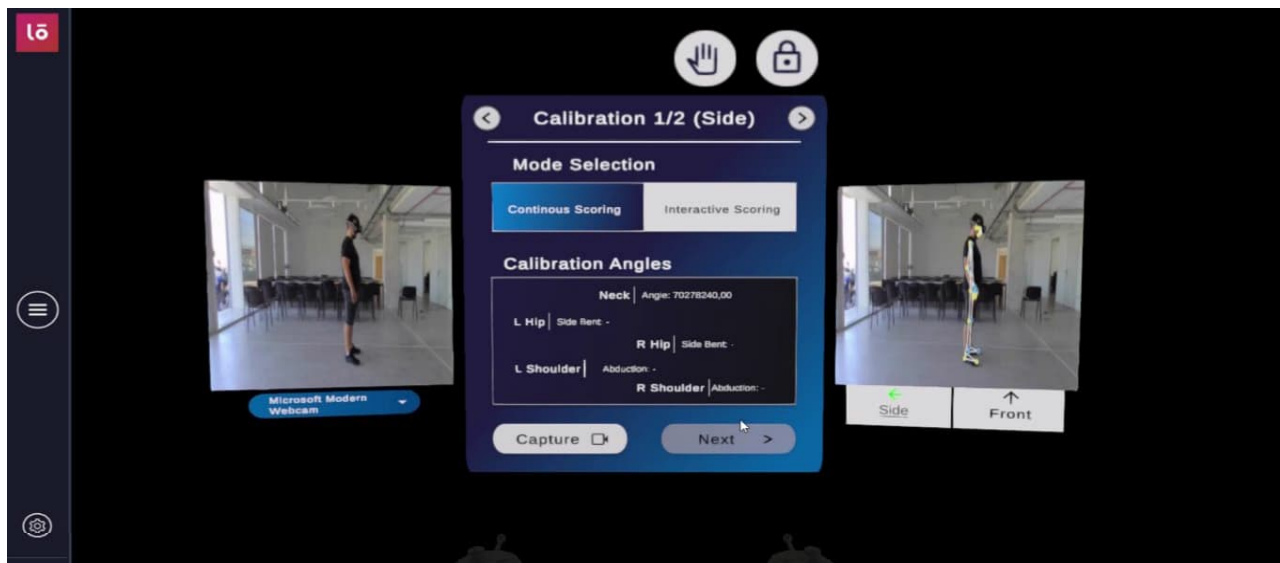


Figure 34. Physical Pose Calibration Flow

Runtime Flow

The runtime flow begins with the selection of the Trainer or Trainee role for each user. The Trainer is driving the process while the Trainee follows along with the verbal instructions of the Trainer and prompts on their HMD UI. The Trainer can select between two modes, either Continuous or Interactive Scoring. Both modes begin with a calibration step (Figure 34) so that the results of the physical ergonomics score are compensated for systematic errors in the joint angle definitions and adjusted for the Trainee's normal posture. The next step in both modes is for the Trainer to select on their UI which joints will be manually assigned a partial REBA score and which will be visually estimated to calculate the total score.

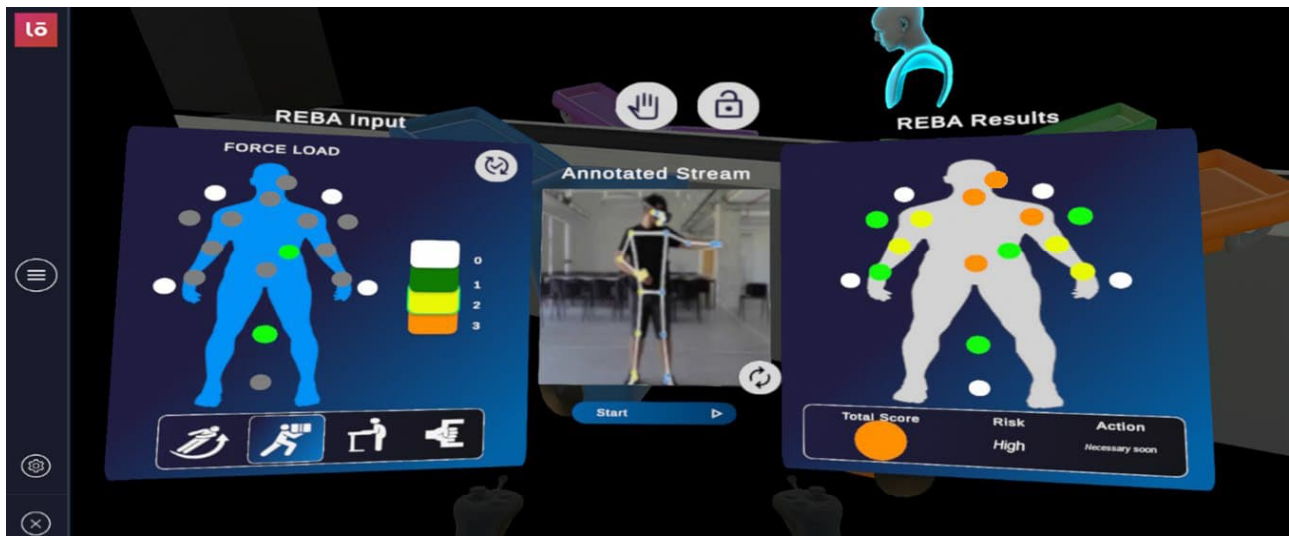


Figure 35. Runtime REBA input and results



Figure 36. Results panel with joint-angle plots and video review of the session

Analysis and Feedback

On the Continuous Scoring mode, the Trainee's pose is detected on each frame and, based on the estimations of the 3D locations of the pose landmarks, a physical ergonomics risk score is estimated and visualized on the Trainer's UI (Figure 35). At the end of the scenario the results for every frame are saved for offline review, the three poses with the highest score are identified and the application flow allows the transition to the Interactive Scoring mode to investigate that specific pose (Figure 36). In the Interactive Scoring mode, the Trainee assumes the identified static pose and the Trainer moves the camera at different points of view from which individual joint scores can be accurately measure to estimate a precise total score for this pose.

Conclusion

The Eron-XR app is a VIROO XR-asset that can be used by a Trainer for ergonomics training in any virtual scenario. The Trainer is provided with a user interface that allows fine control over the REBA score calculation by defining metrics that cannot be automatically calculated from visual input. The asset is using an off-the-shelf RGB camera which makes it simple to install and integrate compared to an RGBD depth camera. The missing three-dimensional aspect is compensated by the capabilities of the pose landmark estimator which infers 3D poses based on the Google MediaPipe foundation model which however is not perfect. To address potential inconsistencies in the estimation, the flow and UI have been adapted appropriately to give the Trainer full control and transparency over the results based on the two available modes and logged data for online and offline review, respectively.

Acknowledgments

The Ergon-XR Project has received funding from the EU Horizon Europe research and innovation action program, via the Open Call #1 – MASTER, issued and executed under project MASTER (grant agreement Nr. 101093079).

Disclosure of Interests

The authors have no competing interests to declare. The VIROO platform is a product of VIRTUALWARE, member of the MASTER Consortium which funded this project via the Financial Support for Third Parties mechanism.

References

- Aberger, J., Shami, S., Häcker, B., Pestana, J., Khodier, K., Sarc, R. 2025. Prototype of AI-powered assistance system for digitalisation of manual waste sorting. *Waste Management* 194, 366–378. <https://doi.org/10.1016/j.wasman.2025.01.027>
- Chu, C.H., Pan, J.K., Chen, Y.W. 2025. Ergonomic workplace design based on real-time integration between virtual and augmented realities. *Robotics & Computer-Integrated Manufacturing* 92, 102859. <https://doi.org/10.1016/j.rcim.2024.102859>
- Emmatty, F.J., Panicker, V.V., Baradwaj, K.C. 2021. Ergonomic evaluation of work table for waste sorting tasks using digital human modelling. *International Journal of Industrial Ergonomics* 84, 103146. <https://doi.org/10.1016/j.ergon.2021.103146>
- Hignett, S., McAtamney, L. 2000. Rapid Entire Body Assessment (REBA). *Applied Ergonomics* 31(2), 201–205. [https://doi.org/10.1016/S0003-6870\(99\)00039-3](https://doi.org/10.1016/S0003-6870(99)00039-3)
- Hossain, M.S., Azam, S., Karim, A., Montaha, S., Quadir, R., De Boer, F., Altaf-UI-Amin, M. 2023. Ergonomic Risk Prediction for Awkward Postures From 3D Keypoints Using Deep Learning. *IEEE Access* 11, 114497–114508.
- Lowe, B.D., Dempsey, P.G., Jones, E.M. 2019. Ergonomics assessment methods used by ergonomics professionals. *Applied Ergonomics* 81, 102882. <https://doi.org/10.1016/j.apergo.2019.102882>
- Manghisi, V.M., Evangelista, A. Uva, A.E. 2022. A Virtual Reality Approach for Assisting Sustainable Human Centered Ergonomic Design: The ErgoVR tool. *Procedia Computer Science*, 1338–1346. <https://doi.org/10.1016/j.procs.2022.01.335>
- Yang, Z., Song, D., Ning, J., Wu, Z. 2024. A Systematic Review: Advancing Ergonomic Posture Risk Assessment Through the Integration of Computer Vision and Machine Learning Techniques. *IEEE Access* 12, 180481–180519.

SRswift: A Spatial Reasoning Library in Swift

Philipp Ackermann

Metason & ZHAW School of Engineering, Switzerland

Corresponding author: philipp@metason.net & philipp.ackermann@zhaw.ch

Keywords: Spatial Computing, Knowledge Representation, Reactive AR Programming

Hybrid AI for Spatial Reasoning

Scene Understanding based on Deep-Learning Models

Research in 3D scene understanding has made significant progress recently. Annotated RGB-D datasets such as ARKitScenes² and Locate 3D³ are increasingly being available to train semantic 3D segmentation using deep learning (DL) methods. Although computing intensive, DL-based 3D object detection can run on mobile devices as demonstrated by Apple's RoomPlan⁴ library. In case DL-based scene understanding will improve further in the future, the need for semantic processing of recognized 3D objects will also increase. By choosing a hybrid AI approach that combines the advantages of deep learning recognition and symbolic processing, eXtended Reality (XR) applications will therefore gain in performance. Instead of managing spatial objects with (hard-coded) geometric calculations, SRswift leverages this to a symbolic level that allows dynamic loading of spatial logic during run-time.

Actionable Spatial Knowledge using Symbolic Processing

SRswift is a declarative spatial reasoning library that combines geometric facts of recognized objects with symbolic predicates and relationships to perform important tasks such as determining the arrangement of 3D objects among each other ("on", "behind", "near", etc.). Its foundation relies on oriented 3D bounding box representations (provided by DL object detectors), extended by a comprehensive set of spatial predicates, ranging from topology and connectivity to directionality and orientation, expressed in a formalism related to natural language. The derived predicates form a spatial knowledge graph and, in combination with a pipeline-based inference model, enable spatial queries and dynamic rule evaluation on detected and virtual 3D objects.

² <https://machinelearning.apple.com/research/arkitscenes>

³ <https://locate3d.atmeta.com>

⁴ <https://developer.apple.com/augmented-reality/roomplan>

Implementation of SRswift

SRswift is a library in the Swift programming language that implements the Spatial Reasoner⁵ framework which is documented in (Häsler 2025) where its motivation and related work can be found. SRswift⁶ is available as free open-source library and can be integrated as a package into macOS, iOS and visionOS projects.

Using SRswift with ARKit

Apple's ARKit⁷ library (Nhan 2024) enables the development of Augmented Reality (AR) applications for iOS mobile devices (iPad and iPhone) and for visionOS-based head-mounted displays. ARKit captures the environment by detecting planes and may additionally track images, faces, and body poses. SRswift includes a sample implementation of generating spatial object instances into the SRswift fact base from detected planes that are classified by ARKit with 7 labels (wall, floor, ceiling, table, seat, door, window). The user of the AR application is assigned to a spatial object that represents the bbox of his avatar. In addition, virtual 3D objects that are anchored in the AR scene are also mapped to spatial objects. SRswift with ARKit have been integrated and validated in the ARchi VR App⁸ (see Figure 37 as an example of this setup).

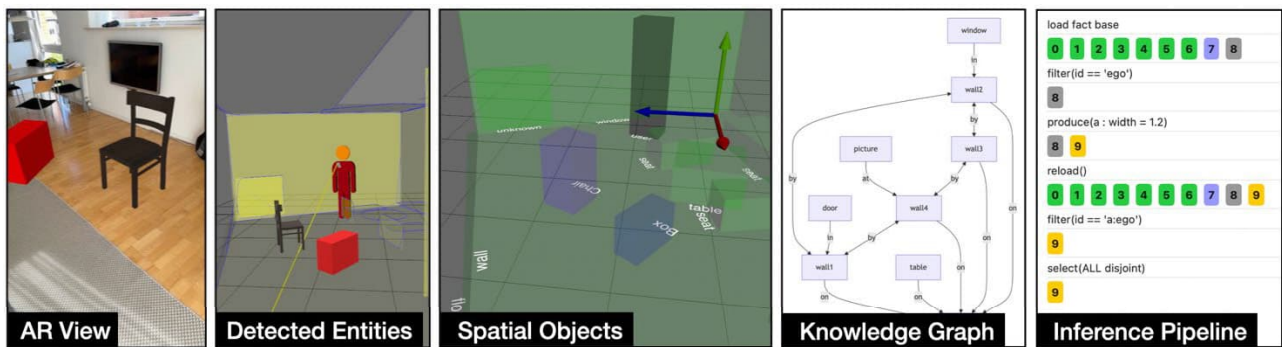


Figure 37. *AR View:* the user's augmented reality view; *Detected Entities:* detected entities such as planes or objects, anchored virtual 3D elements, and the user's avatar; *Spatial Objects:* captured and virtual items loaded into the fact base as spatial objects; *Knowledge Graph:* deduced spatial relations between the objects; *Inference Pipeline:* running a spatial query on symbolic predicates specified as inference pipeline.

Semantic Spatial Reasoning

By loading the generated spatial objects into the fact base of SRswift, inference pipelines can be executed on the automatically generated spatial knowledge graph to gain meaningful insights. The spatial objects in the output list of the last inference operator in the pipeline can be interpreted as search result, used to make decisions (empty list means failure), improve scene understanding, and generate new spatial artifacts. Due to the declarative specification of inference pipelines as text strings, they can be dynamically loaded and executed in XR applications at runtime.

⁵ <https://github.com/metason/SpatialReasoner>

⁶ <https://github.com/metason/SRswift>

⁷ <https://developer.apple.com/augmented-reality/arkit/>

⁸ <https://archi.metason.net>

Use-Cases of Spatial Reasoning with SRswift

Refining Object Classification

DL-based 3D object detectors support object classification limited to the number of trained labels. If the trained labels do not exactly match the domain requirements, spatial reasoning can be used to reassign types (that fit with the loaded taxonomy) and improve classification accuracy by incorporating geometric and spatial knowledge. The following example refines a detected object of type "seat" (as detected by ARKit) and details the classification to "couch" if its length exceeds 1.25 meters:

```
filter(type == 'seat' AND length > 1.25) | map(type = 'couch')
```

Reactive AR Programming

AR experiences rely on understanding the current spatial context and are heavily driven by events triggered by the detection of markers, images, objects, and people. In (Ackermann 2023) the Event-Condition-Action pattern is leveraged as a generic abstraction from the reactive behavior of AR software systems. By constantly evaluating inferences on the spatial knowledge graph of the actual AR scene, the first two aspects of an Event-Condition-Action rule are covered. When the spatial reasoner is delivering a result from the inference, the Action should be executed. This approach is supported in the ARchi VR App with so-called "Spatial Detectors" that are declared in a JSON data structure, e.g.:

```
{
  "do" : "detect",
  "spatial" : "filter(type == 'ceiling' AND footprint > 2.5)",
  "op" : "say('ceiling detected')"
```

Spatial Rule System

An inference pipeline of a spatial reasoner can be viewed as a spatial rule. The presence of a result after execution can be interpreted as a pass or fail. A collection of inference pipelines forms a rule set. A rule system executes all rules in a rule set in a loop until no more changes occur. SRswift tracks changes and can therefore easily serve as the basis for implementing a spatial rule system. By continuously evaluating the scenery of an AR experience with a spatial rule system, a conceptual model (Muff 2025) of the situational context can be generated thanks to the semantically enriched understanding.

Authoring and Testing of Spatial Logic in AR Applications

Creating meaningful inference pipelines and spatial rule sets is challenging with-out supporting tools. Using the SRswift library on macOS, a desktop application was developed that integrates an inference editor, a ruleset composer, a 3D scene viewer, an inference execution visualizer, and an inspector of spatial objects in the fact base (**Figure 38**).

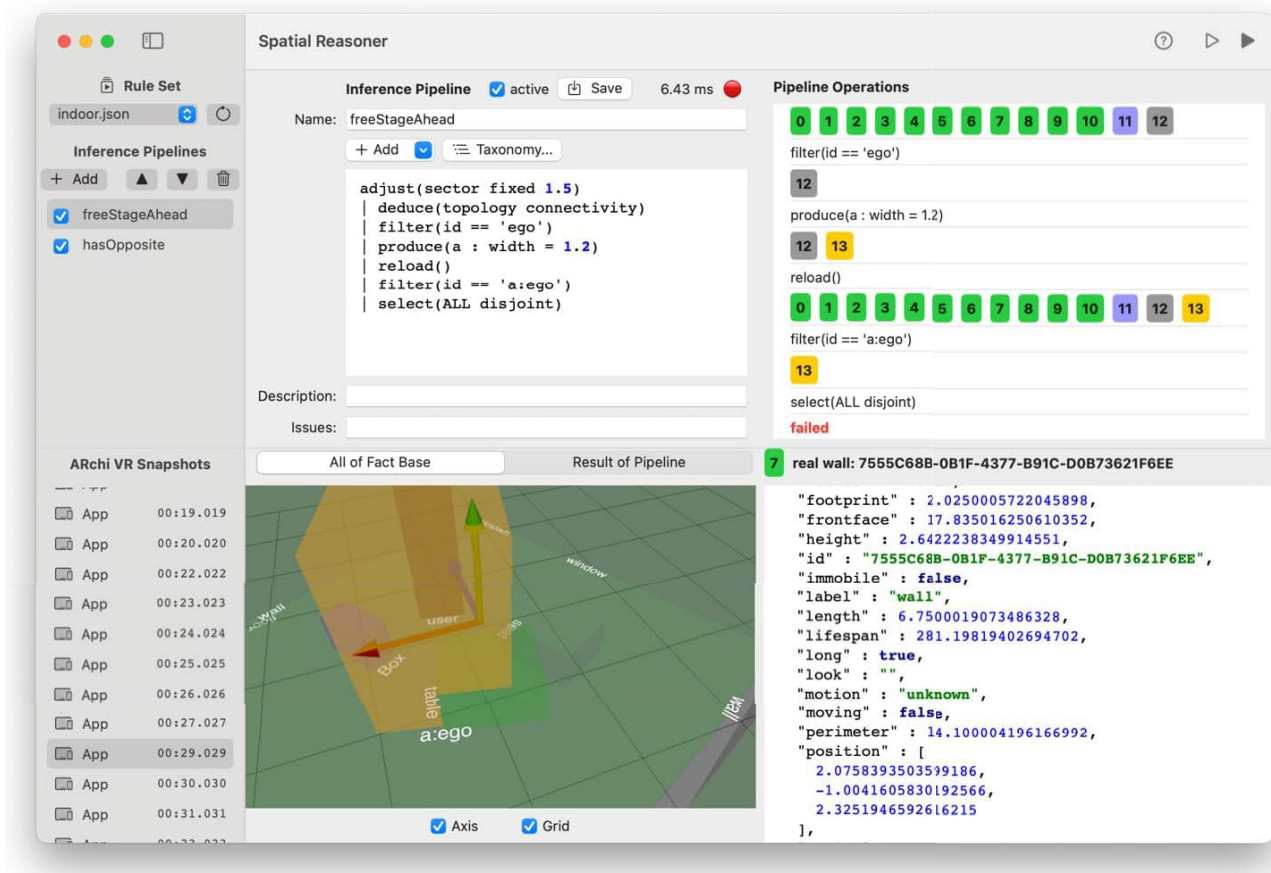


Figure 38. Interactive editor and test environment for spatial rules.

Iteratively editing and testing spatial rules is a central element of a productive authoring process. However, it presents challenges when authoring (on the desktop device) is separated from runtime execution (on the mobile device or HMD). To best enable testing in the authoring tool, snapshots of the AR application's runtime environment are wirelessly sent to and stored in the desktop application, allowing tests to be executed later without a mobile device. Snapshots contain objects captured by ARKit (e.g., detected planes), virtual 3D augmentations, user poses, screenshots, and depth maps. During a recording session, the AR app sends a snapshot every second. By making the recorded snapshots available in the authoring tool, a rule developer can simulate and test the behavior of the rule set as the number of detected objects grows and becomes more precise. Test results can be evaluated visually in 3D and semantically by inspecting the fact base and the executed inference pipelines.

References

- Ackermann, P. 2023. AR Patterns: Event-Driven Design Patterns in Creating Augmented Reality Experiences. In: Zachmann, G., et al. Virtual Reality and Mixed Reality. EuroXR 2023.
- Häsler, S., & Ackermann, P. 2025. Spatial Reasoner: A 3D Inference Pipeline for XR Applications. In: Proceedings of the ICVARs '25. Association for Computing Machinery
- Muff, F. 2025. Metamodeling for Extended Reality. Springer. <https://doi.org/10.1007/978-3-031-76762-3>
- Nhan, J. 2024. Mastering ARKit: Apple's Augmented Reality App Development Platform. Apress: Clarendon, Oxford, UK. <https://doi.org/10.1007/979-8-8688-0847-0>

V CARE: A Virtual Reality (VR) Artificial Intelligence (AI) Assisted Training Tool for Alzheimer's Caregivers

Annie Wan

Faculty of Creative and Critical Studies, University of British Columbia (Okanagan Campus), Canada

Corresponding author: annie.wan@ubc.ca

Keywords: Alzheimer, healthy aging, virtual reality training, artificial intelligence

Introduction

Medical Humanities represents interdisciplinary research at the confluence of the arts, humanities, social sciences, and health sciences. This field critically examines the social, historical, cultural, and ethical aspects of health sciences and healthcare. It aims to unearth more profound insights, intricately exploring the nuances of healthcare experience and the intricacies of medical practice. Through its comprehensive lens, medical humanities endeavors to deepen our understanding of patient experiences, fostering empathy within healthcare contexts and contributing to the development of more ethical medical practices.

V Care promotes the understanding and inclusion of caregivers for individuals with Alzheimer's. We adopted a co-design approach that included people with lived experience in the research and design process, which aligns with the idea of developing the capacity for positive change in health through immersive technologies. The project also collaborated with practicing health professionals and included interviewees and participants from diverse cultural backgrounds, thereby fostering diversity throughout the research. Virtual reality (VR) can be used for Alzheimer's caregiver training to help prepare caregivers for the challenges taking care of someone with the condition.

Methodologies

VR provides an immersive and interactive experience that simulates various scenarios and situations that may arise when caring for Alzheimer's. Our VR training application suggested here could simulate an episode of confusion or disorientation that a person with Alzheimer's may experience. This can help the caregiver understand what the person is going through and how to respond in a way that is calming and supportive. Other VR scenarios included demonstrating how to assist with daily activities, such as cooking and eating, or how to handle challenging behaviors, such as wandering or aggression. In addition to helping caregivers better understand the needs of the Alzheimer's, our VR training can also be a valuable tool for reducing stress and increasing confidence in their caregiving role. By providing a safe and controlled environment to practice different scenarios, our VR application helped prepare caregivers for the challenges they may face in their day-to-day caregiving duties. To enhance the

immersive experience, the application also implemented OpenAI's speech-to-text library during training. Desmond O'Neill, professor of Medical Gerontology at the Trinity College Institute of Neuroscience in Dublin, Ireland, suggests how "further develop[ment] [of] the medical humanities for geriatric medicine represents an important addition to our repertoire and articulacy, facilitating participation in important trends such as reframing aging, an initiative to emphasize the positivities and worth of later life" (O'Neill, 228). This project is an interdisciplinary research-creation that bridges the fields of medical humanities, arts and humanities, artificial intelligence (AI) technology, and immersive technologies. The objectives are: 1. to develop a training application that equips caregivers with skills to handle different situations of Alzheimer's and builds their confidence; 2. to promote healthy aging as "the process of developing and maintaining the functional ability that enables wellbeing in older age., 3. to foster caregivers' relationships with Alzheimer's and encourage them to share their VR training experiences through interviews, focus groups, and a co-design process.



Figure 39. One of the VR scenarios which trained caregivers to be aware of food details



Figure 40. One of the VR scenarios which provides choices for the trainees.

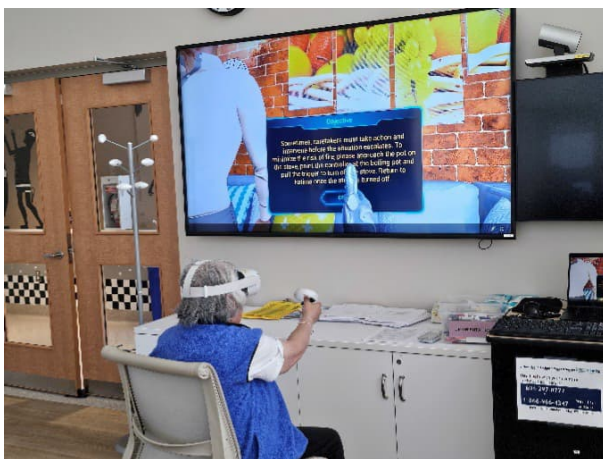


Figure 41. A lived experience in training



Figure 42. A practicing nurse in training

Result, Observation and Suggestions

In the research, we conducted interviews with the trainees, including caregivers with lived experience, practicing caregivers, and practicing nurses. However, in this poster, we will focus on the results gathered from training with six practicing nurses. In general, most of them face similar situations very often, i.e., taking care of Alzheimer and they mostly feel the application will be very beneficial, particularly for student nurses; as they recalled, during their nursing education, there were low-quality training manikins, and no one ever had an actor as a patient during the education. In addition, they found that the application increased their confidence in handling Alzheimer's, even though they were practicing nurses when they received their training. They also acknowledged that the application would remind them of the details of caring for Alzheimer's patients if they had the chance to be trained every one or two months. Initial observations of user experiences revealed significant acclimatization challenges, with nurses exhibiting pronounced disorientation upon entering the VR environment, manifesting as physical imbalance, visual shock, and hesitation.

While object interaction (e.g., grabbing) proved moderately intuitive for some, navigation mechanics universally required extensive verbal guidance, indicating that pre-training specifically in VR interaction will be required in other similar experiments. Spatial design significantly impacted usability; a larger physical workshop area enabled natural walking-based navigation, reducing reliance on artificial teleportation—which some users (e.g., Nurse 2; Nurse 5) mastered quickly while others struggled with depth perception (Nurse 6)—critical safety features like boundary warnings induced alarm, indicating further contextual or verbal instructions should be integrated.

Different people varied substantially in interacting with the VR user interface. Nurse 1 disregarded UI text entirely, despite verbal instructions being given; however, they responded positively to condensed notification banners, suggesting that we might provide too much textual information in UI elements. Conversely, Nurse 2 relied heavily on reading UI text aloud to process objectives, highlighting inconsistencies in the verbal clarity of instructions. Nurse 4 adopted a deliberate, introspective approach—thoroughly reading UI, exploring non-essential interactions (e.g., reading extended dialogue of one of the characters), and treating scenarios with clinical gravity, though disorientation persisted post-session. Later, participants (Nurse 5 and Nurse 6) progressed markedly faster, with Nurse 5 exhibiting anticipatory task completion and fluid use of the controller, raising questions about the relationship between knowledge retention and procedural familiarity. Notably, Nurse 6 displayed high exploratory engagement—interacting with non-highlighted objects and emotionally connecting with scenarios—yet struggled with spatial judgment.

To conclude, the application successfully stimulated engagement, with few nurses describing it as "fun" or "cool" (Nurse 3, 5, 6), as well as performance acceleration across later users signals that structured exposure—not merely repetition—drives proficiency, advocating for scaffolded onboarding that progressively introduces mechanics before complex clinical decision-making. Nurses exhibited diverse interaction patterns, which were tied to their prior exposure, learning preferences, and engagement levels. Nurse 3 initially required extensive support for basic mechanics (e.g., grasping) but demonstrated fluency after Level 1, suggesting rapid skill acquisition following orientation. However, hardware-induced discomfort (disorientation upon headset removal) and persistent UI navigation difficulties (Nurse 4) remained barriers. Therefore, pre-training on basic interaction and navigation in VR should be adopted in future iterations. Nurse 3's characterization of cutting mechanics as a "game" and Nurse 6's emotional response to patient scenarios suggest an immersive potential. Although novelty-driven exploration may sometimes divert focus from core objectives, it has been proven that the novelty and gaming elements are crucial in engaging trainees.

The OpenAI speech-to-text recognition triggered uniform hesitation, revealing inadequate preparation for voice-based interaction; therefore, it is recommended not to use voice commands or voice-activated technology when engaging communities and instead, use AI to generate 'more useful' guidance in text, especially when trainees experience difficulties in either UI interaction or VR interactions in general. Apart from potential novelty, future iterations should incorporate more gaming elements. Nurse 6's reported feeling relaxed after interacting with their peers suggested that social scaffolding enhances comfort; hence, training activities, similar to focus groups and group interviews, should be continued in the next step.

References

- Caponnetto P., & Casu P. 2022. Update on Cyber Health Psychology: Virtual Reality and Mobile Health Tools in Psychotherapy, Clinical Rehabilitation, and Addiction Treatment. *Int J Environ Res Public Health* <https://www.mdpi.com/1660-4601/19/6/3516>
- Kupczik L., Farrelly W. & Wilson S. 2022. Appraising Virtual Technologies' Impact on Older Citizens' Mental Health—A Comparative between 360° Video and Virtual Reality. *Int J Environ Res Public Health* <https://www.mdpi.com/1660-4601/19/18/11250>
- Sung HC., Su HF., Lee WL., Yamakawa M. & Wang HM. 2022. Effects of a dementia virtual reality-based training with peer support for home care workers: A cluster randomized controlled trial. *Int J Geriatric Psychiatry* <https://doi.org/10.1002/gps.5799>
- Matsangidou M., Frangoudes F., Hadjiaros M., Schiza E., Neokleous N., Papayianni E., Avraamides M. & Pattichis C. 2022 Bring me sunshine, bring me (physical) strength: The case of dementia. Designing and implementing a virtual reality system for physical training during the COVID-19 pandemic. *Int J Human-Computer Studies* <https://doi.org/10.1016/j.ijhcs.2022.102840>
- Matsangidou M., Frangoudes F., Hadjiaros M., Schiza E., Neokleous N., Papayianni E., Avraamides M. & Pattichis C. 2022 Participatory design and evaluation of virtual reality physical rehabilitation for people living with dementia. *Virtual Reality for Therapy, Psychological Interventions, and Physical and Cognitive Rehabilitation* <https://link.springer.com/article/10.1007/s10055-022-00639-1>
- Flynn A., Healy D., Barry M., Brennan A., Redfern S., Houghton C., & Casey D. 2022. Key Stakeholders' Experiences and Perceptions of Virtual Reality for Older Adults Living With Dementia: Systematic Review and Thematic Synthesis. *JMIR Serious Games*. <https://games.jmir.org/2022/4/e37228>
- O'Neill, D. 2015. *Geriatric Medicine and Cultural Gerontology*. *Age and Ageing*, 44:3, 353-355.
- O'Neill, D. 2019. *Geriatric Medical Humanities: Fresh Insights into Ageing and Geriatric Medicine - European Geriatric Medicine*. SpringerLink, Springer International Publishing, 10 Apr. link.springer.com/article/10.1007/s41999-019-00187-z.
- O'Neill, D. 2021. Lessons for Humanities and Arts in Gerontology and Geriatrics Curricula from the Medical and Health Humanities. *Innovation in Aging*, 5, Supplement_1, 69-69.

INFINIBRICK: A Tool for the Generation of Endless LEGO Assembly Tasks for Studies on Prolonged Immersive Virtual Reality Exposure

Enricoandrea Laviola, Luana Marangelli, Sara Romano, and Michele Gattullo

Department of Mechanics, Mathematics, and Management, Polytechnic University of Bari, Italy

Corresponding author: s.romano3@phd.poliba.it

Keywords: Endless Assembly Task, Industrial Assembly Simulation, Prolonged Exposure, Experiment Design, Immersive Virtual Reality.

Introduction and motivation

Immersive Virtual Reality (IVR) environments have gained prominence in training operators for industrial activities to improve performance (Abidi et al., 2019). Recent research efforts have shifted toward exploring how the design of virtual surroundings influences user experience during training tasks (Kiluk et al., 2023). A well-designed IVR training environment should not hinder operator performance, but rather enhance various dimensions of user experience, including well-being and sense of presence. However, evaluating these effects entails studies in which users engage in tasks over extended durations to assess different virtual environment designs. Current setups often rely on time-limited or step-based tasks, which may not effectively support long-term evaluation. Tasks of potentially unlimited duration are needed to ensure temporal flexibility and enable examination of how specific elements, such as biophilic ones, influence user experience in a virtual workplace.

To the best of our knowledge, no existing tool enables researchers to easily and rapidly create experimental setups for studies without fixed time constraints. To address this gap, we proposed INFINIBRICK, an IVR-based tool for abstract object construction using LEGO bricks. LEGO has been extensively employed in research to simulate assembly tasks (Alves et al., 2019). The tool provides two modes of operation: experimenter mode and user mode. In experimenter mode, researchers can configure the assembly task directly in the IVR environment. In user mode, participants engage in the assembly task. The goal is to provide the scientific community with a ready-to-use tool that reduces technical complexity, making IVR experimentation more accessible. The tool facilitates the design of replicable IVR studies aimed at investigating parameters requiring prolonged exposure. These include measures such as user attention and focus, potentially assessed through eye-tracking technology, as well as psycho-physiological factors including cognitive load, emotional state, stress, and anxiety.

System overview

INFINIBRICK was developed for the HTC Vive Pro Eye, a video see-through head-mounted display, and implemented using Unity 3D Engine version 2022.3.19. A video showing the main features is available at <https://youtu.be/LLV0Vt75Ew8>. The current version of the system includes an empty IVR environment within which a table is centrally placed with a green 60x20 LEGO Duplo baseplate serving as the main assembly ground for the bricks. Grey boxes for storing LEGO bricks are distributed along the table, with a red box for discarding incorrectly selected bricks during the task. Above the table, a Graphical User Interface (GUI) is positioned to enable the experimenter to design the endless assembly task, configuring the LEGO set, i.e., the number of different LEGO bricks used for the task, with their shape and color (see Figure 43, Left). This interface is positioned at an appropriate height for standing use, anticipating future implementations in which the experimenter will configure the environment by inserting specific elements (e.g., furniture or biophilic elements) to evaluate various aspects of user experience during task performance. The execution of the assembly task is designed for prolonged exposure and intended to be performed while seated.

When the application starts, the GUI offers the experimenter a textual overview of the main INFINIBRICK features. Then, the experimenter can select whether to create a new LEGO set or load an existing one to use or modify. In the LEGO set configuration, the experimenter's first choice is to determine the number of boxes to include. Afterwards, a menu with several buttons, each representing a box, is displayed for the LEGO selection and, after pressing one button, the experimenter has access to 20 different shapes of LEGO bricks. After the selection, a copy of the chosen LEGO appears on a new interface for the next step, which concerns assigning a color from ten available options. Then, the experimenter can choose another box to select a new LEGO brick. Assigning the same shape of LEGO brick to multiple boxes is possible, but not the same color for the same shape. Once all LEGO bricks have been assigned to their respective boxes, i.e., when the LEGO set is fully configured, the experimenter can specify the time limit for the endless assembly task. The experimenter proceeds to the summary page and lets the user begin the assembly task with a potentially unlimited number of bricks within the time limit.

During the task, the user can be free to assemble any LEGO construction as desired or can be guided by visual instructions on which LEGO brick to pick up and where to place it. To convey the required information types, a product model of the brick with the same shape and color as the brick to be picked up was used. Specifically, two *LEGO cues* are provided: animated and static. The *animated LEGO cue* helps the user quickly identify the target brick position with a translation effect following the guidelines proposed in (Laviola et al., 2025). The *static LEGO cue* supports precise alignment by guiding the user on correctly positioning the brick. To differentiate these cues from already placed bricks and the one picked up, the *LEGO cues* are shown with an opacity set to 0.5 (see Figure 1, Right). The shape and color for each new LEGO brick to assemble, as well as its position and orientation, are randomly determined. The user can take one brick at a time from the boxes and place it on the target position on the green baseplate. If the user completes the step correctly, the system provides audio feedback and displays the next brick to be placed. If an incorrect brick is picked up, the system does not accept it at the target position, and audio feedback is not provided, prompting the user to discard it and pick the correct one. To maintain an endless assembly task, the system generates a new LEGO brick in the box after one is taken. This process repeats until the time expires, after which a GUI shows the count of correctly placed bricks and the incorrectly picked ones.

System evaluation

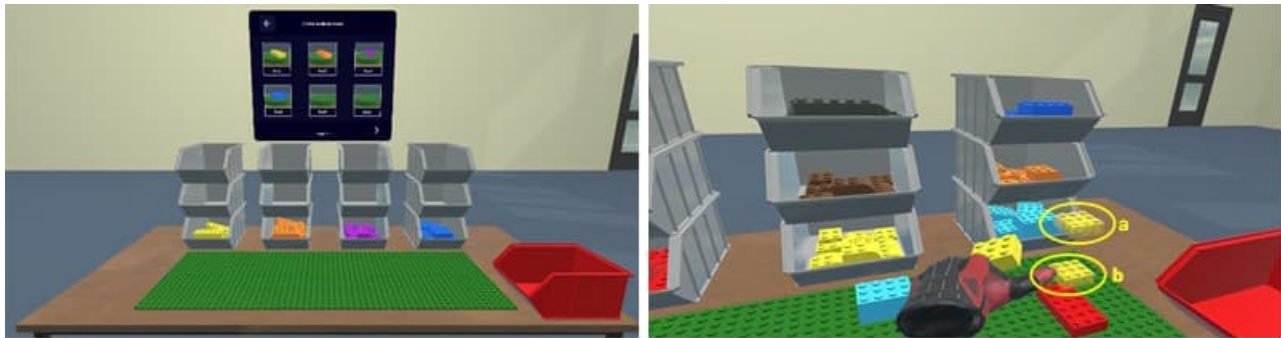


Figure 43. Left: The INFINIBRICK setup with the storage boxes, a red discard box, and the GUI for the LEGO set configuration. Right: A user engaging the endless assembly task following the LEGO cues: animated (a), static (b).

A user study was designed to evaluate INFINIBRICK, assessing experimenter-side usability, as well as user experience and perceived time on the user-side. The study tested all the implemented features and the execution of the endless assembly task using visual instructions. The experiment consisted of two phases: in one, the participant acted as the experimenter; in the other, as the end user. 20 participants (8 females, aged 24-49, mean=26.40, SD=5.38) took part in the experiment with a mean IVR familiarity level of 3.50 (SD=1.80, Min=1, Max=7).

In the role of experimenter (Task 1), participants completed two subtasks: (i) the guided creation of a LEGO set, and (ii) the modification of a pre-existing LEGO set, following the experimenter's verbal instructions on selecting boxes in a specified order, assigning LEGO shapes and colors to bricks in each box, and performing any necessary actions, such as adding/removing boxes or modifying the bricks in a given box. Execution time was collected, followed by the System Usability Scale (SUS) upon completion. In the user role (Task 2), participants performed a 5-minute LEGO assembly task, without knowing the duration, then reported perceived task execution time and filled out the User Experience Questionnaire (UEQ).

Analyzing the results, the average execution time for Task 1 was 5.58 min (SD=1.36, Min=3.10, Max=9.30). The mean SUS score was 91 (SD=9.22, Min=60, Max=100), rated 'Excellent' usability according to (Bangor et al., 2008). In Task 2, participants reported a highly favorable user experience, with all UEQ factors categorized as "Excellent" compared to the UEQ benchmark (Schrepp et al., 2017). Participants perceived the duration of the task to be 5 min on average (SD=2.65, Min=1.00, Max=11.30), which was the actual duration.

Findings indicate that all participants successfully configured the LEGO sets quickly and autonomously, confirming the effectiveness of the user-friendly interface. The interaction method for picking up and placing the LEGO bricks, and all the features employed during the assembly task (such as LEGO cues and sound feedback), had no notable impact on the sense of presence within the empty IVR scenario. This finding is important because INFINIBRICK offers a neutral workbench within an empty IVR environment. By avoiding introduced variables (e.g., differing levels of realism in furnishings), it ensures a stable context for investigating parameters that depend on prolonged exposure, reducing the risk of biased outcomes (Wearden 2016).

Conclusion and future improvements

This study introduced INFINIBRICK, a tool designed to facilitate the simulation of endless assembly tasks within an IVR environment. By enabling researchers to configure assembly tasks directly in IVR, INFINIBRICK can significantly reduce the technical complexity required for conducting controlled user studies that explore changes in environmental conditions. The tool was validated in experimenter and user modes, demonstrating excellent usability and user experience, ensuring a neutral and consistent task, unaffected by external experimental variables. Future developments will focus on enhancing the configurability of the IVR environment, introducing a GUI that allows the experimenter to customize it (e.g., with furniture and biophilic elements) and to make INFINIBRICK publicly available for the scientific community. We aim to place INFINIBRICK as a versatile tool for conducting replicable assembly studies, extending its applicability to a broader range of experimental contexts.

Acknowledgments

This work was supported by European Union - Next Generation EU, NATIONAL RECOVERY AND RESILIENCE PLAN (NRRP) - MISSION 4 Component 2 Investment 1.1 – “Fund for the National Research Program and for Projects of National Interest (NRP)” - CUP D53D23017380001.

References

- Abidi, Mustufa Haider, Abdulrahman Al-Ahmari, Ali Ahmad, Wadea Ameen, & Hisham Alkhalefah. 2019. Assessment of Virtual Reality-Based Manufacturing Assembly Training System. *International Journal of Advanced Manufacturing Technology* 105(9): 3743–59. <https://doi.org/10.1007/S00170-019-03801-3>.
- Alves, J, B Marques, M Oliveira, T Araujo, P Dias, & B Santos. 2019. Comparing Spatial and Mobile Augmented Reality for Guiding Assembling Procedures with Task Validation. *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. <https://doi.org/10.1109/ICARSC.2019.8733642>.
- Bangor, Aaron, Philip T. Kortum, & James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24 (6): 574–94. <https://doi.org/10.1080/10447310802205776>.
- Kiluk, A, V Paneva, S Seinfeld, & J Muller. 2023. The Impact of Different Virtual Work Environments on Flow, Performance, User Emotions, and Preferences. *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. <https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00063>.
- Laviola, E, M Gattullo, S Romano, & AE Uva. 2025. Which Side Is the Top? A User Study to Compare Visual Assets for Component Orientation in Assembly with Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/TVCG.2025.3549164>.
- Schrepp, M, Hinderks A, & Thomaschewski J. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4:4–40. <https://doi.org/https://doi.org/10.25968/opus-3397>.
- Wearden, John. 2016. The Psychology of Time Perception. <https://link.springer.com/content/pdf/10.1057/978-1-137-40883-9.pdf>.

Fostering inclusive work environments through immersive training solutions for visual impairment awareness

Kalliopi-Evangelia Stavroulia¹, Maria Christofi¹, and Andreas Lanitis^{1,2}

¹*CYENS Centre of Excellence, Cyprus*

²*Cyprus University of Technology, Cyprus*

Corresponding author: k.stavroulia@cyens.org.cy

Keywords: Virtual Reality, Inclusive Workplaces, People with Disabilities, Visual Impairments

Introduction

This paper presents an immersive-based training initiative that forms part of a broader effort to build a comprehensive framework for disability inclusion training in Vocational Education and Training (VET) and workplace settings, under the InclusiVR@Work project. The InclusiVR@Work (Using VR technology to achieve reasonable accommodation in the workplace) project, aims to reduce the employability gap of People with Disabilities (PwD) in European workplaces by promoting understanding and empathy through a resource training toolkit consisting of immersive experiences, a disability inclusion Massive Open Online Course (MOOC) and a Continuous Professional Development (CPD) Training Programme. The project offers a novel approach to promoting disability awareness and inclusion in the workplace by leveraging Virtual Reality (VR) based perspective-changing experiences that allow end-users to experience first-hand the challenges faced by colleagues with disabilities such as sensory, physical and intellectual impairments, building their empathy and cultivating a more inclusive mindset.

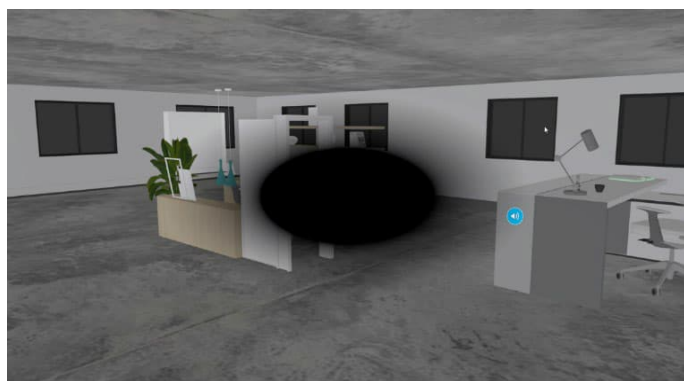
The right to work is a fundamental human right that is not only associated with earning an income, but it also plays a crucial role in promoting mental health, social inclusion and a sense of purpose and well-being (European Commission, 2021; Hanga et al., 2015; Vornholt et al., 2013). Promoting inclusive workplace practices is also aligned with Sustainable Development Goal 3 (SDG 3), which aims to promote mental health and well-being for more resilient societies. The European Union (EU) has taken significant steps to promote accessibility for all people (European Commission, 2021); however, the labor market continues to overlook the realities faced by PwD, and there is a noticeable lack of awareness regarding the challenges those people face during their daily work routine. Workplaces often remain physically or digitally inaccessible, and PwD still do not experience the same access to work opportunities as others (European Commission, 2021; Hanga et al., 2015; Bonaccio et al., 2019). For example, buildings may lack ramps, elevators, or accessible restrooms, while company websites and internal systems may not be compatible with assistive technologies such as screen readers, voice recognition and more. These barriers can make it difficult or even impossible for PwD to participate in

the workforce and sustain their work in the long term. Moreover, hiring practices often underestimate the abilities and potential of PwD (Bonaccio et al., 2019), contributing to higher unemployment rates among disabled populations. Additionally, even when PwD are hired, they may not receive the accommodation and support they need to succeed. Human Resources (HR) managers lack the training and awareness to effectively support employees with disabilities (Bonaccio et al., 2019; McKinney & Swartz, 2021; Syma, 2018). This lack of understanding leads to meaningless changes to workplace culture or infrastructure, making employees with disabilities experience feelings of isolation and marginalization (Vornholt et al., 2013). This not only impacts on their job satisfaction and performance but also prevents the development of truly inclusive workspaces.

Raising awareness of HR staff on the challenges that PwD face is of paramount importance, as their decision-making impacts the development of a positive and inclusive workplace environment. To this end, the InclusiVR@Work project proposes VR-based perspective-changing training that allows HR managers to experience first-hand the challenges faced by colleagues with disabilities and receive practical guidance and suggestions on how to handle various life scenarios, such as facing visual, sensory or intellectual impairments. By emphasizing perspective-taking during the immersive experience, participants will be able to gain deep emotional responses and empathy (Diemer et al., 2015; Ventura et al., 2020; Zavlanou et al., 2021) and reshape their attitudes towards people facing impairments. The project adds real-world value by improving soft skills such as empathy and decision-making of HR managers in supporting people with disabilities, fostering a more inclusive workplace culture. The integration of VR-based training will contribute to the academic body of knowledge by providing scientific data on the impact of immersive experiences on HR managers' training across five EU countries. The remainder of the paper presents the rationale behind the project and its objectives, the development of the scenario and VR experience, and concludes with key reflections and implications.

Scenario Design and Immersive Experience Development

The broader training framework of InclusiVR@Work includes the development of three distinct VR immersive experiences, each simulating a different type of disability: sensory, physical and intellectual impairments. While the overall project will encompass all three areas, this paper focuses specifically on the visual impairment scenario, detailing its conceptual design and technical development. The training scenario offers a sensory-immersive learning experience based on first-person perspective-taking, allowing users to step into the life of a person facing visual impairments such as macular degeneration (Figure 44a) and glaucoma (Figure 44b). The VR scenario developed for this project reflects a workplace evacuation drill, which is a critical safety procedure. The user is placed in the role of an employee with visual impairment, simulating the effects of macular degeneration and glaucoma. The experience begins with the user attempting to read and interpret an evacuation plan, which becomes challenging due to the vision loss. Following the instructions, the user must navigate the office to find the exit, an exercise designed as a maze-like experience, illustrating how the inability to read and interpret essential safety instructions due to vision loss can become a significant barrier to emergency situations. The scenario highlights the need for inclusive safety design, such as auditory guidance systems. More importantly, it allows non-disabled participants such as VET educators and HR managers to "step into the shoes" of their colleagues with visual impairments, fostering their understanding, empathy and awareness in designing inclusive work environments.



a. Experiencing macular degeneration vision disorder



b. Experiencing glaucoma vision disorder

Figure 44. *InclusiVR@Work immersive experiences reflecting visual impairments*

The VR experiences will be integrated into the project's website, so that end-users can access the experience directly through a mobile browser (like Chrome), without needing to install a separate app, ensuring easy access for diverse users across different institutions and EU countries. By designing the VR experiences to be compatible with low-cost mobile VR headsets, the project ensures accessibility and scalability, allowing VET institutions or SMEs to adopt immersive inclusion training as a sustainable, long-term practice.

Conclusions

This paper presented a VR-based training solution focused on disability inclusion in the workplace, specifically simulating a scenario reflecting visual impairments during an evacuation drill. By leveraging VR technology, the project demonstrates how such immersive tools can be used for high-impact empathy-building experiences by VET professionals and HR managers. The proposed prototype is under evaluation to assess its effectiveness, usability, and impact. Feedback from end-users will guide iterative improvements to refine the VR experience, ensuring it meets both technical and pedagogical standards. The final version of the VR experience will be implemented across partner countries to train HR managers and VET professionals to promote a shift in attitudes and practices, encouraging more inclusive workplace environments.

Acknowledgements

This project is co-funded by the Erasmus+ Programme of the European Union under Key Action 2: Cooperation Partnerships, Grant Agreement No. 2024-1-DE02-KA220-VET-000255799.

References

- Bonaccio, S., Connelly, C. E., Gellatly, I. R., Jetha, A., & Martin Ginis, K. A. 2020. The participation of people with disabilities in the workplace across the employment cycle: Employer concerns and research evidence. *Journal of business and psychology*, 35(2), 135-158.
- Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. 2015. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology*, 6, 26. <https://doi.org/10.3389/fpsyg.2015.00026>

- European Commission: Directorate-General for Employment, Social Affairs and Inclusion, Union of equality – Strategy for the rights of persons with disabilities 2021-2030, Publications Office, 2021, <https://data.europa.eu/doi/10.2767/31633>
- Hanga, K., DiNitto, D. M., & Wilken, J. P. 2015. Promoting employment among people with disabilities: Challenges and solutions. *Social Work and Social Sciences Review*, 18(1), 31-51.
- McKinney, E. L., & Swartz, L. 2021. Employment integration barriers: Experiences of people with disabilities. *The International Journal of Human Resource Management*, 32(10), 2298-2320. <https://doi.org/10.1080/09585192.2019.1579749>
- Syma, C. 2018. Invisible disabilities: Perceptions and barriers to reasonable accommodations in the workplace. *Library Management*, 40(1/2), 113-120. <https://doi.org/10.1108/LM-10-2017-0101>
- Ventura, S., Badenes-Ribera, L., Herrero, R., Cebolla, A., Galiana, L., & Baños, R. 2020. Virtual reality as a medium to elicit empathy: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, 23(10), 667-676. <https://doi.org/10.1089/cyber.2019.0681>
- Vornholt, K., Uitdewilligen, S., & Nijhuis, F. J. 2013. Factors affecting the acceptance of people with disabilities at work: A literature review. *Journal of occupational rehabilitation*, 23, 463-475.
- Zavlanou, C., Huber, P., Tisserand, Y., Rudrauf, D., & Lanitis, A. 2021. Procedurally Generated Age-related Visual Deficits in Virtual Reality Environments. In ICAT-EGVE (pp. 19-24).

Globe Viewer for Location-based Augmented Reality

Balázs Ladjánszki¹, Gábor Sörös^{1,2}, Andor Kovács², and Alina Kadlubsky^{1*}

¹Open AR Cloud Europe, Germany

²Nokia Bell Labs, Finland

*Corresponding author: communication@openarcloud.org

Keywords: Mixed/augmented reality, Ubiquitous and mobile computing systems and tools, Location-based services

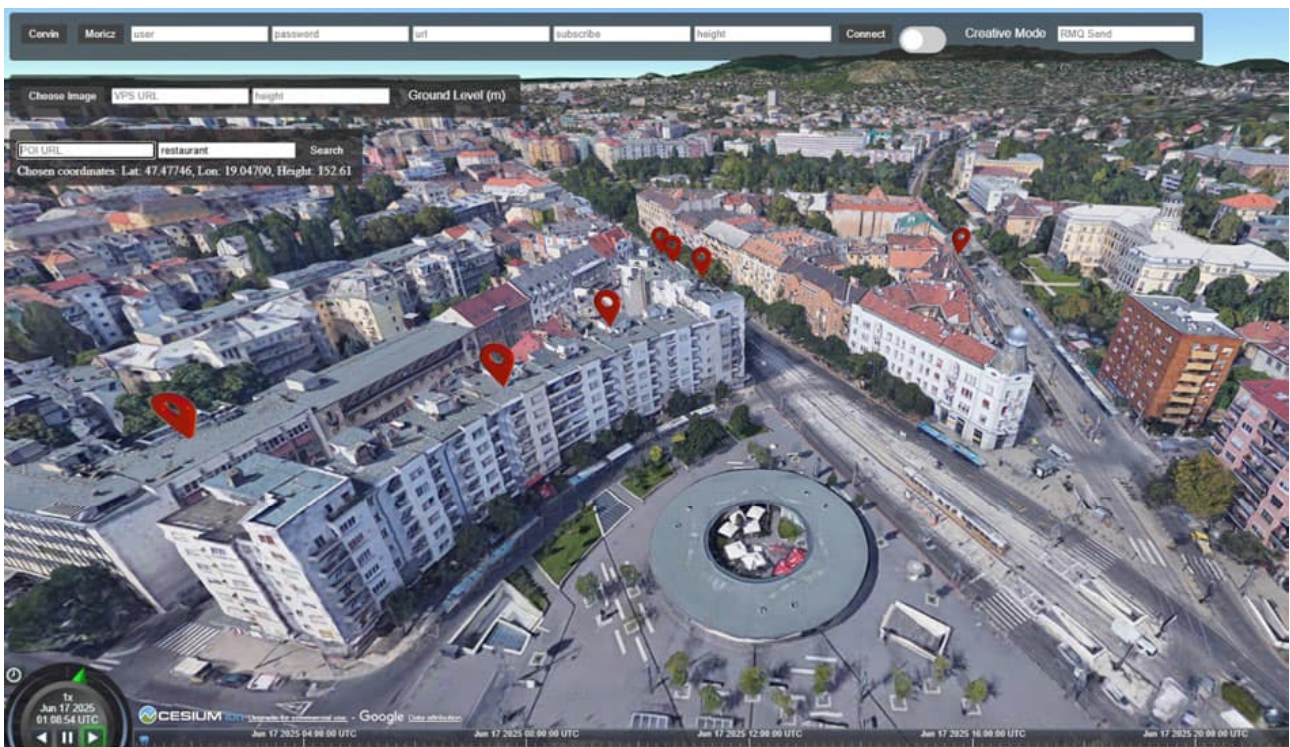


Figure 45. Our open globe viewer tool for developers of location-based AR offers visualization of various OGC GeoPose sources. We demonstrate the visualization of the results of an open visual positioning service, live pose streams from an open WebXR client, and points of interest pins retrieved by an open POI search service.

Abstract

We demonstrate an external globe viewer tool for developing location-based augmented reality. Our globe viewer can visualize camera GeoPose estimates from a visual positioning service, live GeoPose streams of AR clients, points of interest given in OGC POI format, and we can also share waypoints among the AR users. The tool is based on CesiumJS and shows Google's 3DTiles models of buildings, terrain, and the globe surface.

Introduction

Location-based augmented reality (AR) offers geographically anchored digital content blended into our physical world. Continuous contextual search discovers digital objects and points of interest (POIs) relevant at the user's current geo-location, and these digital assets appear within the user's view as if co-located with real-world elements.

Delivering geospatial AR requires several technologies: high-fidelity 3D reconstruction, real-time 6DoF user localisation, context-aware content discovery, semantic and/or photorealistic modelling, and interactive AR rendering. Currently, these functions exist within isolated, proprietary platforms with closed mapping and content pipelines, hindering cross-platform AR experiences and agent interoperability. Scalable, decentralised spatial computing requires open standards and interoperable systems to exchange and render AR assets using shared spatial references, such as user pose and location.

The Open AR Cloud Europe association advances this through the Open Spatial Computing Platform (OSCP) (Sörös et al., 2024), a modular architecture for interoperable spatial computing infrastructure. As part of the Open Source Scalable and Continuous Augmented Reality Services for Urban Spaces (OSCAR4US) project funded by the NGI Search, the project team extended the OSCP with an open-source visual positioning system (OpenVPS), an open-source POI search service, and a reality modeling language. In addition, we created a globe visualization tool for developers of location-based AR experiences, which we present in this paper. We extended a 3D geospatial viewer CesiumJS, with various OGC GeoPose (Open Geospatial Consortium) sources: we visualise VPS localization results, point of interest search results, as well as live GeoPose streams coming from our open WebXR client, together with waypoints shared between users. Our globe viewer enables live debugging of AR experiences by showing the AR users visible in geospatial coordinate space. Furthermore, the globe viewer will allow remote content authoring and remote testing without physical presence in the actual scene.

Related Work

Geospatial AR applications require a shared global coordinate frame, for example, the widely used WGS84. Bringing AR clients into the universal WGS84 coordinate system requires precise (typically visual) localization using both position and orientation (six degrees of freedom) and continuous onboard motion tracking. Mobile SDKs like ARKit (Apple) and ARCore (Google) provide robust motion tracking pipelines, and WebXR can even make this available in Web applications. Given the local and global pose pair at the time of (even as few as a single) localization, the client device can estimate its own GeoPose at camera frame rate, which enables interactive geospatial AR applications.

Geographic content anchoring solutions, including Niantic Lightship VPS, Google Geospatial Anchors, and Apple GeoAnchors enable content placement at geospatial coordinates. However, these

solutions are closed, lacking interoperability and composability, which prevents them from being integrated into a unified application or combined within a single AR session.

There are multiple existing globe visualization tools available for free (Google Earth, CesiumJS, HERE Maps, etc), and Google's one is even connected with their AR solution stack via the Geospatial API. We are targeting a viewer that is fully open and can be extended by developers and creators.

Globe view of the estimated camera GeoPose

As part of the NGI Search project (oscar4us.eu), we have created an Open Visual Positioning System (VPS) which can be tested using our Globe viewer (see **Figure 46**). The OpenVPS requires only a single photo for localization, and it returns the position and the orientation where the photo was taken. When the result is displayed, the viewer enters a first-person view of the location where the picture was taken. Without precise geodetic measurement equipment, we have no way to benchmark the accuracy of the OpenVPS, but in practice, the geo-anchored contents in our globe and AR viewers appear close to where they should be. The estimated error should be below 1 meter, but it depends primarily on the quality of the visual features in the scene and the accuracy of the map's geo-registration.

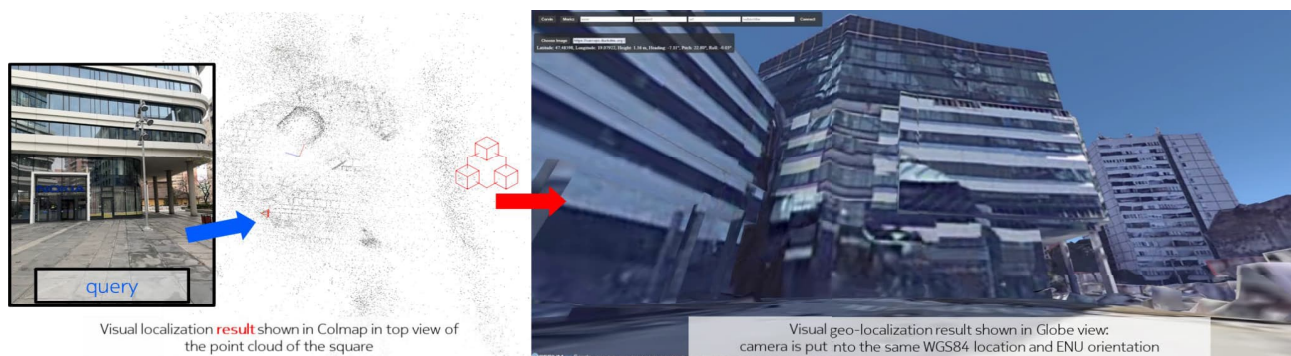


Figure 46. Left: localization query. Middle: First-person view of the estimated camera pose. The globe view is navigated to the estimated camera pose. In an ideal case (perfect localization and assuming a fresh 3D model), the 3D model depicts what is in the query.

It is important to note that the image needs to contain EXIF data regarding the camera's parameters. Since it is not (yet) possible to decide the height of a point using the Google 3D Tile set, we have added a height parameter. To augment the result from the OpenVPS with real altitude instead of height w.r.t the ground, we also need to set the ground altitude. A good starting point is to click on the ground, and the interface will display an approximation of the altitude at that position. However, this works only if the location is within the frame.

Globe view of user pose streams

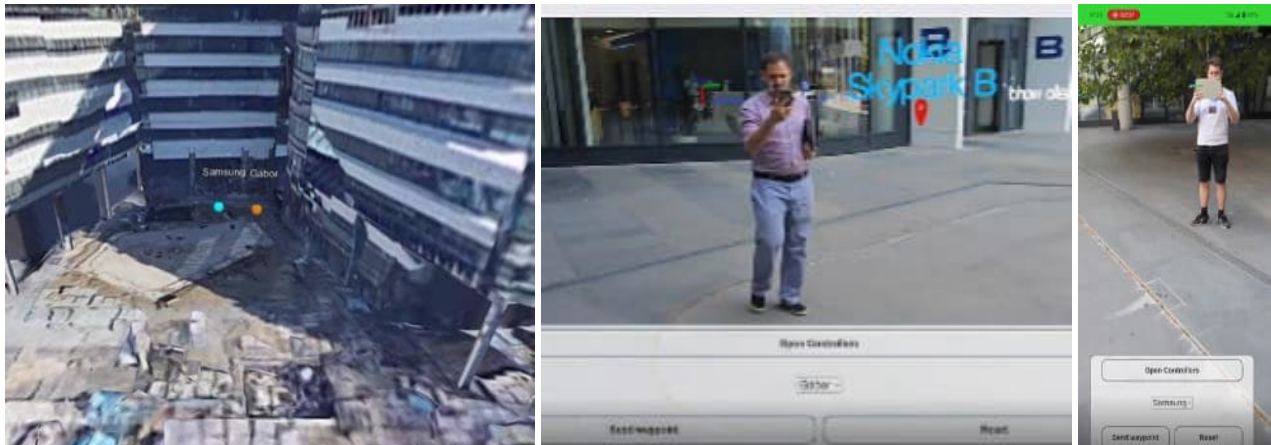


Figure 47. The GeoPose streams of two users are visualised in our globe view (left) and in mobile AR sessions via small RGB coordinate axes. Middle: AR view from device 1 (blue) Right: AR view from device 2 (orange).

In the globe viewer, the movements of smartphone AR users are visible (see Figure 47). Our open-source WebXR client can connect to a RabbitMQ message broker to share and stream the estimated GeoPose at every frame after visual localization. The globe viewer can display any GeoPose and avatar coming from the message broker in the following JSON format:

```
{
  "agent_id": "afd403d0-af55-4e4b-92b3-b42e42313437",
  "avatar": { "name": "Gabor", "color": { "r": 255, "g": 195, "b": 0, "a": 1 } },
  "geopose": {
    "position": { "lat": 47.485999, "lon": 19.074005, "h": 1.269483 },
    "quaternion": { "x": 0.120613, "y": 0.623825, "z": -0.084652, "w": 0.767547 }
  },
  "timestamp": 1747658972982
}
```

Globe view of points of interest

In the viewer, we can visualise the results from the open OSCP Point of Interest (PoI) collection service. The user can provide the area of the search, the URL of the service, and the keyword to be searched. The user can click on the ground to set the latitude and longitude as the center of the search circle with a radius of 200m. The search results are returned in OGC POI format, and the pins of POIs related to the search keyword will be added to the globe view. We chose this search radius because if we increase the search area, the content can quickly become crowded. Figure 1 illustrates all *restaurants* retrieved from the PoI collection service.

Sending waypoints to AR users

It is not only possible to stream the GeoPose from mobile clients to the globe viewer, but we can also send waypoints between clients. The developer/debugger can also send waypoints from the globe viewer to WebXR clients that then appear in the AR session(s) at the specified GeoPose (see Figure 48). Waypoints are also exchanged via the message broker.

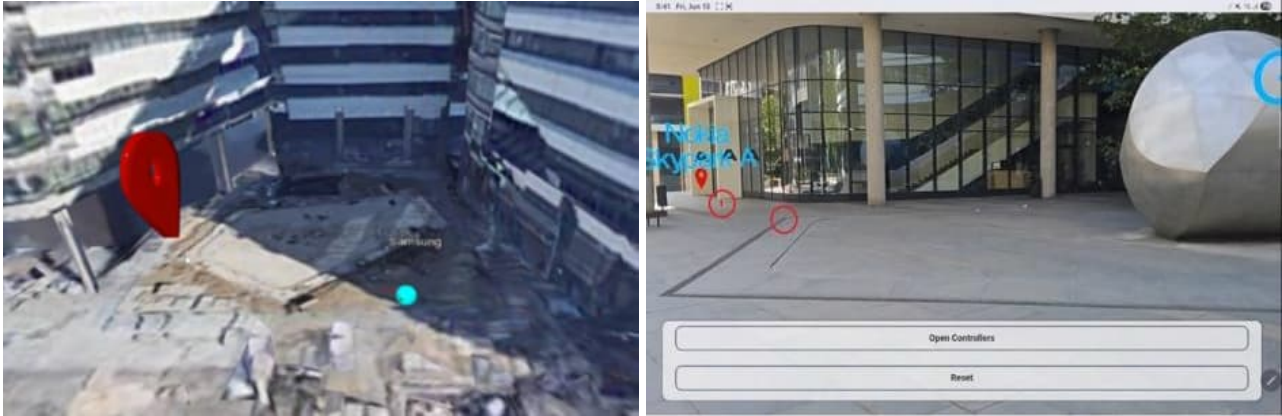


Figure 48 Sending waypoints from the globe viewer directly into a WebXR session on a mobile client. Left: clicked waypoint given by GeoPose. Right: Mobile AR client view of received waypoint (red stick in circle) and the empty circle marks where it should appear.

Demonstration

We will demonstrate all features of the Cesium viewer, emphasizing its rapid deployability at any new location. The demonstration will include (as illustrated above) user geo-localisation using our open-source Visual Positioning System (VPS), discovery of nearby points of interest (POIs) with our open-source POI search service, and sharing the user poses remotely and on location. The globe viewer requires no dedicated application installation and operates directly in a Web browser. We also note that our mobile AR client can run any WebXR-compatible browser. The source code of the globe viewer: <https://github.com/openarcloud/cesium-viewer>

References

- Open Geospatial Consortium – GeoPose standard, <https://www.ogc/standards/geopose>
- Sörös, G., Jackson, J., Vogt, M., Salazar, M., Kadlubsky, A., Vinje, J-E. 2024. An Open Spatial Computing Platform. In IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom), Hong Kong

SITE-VR: A Framework for Improving User Experiences in Immersive Training

Luciana Nedel¹, Anderson Maciel^{1,2}, Carla M.D.S. Freitas¹, Leonardo Guths¹, Lucas Pereira¹, Matheus Negrão¹, Rafael Torchelsen^{1,3}, Wesley Ferreira¹,
and Luciana Berretta⁴

¹*Akcit Lab, Federal University of Rio Grande do Sul, Brazil*

²*Instituto Superior Técnico, University of Lisbon, Portugal*

³*Federal University of Pelotas, Brazil*

⁴*Akcit Lab, Federal University of Goiás, Brazil*

Corresponding author: nedel@inf.ufrgs.br

Keywords: immersive training, education, 3D user interfaces

Introduction

Immersive simulators are natural tools to practice complex interactive tasks, being particularly helpful in education and training. It is widely recognized that effective learning occurs when the senses are stimulated (Laird, 1985). Previous works (Menin et al., 2018) demonstrated that, in addition to saving resources and providing safe training environments, VR promotes knowledge acquisition as efficiently as conventional methods, with the added benefits of improving learning retention and increasing user engagement. In this context, SITE-VR (Simulation, Immersive Training, and Education in Virtual Reality) is an under-development framework designed to improve immersive simulators by addressing challenges such as reducing cognitive overload during interactions, increasing the sense of presence, and understanding the factors influencing user adoption. Our goals are to propose and evaluate interaction techniques that facilitate selection, manipulation, and navigation tasks in immersive virtual environments; to apply and test the proposed interaction techniques and modalities in immersive simulation environments to validate their applicability, effectiveness, and user acceptance; and to develop an understanding of the usability of immersive simulators and their adoption by target users.

Methodology

The framework development relies on studying the users' needs to identify problems and areas for improvement in existing solutions. Therefore, the innovations proposed are, whenever possible, integrated into existing simulators so we can test them with target users. In this paper, we share our recent developments and experiences on building interfaces for a colonoscopy surgery simulator (Martins *et al.*, 2024), and a clinical consultation simulator (Negrão *et al.*, 2023). While modeling complex virtual environments is a challenge *per se*, we will focus exclusively on the user-interaction aspects.

Interaction in training immersive environments is naturally 3D. Recent VR hardware allows hand and head tracking natively, reducing the need for complex mapping and cognitively demanding metaphors that were common in the early days of 3DUI. Nevertheless, several problems remain active, and the problem space varies significantly between different simulators. Surgical simulators often require little to no navigation but are demanding in terms of faithful manipulation. Clinical simulation, in turn, is less demanding in terms of hand-eye coordination but requires navigation around the office or hospital and realistic communicative virtual humans.

Manipulation in a colonoscopy simulator

In a colonoscopy, the surgeon bimanually manipulates an endoscope. One hand guides the insertion section through the rectum, while the other handles the control section. The two sections are linked in such a way that the rotation of one implies the simultaneous rotation of the other. The control section contains knobs to rotate the camera and light in the insertion section, and buttons to activate water and gas injection. The technique the surgeon uses to manipulate the insertion part requires individual finger functions to grasp and bend the tube.

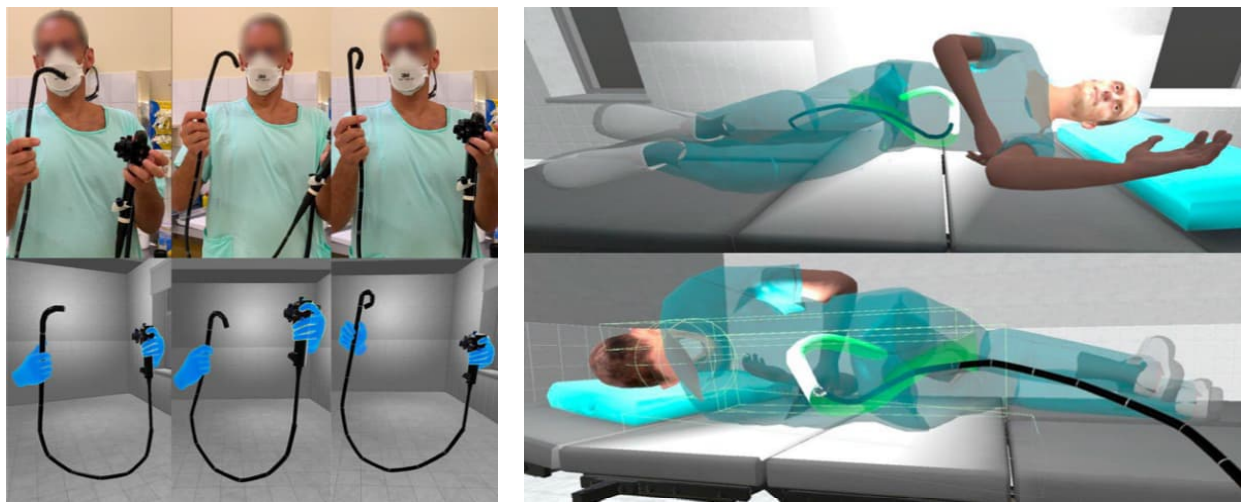


Figure 49. Bimanual asymmetric interaction with an endoscope.

Our solution for this problem is an asymmetric interface combining a number of constraints. This interface was co-designed with gastroenterologists (Figure 49). The first decision was to use the VR controller on the left hand and hand tracking on the right hand. The controller provides haptics, offering security, and allows mapping the controller's thumbstick to knobs and the buttons to other functions. For the right hand, closing the pinch (opposing the thumb to the other fingers) activates grasping, while the index finger is free to guide and help bend the tube for maneuvering. To implement these interactions, we leverage components of a game engine, Unity, for instance. However, while the simulator requires interaction to be sent as external forces, hand tracking is kinematic. To convert hand motion into forces, we have built a twin model of the endoscope with Unity components. To prevent extensive hand motions from breaking the model, a decoupling of the physical and the virtual hand is implemented using a god object, i.e., as the real hand moves away, the virtual hand remains on the endoscope.

Navigation in a hospital environment

In VR navigation, natural walking is hindered by the disparity between the vastness of virtual environments and the limitations of the user's real physical space. To overcome this difference, methods

like teleportation and continuous directional movement, inspired by established video game interactions (Cruz-Neira et al., 2018; Cmentowski et al., 2019), are commonly employed. However, continuous directional movement is frequently linked to cybersickness, and teleportation may fragment the sense of presence and not always guarantees a suitable positioning and orientation for the task. Users who try to compensate for this inaccuracy by physically walking in their real space may collide with (or pass through) the 'guardian' system, breaking immersion and hindering task completion.

We introduced a new locomotion approach to improve navigation in VEs considerably larger than the available physical space. It leverages the available physical space for increased real movement and uses teleportation for longer distances (see Figure 50). To achieve this, we introduce the concept of S-zones, which are pre-defined safe zones that encompass disjoint sets of interactive elements. This approach encourages and allows users to physically move naturally within these S-zones, reserving teleportation for reaching more distant areas. This minimizes short range teleportation, reducing spatial disorientation.

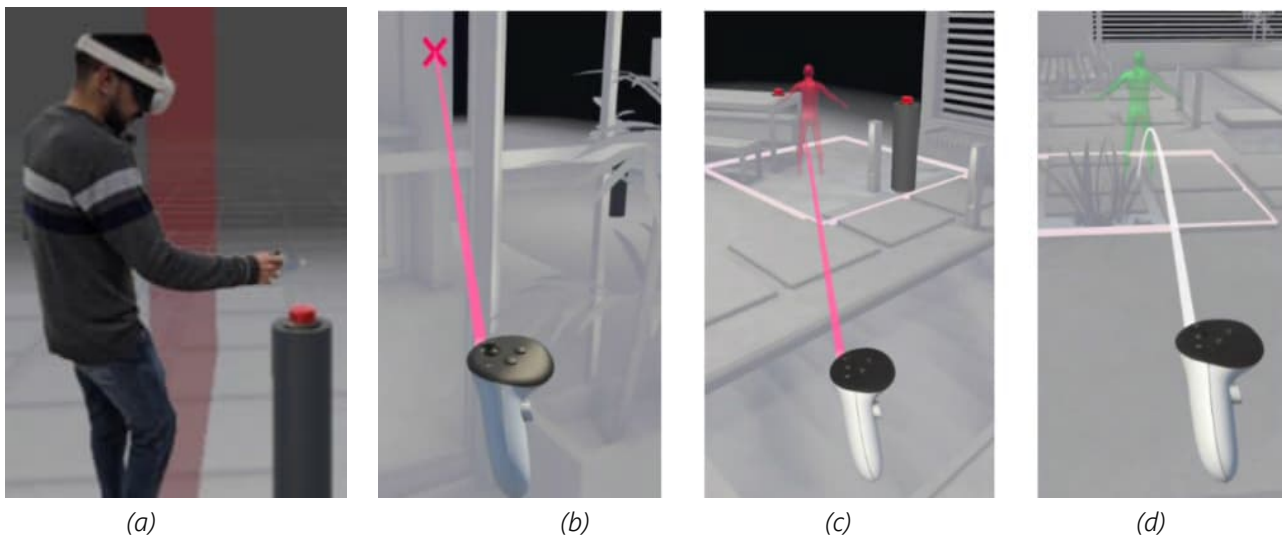


Figure 50. (a) A user trying to press a button that is outside his safety limits (indicated by the red plane); (b) impossible teleportation due to a wall obstacle; (c) invalid destination position due to a collision with an object in the target S-zone; (d) user's spatial orientation in the current and destination S-zone.

Communication with virtual humans

The use of autonomous virtual humans in immersive simulators is a key factor for presence. With the emergence of large language models (LLMs), the use of artificial intelligence-driven agents have gained prominence. We explored the user's perception in conversations with an agent driven by a large language model in comparison with a real human through a wizard of Oz approach. A synthesized voice is used in both cases for fairness. We developed a doctor consultation scenario for experimentation (see Figure 51). The LLM and a human participant alternately simulate the role of a patient, while the subject user always takes the role of the physician. We conducted a user experiment with doctors and medical students to assess their reactions to each condition, focusing on user perceptions of realism, empathy, and utility. The findings suggest that LLM-based virtual patients hold strong potential for scalable clinical training in terms of verbal capability. However, improvements in expressiveness (emotions and body language) and interaction flow are essential for broader adoption.



Figure 51. Reconstruction of a medical consultation room presenting the environment and position of the subject user and the virtual patient, and appearance of the avatars. Lip-synch and idle movements are activated.



Figure 52 A user visualizing and annotating a previously recorded simulated training session in a doctor's office. Besides annotations, temporal navigation, path travelled, and avatar with focused field of view are shown.

Debriefing

Training skills, either in the real world or in VR, require supervision by an expert or teacher who will identify errors and provide feedback. A common instrument for this is the debriefing sessions, carried out just after the task is performed by the trainee. We developed an immersive visualization-based debriefing method for capturing, representing, annotating and replaying VR training sessions. It collects, stores, and reconstructs user interaction data, recording hand movements, 6-DoF positioning, and object manipulation. Unlike traditional approaches, restricted to 2D video reviews (passive, perspective-limited) or instructor-led feedback (subjective, attention-dependent), our system enables standardized, large-scale training analysis. Instructors and trainees review sessions immersively using interactive controls to analyze procedural steps, clinical decisions, and deviations from protocol. Figure 52 illustrates a debriefing session. An exploratory study in a clinical scenario has shown that the spatially grounded replay enhances perceived feedback precision and can improve training effectiveness compared to conventional methods.

Final comments and next steps

The lessons we learned so far while testing immersive simulators with target users allowed us to redesign the interfaces and improve the user experience. In surgery simulation, the new methods allowed for a more ecological interface with the virtual endoscope without the need to a physical endoscope. In the clinical simulation, the debriefing tool facilitated detailed self-observation and guided reflection. It effectively fostered greater metacognitive awareness, improved perceived learning efficiency, and increased participant confidence. The locomotion in the hospital space using teleport with a predictive projection was effective for anticipating position and orientation in the destination location, being highly rated in pilot experiments. We measured a strong correlation between immersion, experience, and the practical use of zones and teleportation, as the use of pre-defined safe zones encourages natural body movement. This reinforces the degree of freedom and safety, especially in physically limited spaces. Finally, the use of virtual agents powered by LLMs in simulating patients for the doctor's office provides a flexible and scalable alternative to scripted or piloted agents. Our initial study also pointed the way to new developments to include additional non-verbal cues in the setup. Summative evaluation with critical users is an ongoing task.

Acknowledgements

This work has been partially funded by the project SITE-VR: A Framework for Improving User Experiences in Immersive Training, supported by the Advanced Knowledge Center in Immersive Technologies (AKCIT), with financial resources from the PPI IoT/Manufatura 4.0 / PPI HardwareBR of the MCTI grant number 057/2023, signed with EMBRAPPI. Thanks are also due to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) through grant 421962/2023-2, and to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

- Cmentowski, S., Krekhov, A., & Krüger, J. 2019. Outstanding: A multi-perspective travel approach for virtual reality games. In Proceedings of the annual symposium on computer-human interaction in play (pp. 287-299). <https://doi.org/10.1145/3311350.3347183>
- Cruz-Neira, C., Fernández, M., & Portalés, C. 2018. Virtual Reality and Games. *Multimodal Technologies and Interaction*, 2(1), 8. <https://doi.org/10.3390/mti2010008>
- Laird, D. 1985. *Approaches to training and development*, Addison-Wesley. Reading, Mass.
- Menin, A., Torchelsen, R., & Nedel, L. 2018. An analysis of VR technology used in immersive simulations with a serious game perspective. *IEEE computer graphics and applications*, 38(2), 57-73. <https://doi.org/10.1109/MCG.2018.021951633>
- Martins, M., Morais, L., Torchelsen, R., Nedel, L., & Maciel, A. 2024. Efficient Position-Based Deformable Colon Modeling for Endoscopic Procedures Simulation. In *ACM SIGGRAPH 2024 Conference Papers* (pp. 1-10). <https://doi.org/10.1145/3641519.3657454>
- Negrão, M. D., Ferreira, W., Bohrer, B., Freitas, C. M., Maciel, A., & Nedel, L. 2023. Design and think-aloud study of an immersive interface for training health professionals in clinical skills. In Proceedings of the 25th symposium on virtual and augmented reality (pp. 157-165). <https://doi.org/10.1145/3625008.3625037>

Validating an Immersive Virtual Environment for Learning 3D Transformations: From Design to Student Impact to Expert Evaluation

Maha Alobaid, and Michael Manzke

School of Computer Science and Statistics, Trinity College Dublin, Ireland

Corresponding author: alobaidm@tcd.ie

Keywords: Computer Graphics, 3D Transformations, Virtual Reality, Immersive Virtual Environment, Spatial Skills

Abstract

3D transformations are essential in computer graphics but remain difficult for students to master due to their abstract and spatially complex nature. This work presents a three-phase research study on the design, implementation, and validation of an immersive virtual environment that supports the learning of 3D transformations through direct interaction and real-time visual feedback with 3D objects. The study began with the design and pilot study of the tool, followed by a main study that assessed its impact on spatial skills and conceptual understanding among computer science students. To further validate its educational value, a survey was conducted with university educators experienced in teaching computer graphics and related disciplines. Findings across all studies indicate that the immersive environment is effective and engaging for supporting students' understanding of 3D transformations. This work contributes to the growing field of XR in computing education and offers evidence-based insights for future integration into teaching and learning practices.

Introduction

In computer graphics education, 3D transformations form a critical foundation for understanding how objects are visualized and manipulated in three-dimensional space. Despite their importance, these topics are often challenging for students to learn due to their abstract representations and the need for strong spatial reasoning skills (Suselo et al., 2022b). Traditional teaching approaches, which typically rely on static 2D visuals and textbook explanations, can make it difficult for learners to mentally visualize the effects of transformations in a 3D environment. Recent advancements in technologies, particularly virtual reality (VR), offer new possibilities for addressing these learning challenges. Immersive environments allow learners to interact directly with 3D objects (Korkut and Surer, 2023), observe the immediate effects of transformations from multiple perspectives, helping bridge the gap between abstract theory and intuitive spatial understanding. In this context, we developed an immersive virtual environment to support the teaching and learning of 3D transformations. This work presents the

development of the immersive environment, examines its impact on student learning, and explores its validation through expert feedback, positioning the tool as a pedagogically sound and practically relevant solution for XR-supported computer graphics education.

Immersive Environment Design

The immersive learning environment was developed using Unity, C#, and OpenXR to ensure compatibility with a range of VR headsets. It was specifically designed to address the challenges students face in understanding visual representations and OpenGL commands while learning 3D transformations, by offering an interactive environment where users can manipulate 3D objects and receive immediate feedback. A key feature of the environment is its ability to trace the object's state before applying a transformation, allowing users to observe updated positions based on their input (as shown in Figure 53). Users can apply translation, rotation, and scaling individually or in sequence to explore their effects on object position and orientation. The system also integrates real-time OpenGL transformation commands, which dynamically update as users apply changes to objects in the scene (see Figure 2). This dual feedback (visual and textual) was intended to help students form clearer connections between abstract operations and their visual effects in a three-dimensional space. Users interact with the system through VR controllers, enabling them to apply transformations and observe the results from multiple angles. The environment also allows users to walk or fly through the 3D space, providing flexible movement for exploring objects from different perspectives. The interface was designed to reduce unnecessary complexity and cognitive load. The design is guided by the Cognitive Theory of Multimedia Learning (CTML), which emphasizes that learning is more effective when visual representations are paired with relevant textual information (Mayer, 2005). By aligning spatial interactions with immediate visual and code-based feedback, the environment helps students build mental models of 3D transformation processes in a meaningful and engaging way (Alobaid and Manzke, 2024).

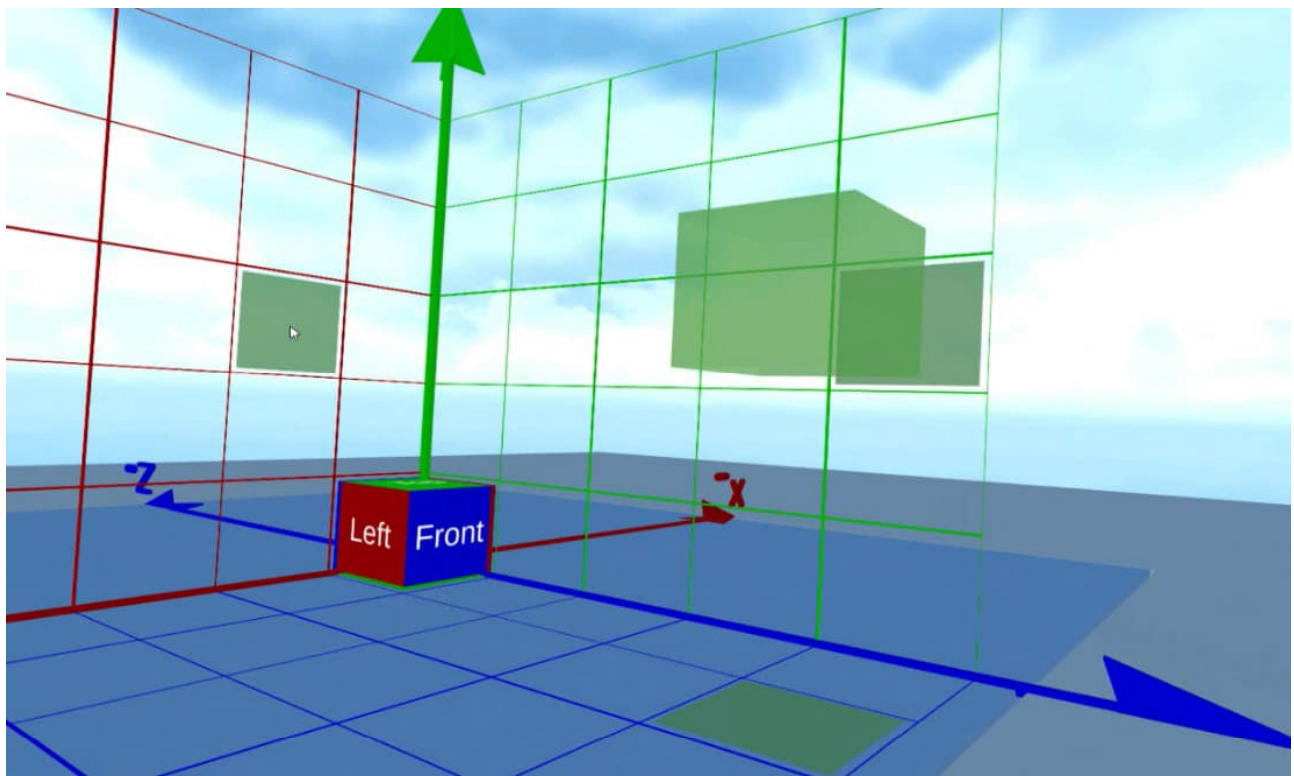


Figure 53 The trace feature to visualize the user's input.

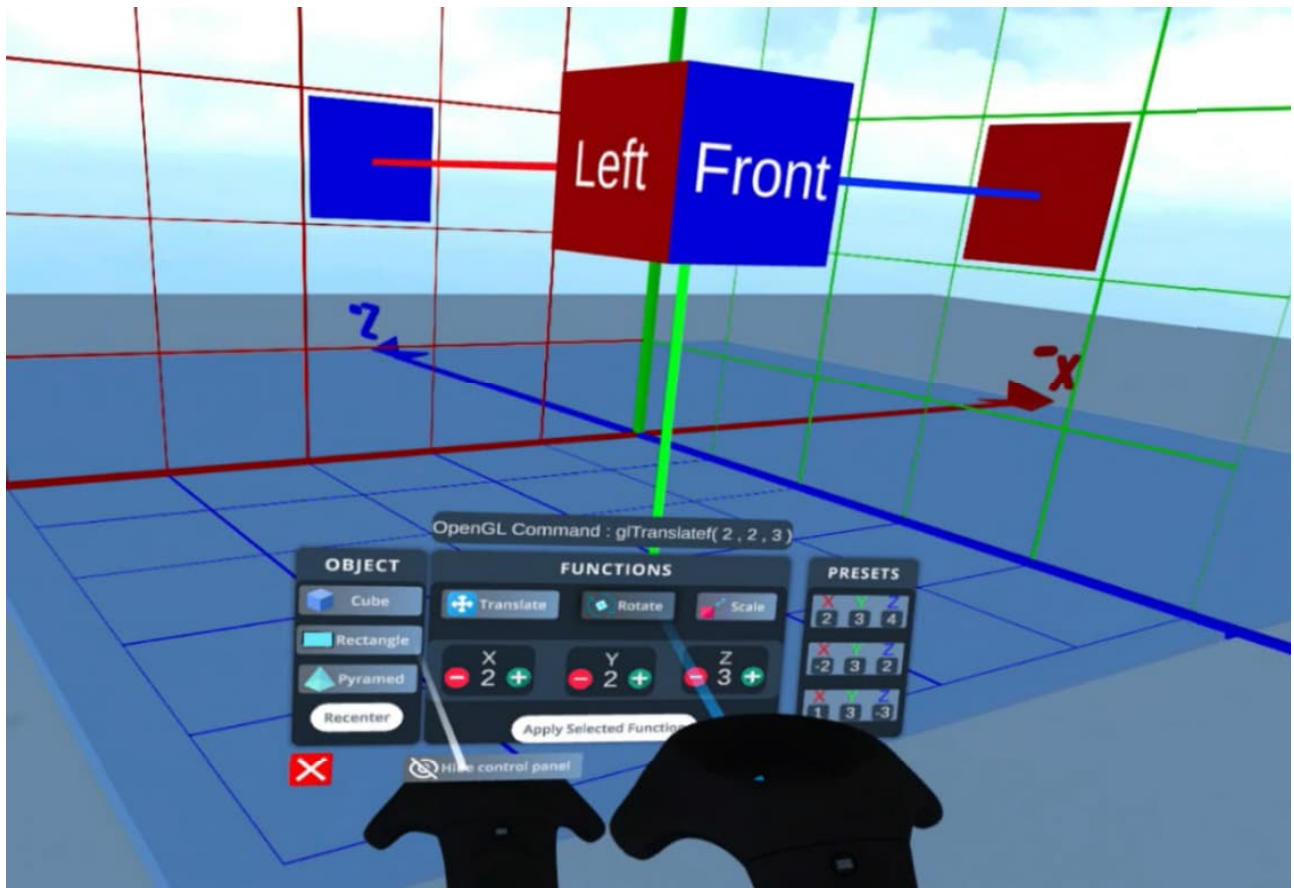


Figure 54. Translation function with real-time OpenGL command in the VR environment.

Pilot Study

The first phase of evaluation involved a pilot study aimed at exploring the usability and initial educational potential of the VR environment. Participants included seven users with extensive experience in VR and computer graphics. After a brief orientation session, they interacted with the environment and completed tasks involving 3D transformations without any time constraints. Feedback was gathered through post-session questionnaires. Findings indicated that participants found the environment intuitive and engaging. Most reported that the immersive setting helped them visualize how transformations affect objects in 3D space. At this stage, usability issues and areas for improvement were identified to enhance the environment for the main study.

Main Study

Building on the pilot, a larger lab controlled study was conducted to measure the effectiveness of the VR tool. The study involved 49 undergraduate and postgraduate computer science students who completed pre- and post-tests measuring their spatial reasoning and conceptual understanding of 3D transformations. They also completed surveys assessing engagement and perceived learning educational effectiveness. Results showed statistically significant improvements in spatial skills and contributed to deeper conceptual understanding after using the tool. Students expressed high levels of satisfaction with the VR experience and highlighted the value of being able to see and interact with transformations directly from different perspectives (Alobaid et al., 2025).

Expert Evaluation

To further evaluate the tool's pedagogical relevance from an expert perspective, a survey was conducted with 16 university educators from the Department of Computer Science. Participants included Professors, Associate Professors, Assistant Professors, and Lecturers, all of whom had varying years of teaching experience and expertise in computer graphics, programming, or VR. Experts first interacted with the immersive environment, then completed a questionnaire assessing its educational value, usability, and potential to enhance spatial ability in learning 3D transformations. Overall feedback was highly positive and strongly agreed that the VR tool supports learning by enabling students to directly engage with 3D transformations and improve spatial visualization. Several experts emphasized its potential impact and recommending its adoption as part of computer graphics courses at the college and suggestions to improve usability for novice students and sensitivity settings of the VR controller.

Conclusion and Future Work

This work presents the development and validation of an immersive VR environment designed to enhance the learning of 3D transformations in computer graphics. Across three studies a pilot, a main student evaluation, and an expert survey—the tool has demonstrated strong potential to improve student learning, engagement, and spatial reasoning. Future work will build on expert feedback by enhancing usability and further integrating the tool into computer graphics modules. An open-source version of the tool is also planned to support broader use and collaborative refinement in the computing education community. Additionally, further studies should be conducted to compare this approach with traditional teaching tools and to evaluate the long-term retention of learning outcomes, which would provide deeper insight into the educational effectiveness of the immersive environment.

References

- Suselo, T., Wünsche, B.C. and Luxton-Reilly, A. 2022. Why are 3d transformations in computer graphics difficult? an analysis of a decade of exam questions. In Proceedings of the 24th Australasian Computing Education Conference p. 181–190.
- Korkut, E. H., & Surer, E. 2023. Visualization in virtual reality: a systematic review. *Virtual Reality*, 27(2), 1447– 1480.
- Mayer, R.E., 2005. Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 41(1), 31– 48.
- Alobaid, M., & Manzke, M. 2024. Work-in-Progress—Design of Immersive Virtual Reality Environment for Learning 3D Transformations. *Immersive Learning Research-Academic*, 96-102.
- Alobaid, M., Young, G. and Manzke, M. 2025. Immersive Virtual Reality for Developing Spatial Skills in Learning 3D Transformations in Computer Graphics. *The Eurographics Association*.

Digital Characters Augmenting Cultural Spaces through Handheld and Head-Worn Devices

Spyros Vosinakis, and Athina Bosta

Department of Product & Systems Design Eng, University of the Aegean, Greece

Corresponding author: spyrosv@aegean.gr

Keywords: mixed reality, cultural heritage, physiological measures, digital characters, HoloLens

Introduction

Augmented reality (AR) technology can be beneficial for the dissemination of cultural heritage and recently there has been a notable tendency in utilizing it to enhance the experience of visitors in museums and cultural spaces. A special case of content type in cultural AR applications are animated digital characters, which can bring important gains to the visiting experience (Rizvic et al, 2019; Wang et al, 2019). Nevertheless, the generation of AR cultural experiences with animated digital characters is not without design and technological challenges. The impact of these experiences on visitors in terms of engagement and cultural presence have not been studied extensively, whilst there is little knowledge on how the different AR display devices impact the performance of the application (Dima, 2022; Spierling, 2017). In an attempt to shed more light on these challenges we have designed and implemented an AR experience for the presentation of a cultural space, and we studied the user experience and knowledge gains under two different conditions: a handheld tablet device and Mixed Reality glasses (HoloLens).

The AR experience takes place in a well-preserved mansion that contains a unique ceiling painting with many historic references. The mansion dates back to 1860 and is located in the city of Hermoupolis in Syros island, Greece. The aim of the application is to introduce visitors to the story of the mansion and, through it, the history of the city and the birth of the modern Greek state. We have set up a user study in order to comparatively evaluate the two candidate interaction devices (tablet and MR glasses) in terms of presence, user experience, preference and learning. We adopted a within-subjects design, where each of the participants experienced the interactive presentation on both devices. The application has been split into two parts (almost equal in length) and each user used one device for the first half and the other for the second.

Design and Development of the Mixed Reality Application

Two digital characters, the god Hermes and the Greek revolution hero Kolokotronis, guide the experience, "descending" from the ceiling painting in life-size form, interacting with physical elements in the mansion's living room (Figure 55). The 10-minute story unfolds in five segments, incorporating

interactions with objects like a portrait, radio, and piano, along with historical narration and humorous elements. The application integrates various design elements, including augmented 2D/3D content (images and musical notes), interactive elements prompting user movements and radio control, references to physical objects to enhance immersion, and surprise elements.



Figure 55. A capture of the two digital characters inside the mansion.

The application was developed in Unity, with characters modeled in Blender, based on the ceiling painting and historical artworks. Two actors provided voice and motion, captured in a studio using Vicon cameras and post-processed with Unity's UMotion Pro plugin to correct any erroneous moves and to create the impression that Hermes is flying. Salsa LipSynch plugin for Unity has been used for automatic lip syncing of the digital characters to the recorded voices. A digital twin of the actual room has been created in Unity and has been used for testing and refining the application before its actual deployment. The tablet version of the application is using Unity AR Foundation and runs only in ARCore enabled devices, whilst the Hololens version uses the MRTK library for the hand interactions.

User study with tablet and MR glasses

The study process was the following (Figure 56). Initially we welcomed participants, explained the procedure and collected demographics. Then, users were familiarized with the Hololens through a short demo (duration about 1 minute). Next, they experienced the first half of the application with one device and filled in the SUS Presence questionnaire afterwards. After that, they switched to the other device for the second part of the application and filled in the same questionnaire for the second condition. Next, they had to choose between one of the two devices as their preferred for cultural heritage presentation and had a brief interview with the investigators, where they were asked to comment on the positive and negative aspects of each device according to their experience. Finally, they had to fill in a form with six questions about the content of the presentation to test their learning effectiveness. Three of these questions were related to the first part of the application (and the first device, respectively) and the others to the second. During their experience in both conditions, users were

wearing a wrist sensor that was collecting physiological data. Our intention was to measure EDA (electrodermal activity) during each condition, to identify any important changes or peaks in attention. The evaluation took place in the physical space of the mansion and the devices used were: Xiaomi Tab 6 with the application running in landscape mode, and MS HoloLens 2. For the physiological data we used the Empatica EmbracePlus smartwatch.

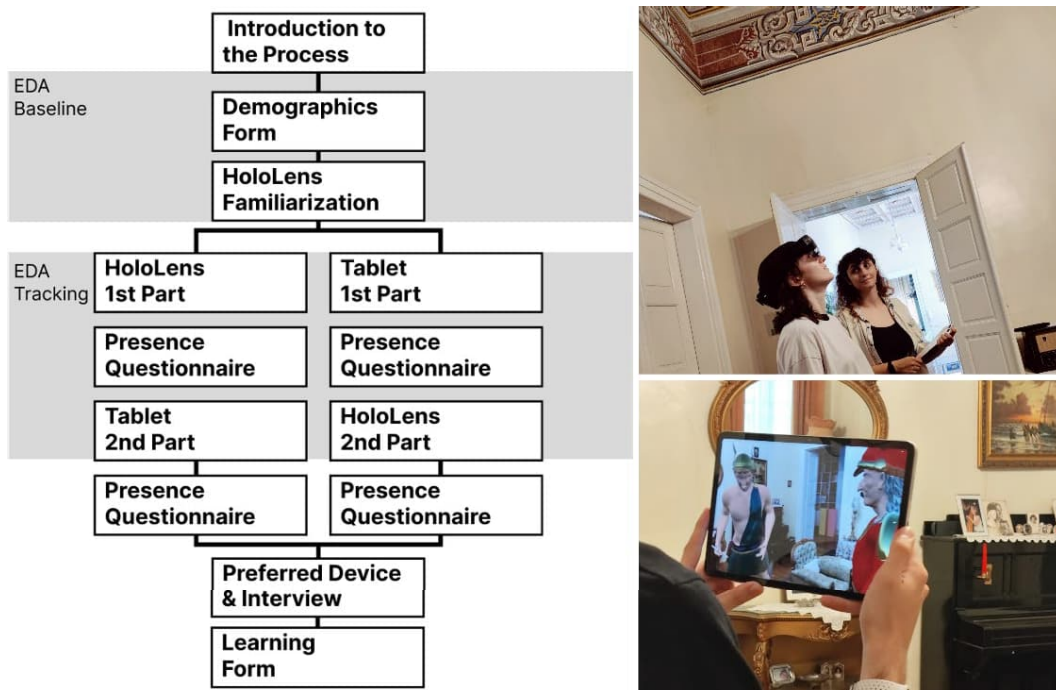


Figure 56. The study process (left) and photos of users during the study (right).

Twenty two (22) users participated in the study (female: 8, male: 14). They had a wide age distribution from 19 to 50 years old (18-19:1, 20-24:9, 25-29:1, 30-34:1, 35-39:1, >40:7) with considerable experience in video games (average: 3.5 on a scale 1 to 5) and use of XR environments (average: 2.9 on a scale 1 to 5). None of the participants had used a HoloLens device before. The study has been executed without any critical issues and all participants managed to complete the process.

The study did not find any important differences in terms of presence and learning. For each participant we calculated the average value of the 6 questions of SUS Presence questionnaire. The average of all participants for the HoloLens experience was 4.48 (on a scale from 1 to 7) and for the Tablet 4.27. These values indicate a medium to strong sense of presence in both conditions, without a statistically significant difference between them. On the learning part, it seems that the users could not easily absorb the information presented to them through the MR experience. On average they answered correctly less than half of the questions (2.81 out of 6). Again, there was no statistical significance in question scoring between the content presented in HoloLens (1.45 out of 3) and that presented on the tablet (1.36).

The results in terms of attention are mixed. We applied a median filter and a low-pass Butterworth filter on the raw EDA data to identify peaks, which are interpreted as moments of emotional intensity. These were detected in nine out of the 22 participants. It is noteworthy that 6 of the users whose EDA data produced one or more peaks claimed that they had no or very little experience with XR environments; probably an indication of the “wow” effect. Next, we calculated the average and the slope of the recorded EDA values during the two conditions. There were notable differences in the average values of users’ EDA between the two conditions, but without following a general trend towards one

device. Similarly, the slopes were mostly positive, indicating a possible rise of arousal during the experience, but not in favor of a specific device. The paired t-tests in both measures did not reveal significant differences.

In terms of preference, it seems that most participants were impressed by Hololens. Sixteen (16) users stated that they would prefer it as a device for MR presentations in cultural heritage, in contrast to the remaining six that found the tablet better. The interviews revealed further findings regarding the users' stance towards the devices. Most users expressed that using Hololens they felt like standing next to the characters, at life-size. Their navigating felt more relaxing and they liked that the device offered dynamic sound and stereoscopic viewing. On the other hand, some users found the colors 'duller', mentioning that the characters were presented as ghosts, and most of the users were irritated by the limited field of view of the device. Regarding the tablet experience, participants described it as the device they were most familiar with, making it easier to use. They commented that the resolution of graphics, character models and animation felt much better. For some users it was more natural to follow and explore the scene using the tablet, whilst for others it was difficult to notice the regions of interest, often missing some parts of the content. One user highlighted the fact that the tablet is an indirect way of looking at the scene, as they had to use their hands to point at the content.

Conclusions

The work demonstrated that AR experiences featuring digital characters have the capacity to enhance visitor engagement and presence in cultural spaces, irrespective of the device used. Although MR glasses provided a more immersive interaction, user familiarity and graphical clarity of tablets also offer notable benefits. These insights highlight the importance of carefully selecting interaction devices based on the intended application context and user group characteristics to maximize cultural learning and engagement.

Acknowledgements

This work has received funding from HORIZON-CL2-2022-HERITAGE-01 (Grant Agreement Number 101094998) under the HeritACT project.

References

- Dima, M. 2022. A design framework for smart glass augmented reality experiences in heritage sites. *Journal on Computing and Cultural Heritage (JOCCH)*, 15(4), 1-19. <https://doi.org/10.1145/3490393>
- Rizvic, S., Boskovic, D., Bruno, F., Petriaggi, B. D., Sljivo, S., & Cozza, M. 2019. Actors in VR storytelling. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)* (pp. 1-8). IEEE.
- Spierling, U., Winzer, P., & Massarczyk, E. 2017. Experiencing the presence of historical stories with location-based augmented reality. In *International conference on interactive digital storytelling* (pp. 49-62). Cham: Springer International Publishing.
- Wang, I., Smith, J., & Ruiz, J. 2019. Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Posters

HOLOVER-5G: Industrial Trainings Integrating Spatial Gaussian Splats with Real-Time 3D Holograms over 5G

Simon N.B. Gunkel¹, Tessa Klunder¹, Alessio Antoniacomi², Glauco Miniussi²,
and Gianluca Cernigliaro¹

¹TNO, The Netherlands

²OVER, Estonia

Corresponding author: Simon.Gunkel@tno.nl

Keywords: AR, MR, training and education at a distance, 3D representation

Abstract

In this (poster) paper, we outline our current work on HOLOVER-5G a scalable holographic communication pipeline deployed in 5G infrastructure and a commercial application. Our work integrates photo realistic 3D communication with environments as Gaussian splats to create high-quality training scenarios that are fully photo realistic and provide natural interaction.

Introduction

Next-generation eXtended Reality (XR), is enabling a revolution in human-centered applications, providing the possibility of enhancing shared digital experiences to become intuitive and realistic as never before. **Holographic communications** (Social XR) are increasing their level of realism and seamlessness, narrowing the gap between the virtual and the real, but encountering numerous challenges related to compression and transmission **with limited real-time capabilities** and **lack of scalable multi-user solutions**. Modern networking technologies such as 5G and 6G are increasing the confidence in the adoption and broad distribution of applications, such as volumetric video processing, especially when real-time is a requirement. HOLOVER-5G aims to provide the key technological contribution to reducing this gap, thanks to the introduction of the **Holographic Multipoint Control Unit (Holo-MCU)**, a virtualized optimization tool capable of i) offloading heavy processing from the client side; ii) maximizing the number of users in the remote experience; iii) providing elasticity to reconfigure and adapt the available network resources. Our work is based on / addresses the following 3 limitations.

Limitation 1, **Human presence**: Bringing real people in XR is key for realistic interactions. Remote XR experiences are normally based on virtual avatars that cannot adapt over time. Holograms provide this capability when dealing with natural content. With HOLOVER-5G, real humans remotely located will be

in XR experiences. In particular in a specifically developed scenario for training in the manufacturing industry⁹, allowing remote interaction, joint assets manipulation and a realistic feeling of being there.

Limitation 2, **Real-time**: holographic communications face challenges due to the large volumes of data (Viola et al., 2023) to be processed and solutions vulnerable to real-time constraints (Graziosi et al., 2020). HOLOVER-5G enables real-time communication between remote humans (Gunkel et al., 2023b) and advances the state of the art for live holographic communications with a network optimization solution (Gunkel et al., 2021).

Limitation 3, **Large-scale scalability**: Current pipelines address mostly peer-to-peer use cases (Jansen et al., 2022) without considering multiuser experiences, precluding the reaching of a bigger audience. We aim to overcome this lack applying multi-user optimization strategies to the volumetric video domain, exploiting 5G capabilities like stable mobile network connections and edge processing capabilities.

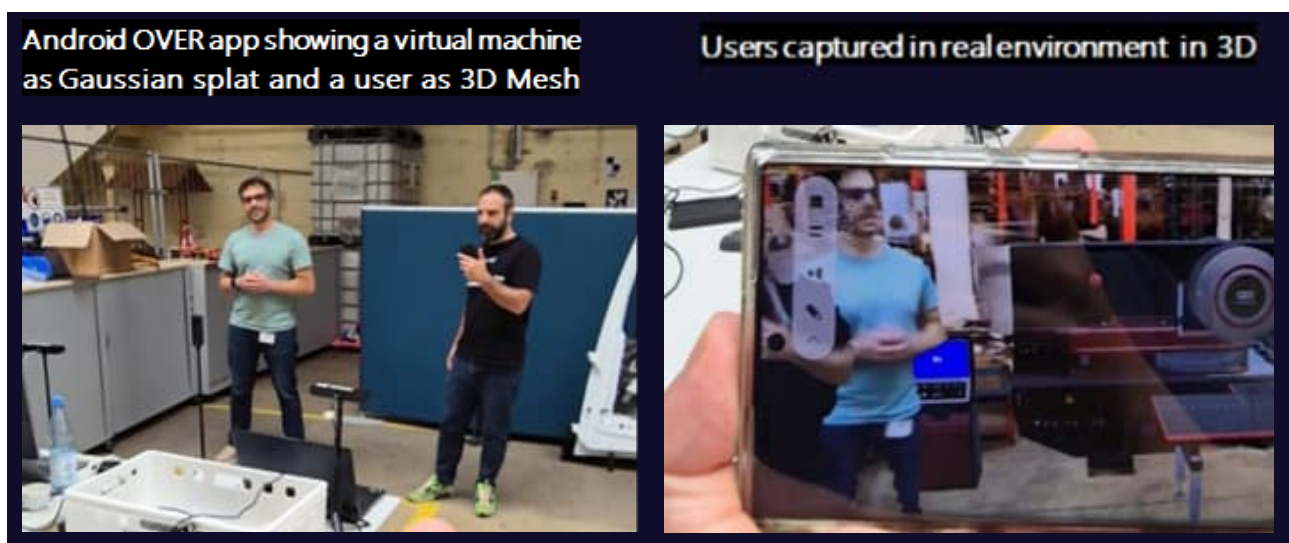


Figure P 1. Initial demonstration of the use case in action (left: real environment, right: virtual mobile app)

Use case

The main objective of HOLOVER-5G is to provide holographic training simulation within a highly realistic industrial environment (see Figure 1). At its core is a **Gaussian Splatting-based** Digital Twin of a factory, providing an **immersive** reconstruction of the facility. Within this environment, a 3D machine asset seamlessly transitions between multiple operational states, normal operation, error mode, and recovery, triggered by specific user actions. The **choice to include a static 3D asset with animated sequences** serves to further enliven the environment, making the demonstration **more engaging** and **immersive** for participants. By overlaying a holographic trainer onto the Digital Twin, users can observe machinery processes up close, practice guided troubleshooting, and refine their skills in a risk-free virtual setting that closely mirrors real-world conditions. Moreover, the scenario is generated with low complexity, allowing portability in multiple devices (i.e. smartphones or XR displays and XR glasses).

⁹ <https://doiotfieldlab.tudelftcampus.nl/virtual-training-in-the-manufacturing-industry/>

Technology

HOLOVER-5G proposes a complete end-to-end pipeline, as shown in Figure . We use a capture module as presented by Gunkel et al. (2021), per user, a singular RGBD (color and depth) capture sensor (ZED 2i) captures the user (and or objects) at color and depth, does some image optimizations (like cleaning and hole filling) and converts the 16bit depth values into our own greyscale color format (Gunkel et al., 2023a). Further, please note that the WebRTC output of the capture module will send our own modified VP9 encoding to enable the bitstream-based composition in the Holo-MCU (Gunkel et al., 2024).

At the core of HOLOVER-5G is a unity client integrated into a commercial application (OVER APP,¹⁰), the OVER APP leverages Gaussian splatting technology to deliver photorealistic, accurately scaled 3D reconstructions for industrial training scenarios. Each scene is pre-processed using an optimized pipeline, resulting in a lightweight, high-fidelity representation (under 30 MB per environment) that can be rendered in real-time on mobile devices. This approach enables immersive, interactive training experiences with realistic spatial cues, minimal storage requirements, and seamless performance, even on smartphones and tablets. Further, the client acts as a WebRTC Rendering Client that uses a custom shader to depict the RGBD points as custom 3D photo realistic mesh.

For the multi-user transmission, we use a Holo-MCU, based on the work presented in Gunkel et al. (2024) with the key difference of supporting WebRTC streams instead of RTP streams. The Holo-MCU connects with the individual upload clients via WebRTC to receive multiple VP9 RGBD streams. These streams follow a specific VP9 encoding (Gunkel et al., 2024). On each frame, the Holo-MCU combines all streams into one bitstream (in a horizontal tile composition). The final singular stream will be sent to all receiving user end clients.

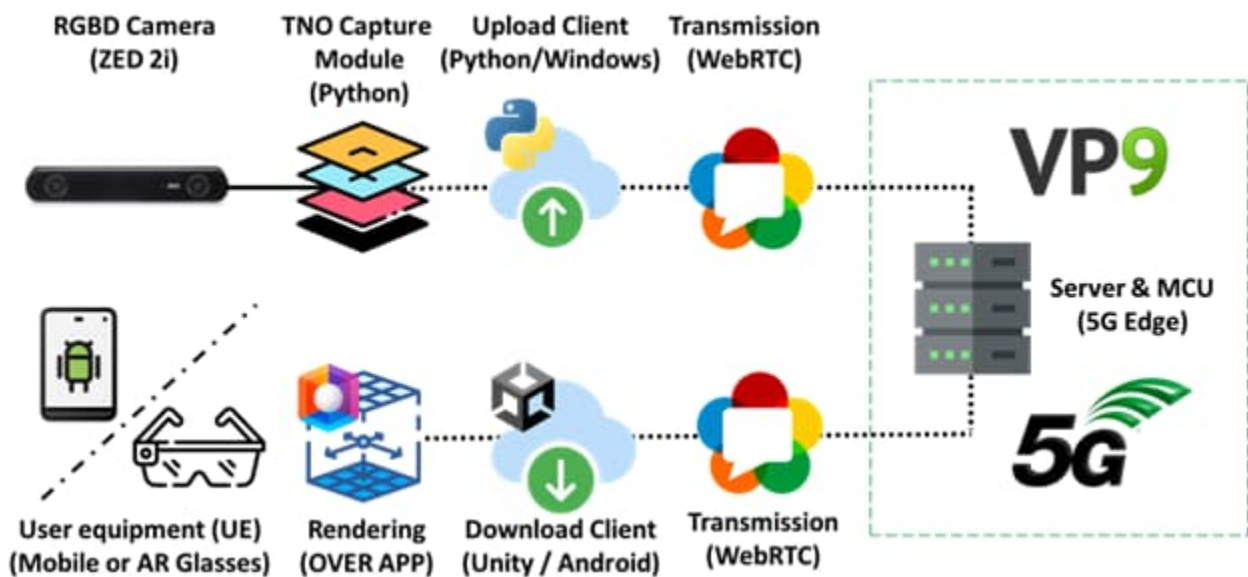


Figure P 2. Architecture and deployment.

¹⁰ <https://www.overthereality.ai/>

5G deployment and initial results

As an initial evaluation of our technical approach, we deployed the Holo-MCU and end clients in a 5G Test Lab facility in Berlin. The Holo-MCU performance was tested locally running on a laptop (Acer Predator Helios 300, PH315-55s-917f) with pre-encoded RGBD streams (512x1024 pixel, 30fps) with 2, 4 and 8 tiles. System delay was tested with the DelAyrUco tool¹¹ with the Holo-MCU deployed in the testbed (edge) server and the DelAyrUco client connected over Wi-Fi. Thus, the setup is similar to Gunkel et al., 2024, but in a more realistic server deployment. Our measurements are shown in Table and indicate similar results as in previous experiments (Gunkel et al., 2024), but now in an operational environment. Important to note, however, is that we expect additional delays if end devices are connected via 5G.

Table P 1. Initial Measurements

Components	2 User	4 User	8 User
Holo-MCU CPU in %	1.77 (+/- 0.35)	2.57 (+/- 0.38)	3.46 (+/- 0.29)
Holo-MCU bitrate in MB/s	1.83 (+/- 0.47)	3.65 (+/- 1.28)	7.42 (+/- 2.09)
System Delay in ms	263 (+/- 30)	305 (+/- 35)	332 (+/- 36)

Conclusion and future work

In this paper we present our design and outline the technical of HOLOVER-5G to support enhanced technical trainings with the help of scalable holographic communication and Gaussian splats. Furthermore, we present initial measurement results evaluating the technical suitability of our approach. Finally, more technical and user tests are planned to test all aspects of our approach in a full 5G deployment.

Acknowledgements

This work was partially funded by the European Union: SPIRIT OC1 (101070672), ETHOS (Enhanced Training over Holographic Scalable Communications)¹².

References

- Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., & Tabatabai, A. 2020. An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC). *APSIPA Transactions on Signal and Information Processing*, 9, e13. <https://doi.org/10.1017/ATSIP.2020.12>
- Gunkel, S.N., Dijkstra-Soudarissanane, S., Stokking, H.M., & Niamut, O.A. 2023a. From 2D to 3D video conferencing: Modular RGB-D capture and reconstruction for interactive natural user representations in immersive extended reality (XR) communication. *Frontiers in Signal Processing*, 3, 1139897. <https://doi.org/10.3389/frsip.2023.1139897>
- Gunkel, S.N., Klunder, T., & Cernigliaro, G. 2025. Ready Expert One: Universal 3D Workbench for Remote Industrial Training. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences* (pp. 359–361). <https://doi.org/10.1145/3706370.3731701>

¹¹ <https://github.com/TNO/DelAyrUco>, 512x1024 pixel, 30fps

¹² <https://spirit-project.eu/>

- Gunkel, S.N.B., Dijkstra-Soudarissanane, S., & Niamut, O. 2023b. "You AR' right in front of me": RGBD-based capture and rendering for remote training. In *Proceedings of the 14th Conference on ACM Multimedia Systems*. ence on ACM Multimedia Systems. <https://doi.org/10.1145/3587819.3593936>
- Gunkel, S. N., Hindriks, R., Assal, K. M. E., Stokking, H. M., Dijkstra-Soudarissanane, S., Haar, F. T., & Niamut, O. 2021. VRComm: An end-to-end web system for real-time photorealistic social VR communication. In *Proceedings of the 12th ACM Multimedia Systems Conference*. <https://doi.org/10.1145/3458305.3459595>
- Gunkel, S.N.B., Hindriks, R., Shiferaw, Y., Dijkstra-Soudarissanane, S., & Niamut, O. 2024. VP9 bitstream-based Tiled Multipoint Control Unit: Scaling simultaneous RGBD user streams in an immersive 3D communication system. In *Proceedings of the 15th ACM Multimedia Systems Conference* (pp. 23–33). <https://doi.org/10.1145/3625468.3647608>
- Jansen, J., Subramanyam, S., Bouqueau, R., Cernigliaro, G., Cabré, M. M., Pérez, F., & Cesar, P. 2020. A pipeline for multiparty volumetric video conferencing: Transmission of point clouds over low latency DASH. In *Proceedings of the 11th ACM Multimedia Systems Conference*. <https://doi.org/10.1145/3339825.3393578>
- Viola, I., & Cesar, P. 2023. Volumetric video streaming: Current approaches and implementations. In *Immersive Video Technologies*, 425–443. <https://doi.org/10.1016/B978-0-32-391755-1.00021-3>

BAZKARIA: Eating Together Through Immersive Telepresence

Andrés Santos-Torres¹, Patricia de Torres Coll¹, Tamás Bukits¹, Ander Elosegui¹,
Iñigo Tamayo¹, Ramón Perisé², Sergio Cabrero Barros¹, and Mikel Zorrilla¹

¹Fundación Vicomtech, Basque Research and Technology Alliance (BRTA), Spain

²Mugaritz, Spain

Corresponding author: asantos@vicomtech.org

Keywords: Volumetric Video, Social XR, Immersive Telepresence.

Abstract

Sharing a meal with others or commensality is a vital, social, and intimate experience. It positively influences our health and mood. However, sometimes the circumstances, such as geographical distance or our daily routines, prevent us from sharing this experience physically. The project BAZKARIA aims to create an immersive and augmented experience to bridge the barriers of remoteness. This project accomplished these aims by: (1) engaging domain experts from commensality and technological experts in co-design sessions to define the experience and the technological requirements; (2) providing an infrastructure that delivers to each remote commensal immersion and copresence using volumetric video; and (3) evaluating the experience with an in-the-wild user study, and the system architecture using Vicomtech and SPIRIT testbed.

Introduction

Eating is a vital, social, and intimate experience. Eating with others influences our health, e.g., prevents eating disorders in adolescents (de la Torre-Moral et al., 2021), while “social isolation and subjective loneliness are risk factors for malnutrition among older people” (Boulos, 2017). However, sometimes we lack the time to meet physically and share a meal.

In this context, BAZKARIA (Basque word for lunch) aims to create an immersive and augmented experience of eating together while apart. We aim to connect people eating in two different locations using immersive telepresence, focusing on three key challenges. The first challenge is understanding how remoteness changes the experience of eating and how technology must be deployed in this specific use case and operational context. The second is to use volumetric video to provide immersion and co-presence, while coping with current technical limitations. The third is to enhance the experience with alternatives for actions that take place while eating, but that are hard to replicate remotely, such as toasting. We call these augmented interactions.

We validated the telepresence infrastructure and the experience for Immersive Digital Commensality. First, we conducted an in-the-wild user study to validate the experience. Second, we deployed the architecture using the SPIRIT and Vicomtech testbeds.

Related Work

BAZKARIA aims to build an *XR table that engages remote diners in a virtually co-located* immersive dinner. In such scenarios, Pan et al. (2017) found that displaying virtual characters and as much contextual information as possible regarding social interactions enhances the sense of social presence in virtual environments. First, virtual characters in an eXtended Reality (XR) scenario enable social variables to enhance social interaction, as they generate human-like feelings such as empathy and emotions, among others. Second, acquiring the participants' behaviour enriches the social experience as they provide cues regarding the social interaction, such as the mood of others.

Regarding cases of volumetric capture for holographic telepresence, there are works, such as those by Pietroszek and Eckhardt (2020), that aim to utilise these systems to capture and create immersive stories. Additionally, studies such as Dijkstra-Soudarissanane (2020) and Irlitti et al. (2024) emphasise the importance of telepresence using volumetric video to establish common ground among remote participants interacting in a virtual co-located experience.

Experience Design

We conducted two participatory co-design workshops to identify user requirements of the experience (Santos-Torres et al., 2025) and its technological feasibility. We collected participants' impressions of the two workshops and systematised them into design requirements using open coding analysis: (C1) Provide coherence in the experience, integrating naturally the elements of the experience into each remote commensal table; (C2) Provide direct communication between the commensals; (C3) The experience should provide context to the shared experience, trying to replicate how a physical experience provides it; (C4) Augmented and substitute interactions to gap the barriers that remoteness.

System Architecture

The XR Table has two key components (See Figure P 3) to provide telepresence of each commensal on the remote side. The crucial elements of both sides are captured and transmitted to the other side through the network and naturally integrated into the remote side. The following paragraphs describe these components.

Telepresence system: composed of a *volumetric capture system* with one or more cameras, along with the calibration file, to reconstruct the element at the remote side, an *object segmentation component* to identify the elements and filter it from the rest of the scene, a *scene composition component* on the receiver side, to recreate the remote side coherently in the local scene, an *object tracking system* that provides spatial interactions.

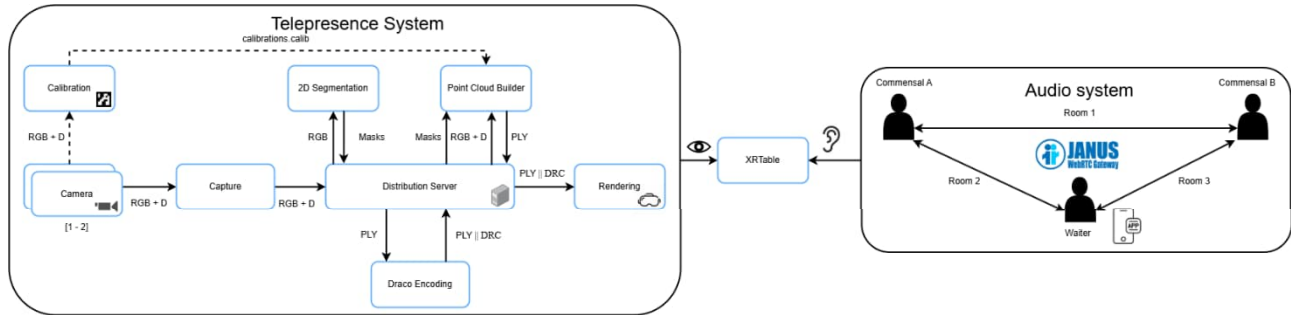


Figure P 3. Architecture of the system

Audio System: the setup consists of three separate rooms to manage audio communications between remote commensals and the waiter. To prevent crosstalk, as the same waiter serves each table in sequence, we deployed an isolated audio routing system that ensures only the active table's microphone and speakers are connected during interactions. This avoids overlapping audio capture between rooms.

Evaluation

User evaluation

To gain qualitative insights into the user experience, we are analysing the post-interviews from the study. Further analysis will be conducted to integrate the qualitative insights with the quantitative findings. These interviews have been transcribed and analysed into themes, each with its respective thematic code.

From preliminary analyses, we can conclude that there is an initial adaptation process to the experience, influenced by social factors such as inhibition and disorientation that shape the experience in the first steps. However, as the experience progresses, these factors become less relevant, according to the participants. Physical factors, such as the ergonomics of the headset and the cognitive load associated with familiarising oneself with the immersive environments, are predominant in the initial phases.

Technical Evaluation

The technical development primarily focused on designing a segmentation system capable of detecting and handling various object types based on specific user scenarios. This research addresses the challenge of transmitting large point clouds efficiently to multiple endpoints. To evaluate the performance of our segmentation pipeline, we established two distinct testbeds: one at the University of Surrey, leveraging 5G connectivity in line with project requirements, and another one at Vicomtech, utilising our local network to resolve and enhance issues previously encountered in the Surrey setup. To reproduce and evaluate the results, we utilised a publicly available dataset from CWI (Centrum Wiskunde & Informatica) (Reimat et al., 2021).

The results from the evaluation using the SPIRIT platform consistently fell below the target of 30 frames per second (fps). As the number of cameras increased, the segmentation process was unable to keep pace with the required streaming speed, resulting in a performance bottleneck. This led to occasional decoding failures on the client side, resulting in dropped or corrupted RGB frames that the YOLO model could not process. Learned lessons in this first evaluation; we enhanced the system design and validated the improvements using Vicomtech testbed. Enhancements such as: Multithreaded

Processing Pipeline to increase GPU Utilisation, Early Image Resizing for Efficiency and reduction of computational load, and parallelised output composition using a separate thread.

Conclusion and Future Work

This project designs, implements, and evaluates an Immersive Digital Commensality experience using volumetric video. We proposed a system to provide immersive telepresence systems to remote commensals using volumetric video. We evaluated the experience through a user study, gaining qualitative insights from the study participants. Furthermore, we tested the proposed system using a 5G testbed to evaluate its capabilities. In our future work, we plan to use the system to provide multisensory cues and a variety of interactions using the object tracking component and iterate on the overall user experience based on the findings of the user evaluations.

Acknowledgment

Funded by the European Union (SPIRIT, 101070672). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. The SPIRIT project has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

References

- de la Torre-Moral, A., Fàbregues, S., Bach-Faig, A., Fornieles-Deu, A., Medina, F. X., Aguilar-Martínez, A., & Sánchez-Carracedo, D. 2021. Family meals, conviviality, and the Mediterranean diet among families with adolescents. *International journal of environmental research and public health*, 18(5), 2499. <https://doi.org/10.3390/ijerph18052499>
- Boulos, C., Salameh, P., & Barberger-Gateau, P. 2017. Social isolation and risk for malnutrition among older people. *Geriatrics & gerontology international*, 17(2), 286-294. <https://doi.org/10.1111/ggi.12711>
- Pan, X., & Hamilton, A. F. D. C. 2018. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3), 395-417. <https://doi.org/10.1111/bjop.12290>
- Ohl, S. (2017). Tele-immersion concepts. *IEEE transactions on visualization and computer graphics*, 24(10), 2827-2842. <https://doi.org/10.1109/TVCG.2017.2767590>
- Reimat, I., Alexiou, E., Jansen, J., Viola, I., Subramanyam, S., & Cesar, P. 2021. Cwipc-sxr: Point cloud dynamic human dataset for social xr. In *Proceedings of the 12th ACM Multimedia Systems Conference* (pp. 300-306). <https://doi.org/10.1145/3458305.3478452>

Virtual Reality Evacuations for Resilient Mobility Systems: Users' Perspectives

Meng Cai, Niklas Suhre, and Eva Kassens-Noor

Technical University of Darmstadt, Germany

Corresponding author: cai@verkehr.tu-darmstadt.de

Keywords: evacuation simulation, user perspectives, resilient mobility, virtual reality

Introduction

The escalating impacts of climate change, coupled with the more frequent occurrence of extreme events, demand future mobility to be resilient. Research shows that critical infrastructures in Europe are facing growing risks from floods, heatwaves, droughts, windstorms, and wildfires (Forzieri et al., 2018), which poses significant challenges to transportation systems and beyond.

Although the advances in technology have brought new opportunities to improve the resilience of transportation systems, large mobility gaps exist among different populations. As new transportation options are penetrating the market, they continue to disproportionately favor the young, the male, the affluent, and those who are more comfortable with technology (Abraham et al., 2017; Bullard, 2003; Dill and McNeil, 2021). Thus, ensuring the future of mobility systems is resilient and inclusive requires not only innovative technologies but also community-centered solutions.

Recent advances in Virtual Reality (VR) technologies offer transportation researchers and practitioners great potential to enhance the resilience and inclusivity in mobility systems, especially in preparing for and responding to extreme events. Unlike traditional evacuation exercises, VR simulations safely recreate realistic crisis situations and allow users to interact with emergencies without physical risk or extensive logistical demands (Hung et al., 2025). A growing body of literature has demonstrated that VR can effectively simulate diverse disruptive scenarios. For instance, Kolb et al. (2019) created an immersive VR visualization of hydrological extreme events based on precipitation data. Most recently, Arizala et al. (2025) proposed an extended reality prototype for visualizing evacuation routes.

However, existing studies have predominantly focused on VR's technical feasibility and operational effectiveness, while largely neglecting users' perspectives on their engagement in simulated environments. This oversight is critical because refining VR interfaces, pacing, and scenario complexity depends directly on such feedback. In addition, integrating users' feedback into VR design aligns with community-centered resilience planning efforts. Therefore, this pilot study explicitly explores users' perspectives of VR evacuations within resilient mobility contexts. Specifically, this research addresses three questions: 1) How do users perceive their overall experience when interacting with VR evacuation

simulations? 2) Do users' perspectives differ by gender? 3) Do users' perspectives vary depending on whether they have interacted with VR before?

Methods

We simulated a flash flood-induced emergency scenario in a central area of Frankfurt am Main, Germany (Figure a). Participants were tasked with evacuating on foot from a designated starting point to a pre-assigned safe location situated in a neighboring block. The shortest evacuation route required crossing a bridge over a river. As participants approached the bridge, it collapsed due to rapidly intensifying currents, forcing them to identify and pursue an alternative route using the remaining pedestrian infrastructure. The simulation was implemented using the Cyberith Virtualizer Elite 2 omnidirectional treadmill for physical locomotion, in combination with HP Reverb G2 VR headsets (Figure b). Prior to entering the simulation, participants completed a brief questionnaire of self-identified gender and prior experience with VR. Upon completion, participants filled out a follow-up survey to describe their experience and provide any additional comments. We translated responses written in German into English using DeepL Translator.

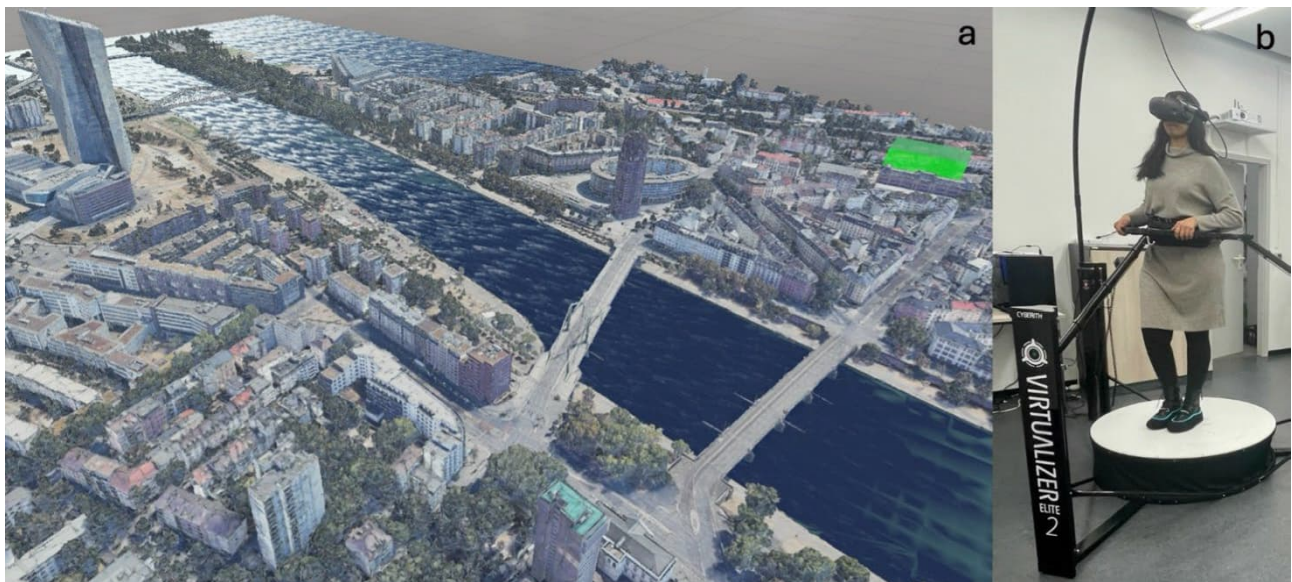


Figure P 4. Simulation setup and equipment. a) VR simulation of a flash flood scenario in central Frankfurt am Main, Germany. The green box indicates the designated evacuation destination. b) Using Cyberith Virtualizer Elite 2 omnidirectional treadmill and HP Reverb G2 VR headset to navigate the simulated environment.

A total of 143 students affiliated with the Technical University of Darmstadt signed up for the study. Of these, 15 participants either withdrew or were unable to complete the evacuation successfully due to reasons such as motion sickness or incompatibility with the VR headset. The final sample consisted of 128 participants who completed the study. Participation in the study was entirely voluntary. The study did not provide monetary or material compensation. The study was approved by the Ethics Commission of the Technical University of Darmstadt (EK 52/2023).

We conducted a manual content analysis to systematically analyze participants' responses, which involves identifying recurring themes and patterns through an iterative coding process (Krippendorff, 2018). In addition, to understand the general user experiences, we classified each participant's combined textual responses as positive, negative, or neutral and examined differences in their responses according to participants' self-identified gender and their prior experience with VR.

Results

In general, participants perceived their VR evacuation experiences positively. Participants frequently described their experiences using terms such as "interesting," "good," and "engaging." 63% (80 out of 128) described their overall experience positively, pointing out aspects such as the immersive realism and engaging crisis scenarios. 31% (40 out of 128) of the participants gave neutral feedback, including minor navigational challenges and initial confusion within the virtual environment; for example, a participant mentioned uncertainty about choosing alternative routes after the bridge collapse. A minority, only 6% (8 out of 128), reported negative experiences, which primarily associated with physical discomfort, motion sickness, or technical issues.

Male participants reported slightly more positive experiences and fewer negative concerns compared to female participants. Among female participants, 59% (29 out of 49) described their experience positively, 31% (15 out of 49) neutrally, and 10% (5 out of 49) negatively. Female participants often cited initial orientation challenges and headset discomfort. One participant stated, "The experience was engaging but slightly confusing at first due to unclear navigation cues." Male participants provided slightly more positive feedback, with 65% (51 out of 79) reporting positively, 32% (25 out of 79) neutrally, and only 4% (3 out of 79) negatively. Male users often emphasized the engaging realism, such as the excitement and immersion created by unexpected events like the bridge collapse.

Furthermore, participants with no prior VR experience expressed higher enthusiasm, whereas experienced VR users tended to provide more critical feedback. For participants without previous VR experience, 65% (55 out of 85) reported positive experiences, including the excitement and novelty of their first immersive virtual scenario. 31% (26 out of 85) gave neutral feedback. Only 5% (4 out of 85) reported negative experiences, typically related to discomfort or initial disorientation. Among participants who had prior VR experience, positive responses were slightly lower at 58% (25 out of 43). Neutral feedback accounted for 33% (14 out of 43), reflecting higher expectations or comparative experiences with other VR platforms. Negative feedback was slightly higher at 9% (4 out of 43), mostly mentioning technical limitations or shortcomings compared to previous VR experiences.

Discussion

By focusing explicitly on users' perspectives, this study contributes to the ongoing discussion of the development of resilient and inclusive mobility systems. Findings derived from user feedback show the importance of designing VR evacuation scenarios that accommodate diverse user groups, thus supporting more equitable and effective evacuation strategies in mobility planning. Gender-based differences in user experiences reveal essential considerations for enhancing resilience in mobility systems. The increased incidence of navigational challenges and discomfort among female participants suggests that tailored interface designs and scenario adjustments could mitigate potential gender-related disparities in evacuation effectiveness. In addition, varying perspectives based on participants' prior VR experience indicate the need for adaptive VR environments that optimize realism and usability according to users' familiarity with the technology. A key limitation of this study is its relatively homogeneous sample, which limits the generalizability of findings. In addition, the simulation considered only one disaster type. Future research should expand on these findings by evaluating broader user groups and more mobility scenarios to better support building resilient and inclusive mobility systems.

Acknowledgement

Funded by the European Union (ERC, scAlnce, 101087218). 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 Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. We thank NUMENA for their support in setting up the equipment.

References

- Abraham, H., Lee, C., Brady, S., Fitzgerald, C., Mehler, B., Reimer, B., Coughlin, J.F. 2017. Autonomous vehicles and alternatives to driving: trust, preferences, and effects of age, in: Proceedings of the Transportation Research Board 96th Annual Meeting. Transportation Research Board Washington, DC, pp. 8–12.
- Arizala, J.L., Permito, J., Escopete, S.E., Niño, J.K., Deja, J.A., 2025. Designing RoutScape: Geospatial Prototyping with XR for Flood Evacuation Planning. <https://doi.org/10.48550/arXiv.2505.06045>
- Bullard, R.D., 2003. Addressing urban transportation equity in the United States. *Fordham Urb LJ* 31, 1183.
- Dill, J., McNeil, N., 2021. Are Shared Vehicles Shared by All? A Review of Equity and Vehicle Sharing. *J. Plan. Lit.* 36, 5–30. <https://doi.org/10.1177/0885412220966732>
- Forzieri, G., Bianchi, A., e Silva, F.B., Herrera, M.A.M., Leblois, A., Lavallo, C., Aerts, J.C., Feyen, L. 2018. Escalating impacts of climate extremes on critical infrastructures in Europe. *Glob. Environ. Change* 48, 97–107. <https://doi.org/10.1016/j.gloenvcha.2017.11.007>
- Hung, M.-C., Lin, C.-Y., Hsiao, G.L.-K., 2025. Virtual Reality in Building Evacuation: A Review. *Fire* 8, 80. <https://doi.org/10.3390/fire8020080>
- Kolb, D., Kurtz, W., Weismüller, J., Ramm, A. von, Ludwig, R., Kranzlmüller, D., 2019. Using virtual reality to visualize extreme rainfall events derived from climate simulations. Presented at the EGU General Assembly 2019, Vienna, Austria. <https://doi.org/10.5282/ubm/epub.71906>
- Krippendorff, K., 2018. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications.

SELFEX²: Real-Time Telepresence Dexterity Training for Manufacturing Using Wearable Technologies and Augmented Reality

Angel Dacal-Nieto^{1*}, Alba Domínguez-Pedreira¹, María Rodríguez-Pérez¹,
Víctor Alonso-Ramos¹, and Oscar Rodríguez-González²

¹CTAG – Automotive Technology Center of Galicia, Spain

²SFC Solutions S.A., Spain

Corresponding author: angel.dacal@ctag.com

Keywords: Augmented-Reality, manufacturing, training, finger tracking, wearable.

Abstract

The process of training a manufacturing operator is usually long and complex, involving time, resources, and expert trainers. SELFEX² is a new approach which aims to improve training processes in manufacturing by using wearable finger-tracking gloves, trackers and extended reality, to allow a self-training procedure that provides a quantifiable degree of readiness to execute the workplace in line. By using the tools of the CORTEX² framework, SELFEX² allows real-time training between one teacher-senior operator in one location, and several junior operators in remote locations, that can learn to execute dexterity-based tasks in real-time, by combining video, voice, and the augmented reality representation of the hands of the teacher. This way, the junior operators can also learn and execute the tasks in their remote locations with finger tracking devices, allowing them to quantitatively measure their dexterity on the task and receiving real-time insights and comments, both in a fully remote way.

Introduction

Despite the efforts of virtual reality technologies (Naranjo et al., 2020), the training of manufacturing operators is still based nowadays on a combination of text, videos, and an incremental presence in the workplace until some reviewer agent subjectively approves the operator to work in line. This approach could be improved if 1) the junior operators could self-educate a large part of the training, 2) the abilities and dexterity of the junior operators could be quantified, and 3) the experience of senior operators could be properly documented. Recording the movements of their hands in dexterity-based operations that are not enough digitized, would benefit future generations. As a solution, SELFEX² introduces finger tracking gloves and XR devices as a complementary method for the training process of manufacturing operators, tailored to high-dexterity workplaces (Triviño-Tonato et al., 2024). However, not every dexterity-based operation can be learned in an autonomous way.

This means that a real-time learning process can also be positive for the training of such manufacturing tasks under the finger-tracking paradigm. CORTEX² (Pagani et al., 2023) is a cooperative platform for extended reality applications, which has been identified as an excellent complementary tool to provide this real-time feature to SELFEX². By using the CORTEX² features, SELFEX² can facilitate this collaborative work in a remote way, using a telepresence approach.

Concept

SELFEX² defines a new way to allow a collaborative bidirectional learning flow, permitting the teacher and learners to reference some common information, reinforcing some complex steps of the learning process that require more depth, pointing out details, or a specific interaction. The learning methodology of SELFEX² offers the possibility of performing real-time live training sessions where a trainer explains how to develop a manufacturing task to a group of trainees. SELFEX² launches a conference between users, in which video, audio and kinematic information are shared. This is performed using the Rainbow teleoperation platform integrated into CORTEX² (see **Figure P 5**).

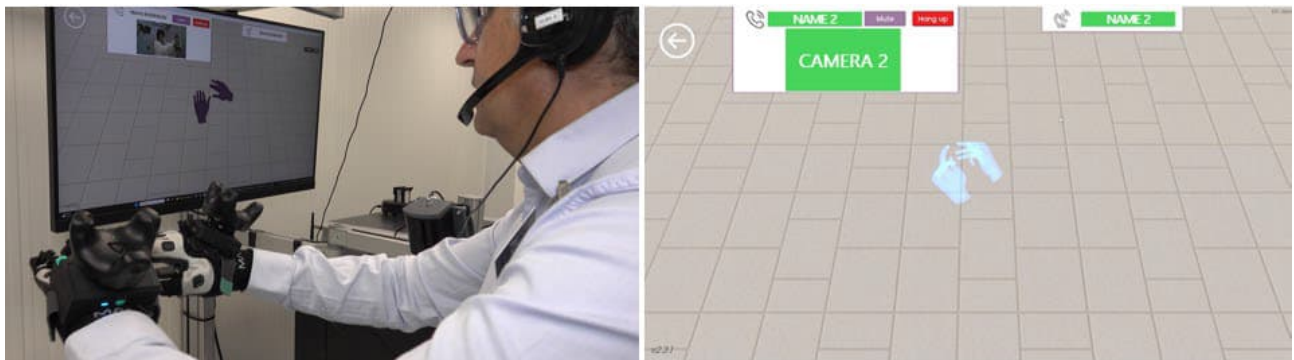


Figure P 5. Telepresence view of a senior operator teaching a task to a junior operator. Left: view from the senior perspective. Right: view from the junior operator using the computer interface.

SELFEX² combines different tools to perform this functionality: 1) finger tracking gloves from Manus (Quantum XR Metagloves¹³ or Haptikos¹⁴, 2) trackers and antennas, 3) augmented reality device (Microsoft HoloLens 2¹⁵), 4) a series of cameras and screens, and 5) a server to execute the software platform. This set is complemented by a workspace to place the object related with the task to learn. SELFEX² is deployed in a portable sensorized cabin, so that this environment is a moving laboratory that can be brought to the manufacturing site and perform in-site experiences (see **Figure P 6**).

¹³ Manus website: <https://www.manus-meta.com/>. Accessed 3rd July 2025

¹⁴ Haptikos website: <http://www.haptikos.tech/>. Accessed 3rd July 2025.

¹⁵ Microsoft HoloLens 2 website: <https://learn.microsoft.com/es-es/hololens/>. Accessed 3rd July 2025.



Figure P 6. SELFEX² workbench (left) and in its portable cabin (right).

Results

SELFEX² has been evaluated with the end user SFC Solutions, a manufacturer, distributor and integrator of manufacturing flexibility tools known as hexapods. This company delivers these mechatronic devices throughout Europe, and this often requires specific travels to perform teaching sessions regarding their functionalities and maintenance procedures.

Using their use cases as benchmark, the evaluation of SELFEX² has been carried out through practical sessions with real users in a controlled environment, measuring technical features (such as delays in the data transmission) and different usability experiences using exhaustive questionnaires (see **Figure P 7**). Some aspects have been compared, including the different perception between senior and junior users, glove model, visualization mode, age, dominant hand, efficiency of the P2P and conference modes, among others.

Overall, the results show that the solution is technically feasible and useful, with acceptable delay levels and a good perception of the experience by the users. This evaluation has allowed us to find improvement areas, including new visualization concepts, accuracy and motion capture range extension.



Figure P 7. HoloLens 2 usage (left) and phantom hands visualization (right).

Conclusions

The end user SFC Solutions indicates that this technology can represent a relevant innovation in terms of efficiency, cost reduction and quality of the training and maintenance experiences for their products. This can be directly applied to their business case, which reinforces the potential of the system to represent a valuable solution for remote training.

SELFEX² facilitates the learning processes, implying less transportation costs and emissions, reducing the number of trips per employee, and subsequently with economic impact on the companies. Moreover, it impacts the reduction in CO₂ emissions and the retention of know-how, thus reducing downtime in the current aged European landscape.

SELFEX² is an interesting technological brick to apply in future Industry 5.0 initiatives, as a cognitive augmentation tool for the manufacturing operator and human-centric factories.

Acknowledgements

The authors want to thank the project SELFEX², composed by CTAG and SFC Solutions, funded by an open call of the project CORTEX², which has received funding from the European Union's Horizon Europe research and innovation program under grant agreement N° 101070192.

References

- Naranjo, J.E., Sanchez, D.G., Robalino-Lopez, A., Robalino-Lopez, P., Alarcon Ortiz, A., Garcia, M.V. 2020. A scoping review on virtual reality-based industrial training. *Appl. Sci.* **10**, 8224. <https://doi.org/10.3390/app10228224>
- Triviño-Tonato, E., Villar-García, M., Dacal-Nieto, A. 2024. Manufacturing Operators Self-training Using Finger and Wrist Kinematics. In: Pons, J.L., Tornero, J., Akay, M. (eds) *Converging Clinical and Engineering Research on Neurorehabilitation V. ICNR 2024. Biosystems & Biorobotics*, vol 32. Springer, Cham.
- Pagani, A., Minaskan, N., Javanmardi, A., Xie, Y., Rivier, S., Bailleau, V., ... & Garcia-Palacios, A. 2023. CORTEX2 - Extended collaborative telepresence for future work and education. *LMDE 2023 - Leading and Managing in the Digital Rea*, Jun 2023, Athènes, Greece. [ff10.5281/zenodo.8065220](https://zenodo.org/record/8065220). [ffcea-04575038](https://doi.org/10.1145/3588888)

ROTIE (Robot Teleoperation through Immersive Environments)

Anabel Díaz-Labrador, Miguel Collazo, Diego Veiga, Antón Concheiro, Santiago Serantes, Diego Socas, Angel Delgado, Pablo Fernández, and Héctor Pérez

ITG - Instituto Tecnológico de Galicia, Spain

Corresponding author: adelgado@itg.es

Keywords: Human-robot interaction, XR, robotics, dogbot, VR, telepresence

Summary

ROTIE is a distributed and fully containerization-compatible platform conceived to allow robot operators to monitor and drive their mobile robots or manipulators through immersive interfaces working in bi-directional mode. It will be integrated in the framework of the European Project SPIRIT¹⁶. The framework offers APIs for advanced Human Interaction Devices that are used to facilitate user-intention-based robot control. And, on the other way round, the user is virtually tele-transported to the robot environment through the use and fusion of sensorial data including principally 2D/3D images and spatial sound. Besides, the platform is designed to allow also the development of digital twins that can extend the performance of robot operations for pre-operation planning and post-operation analysis.

Architecture

Our application provides the necessary interfaces (Web Sockets, Restful APIs, WebRTC clients and server, etc.) to remotely and safely control one or multiple semi-autonomous robots in an unknown environment alongside with a fully immersive recreation of said environment using data acquired by the robot. This application consists of several fully scalable containerized services that can be deployed on a cluster or on the cloud. Figure P 8 depicts the architecture of ROTIE solution.

¹⁶ <https://spirit-project.eu/>

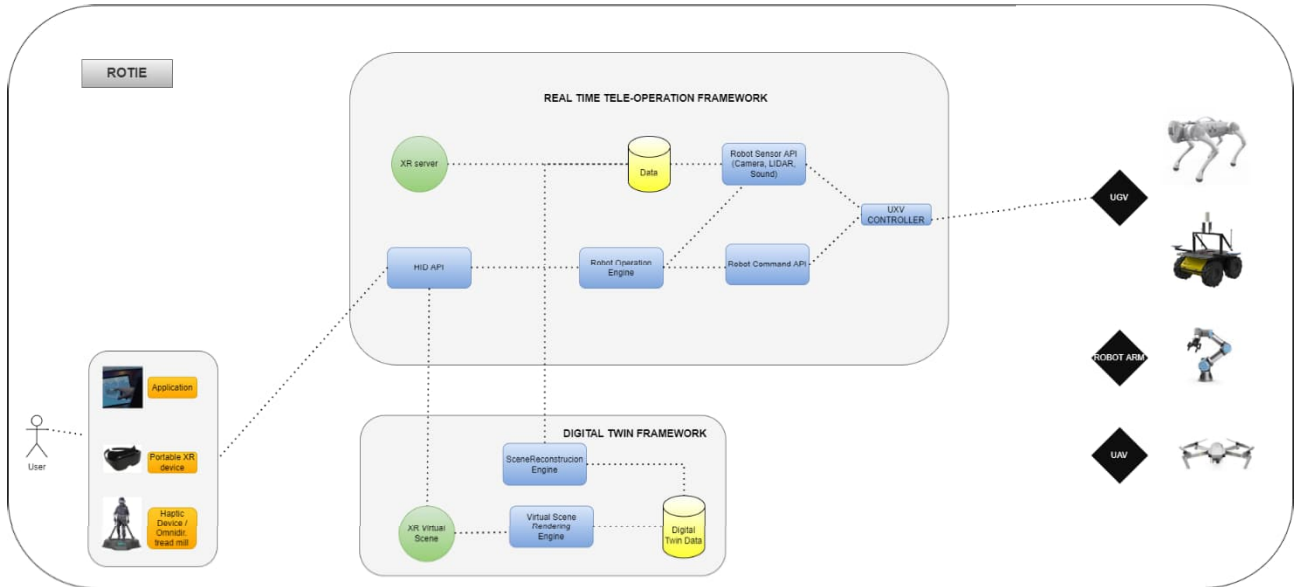


Figure P 8. ROTIE architecture

Devices

- Dogbot Unitree GO1
- Dogbot Unitree GO2
- Intel Real Sense cameras
- AR Glasses Quest 3
- Omnidirectional treadmill Katwalk VR

Features

Dogbot control with HID devices

A HTTP interface (Rest API, WebSocket) has been developed for multi-user control. This interface receives the dogbot's status (battery, IMU, position, temperature, etc.) at a configurable frequency of one second. This interface also allows high-level commands to be sent to the dogbot to move it (Euler for rotations, speed, spin rate, height, etc.).

This interface works as an adapter and allows different HID devices to control the dogbot without the need to know how to manage its control and simplifying control to X-Y displacements, and linear and rotational speed commands.

A service was developed to get information from AR glass (Quest 3) with linked joysticks to convert user movements and actions into real robot movements (Figure 2). Besides, a service was developed to convert the movement of an omnidirectional treadmill (model Katwalk VR) into movements on the dogbot, allowing a user to control the dogbot from it.

Video streaming with on-board cameras

Services to rescale/normalize the videos from the dogbot cameras and stream them using RTSP and MJPEG were developed. This protocol and format provide lower latency (and makes it easier to consume videos from third-party sources). Additionally, it was also tested for the deployment of a media center (mediamtx) within dogbot itself to allow multiple users to consume the RTSP and even convert it to

other streaming formats such as WebRTC. This media center has also been deployed on an external server, which allows users to receive streams from the Internet.

Advanced robot control based on image processing

A person tracker was implemented using YOLO v5, running directly on the dogbot and using the cameras and control API to track the person closest to the dogbot. The same tracker was extended to track objects based on their color. This latter tracker relies on a color filtering algorithm.

Advanced robot control using a voice assistant

A virtual/physical assistant that runs inside the dogbot and does not require internet access was implemented. This assistant constantly picks up sound from the microphone, so it doesn't require a keyword and applies an ASR and NLU model to detect what is being said and the intent of the command. An app was also adapted to communicate with the dogbot from a mobile phone, allowing it to speak and give commands from its microphone. A service was also developed to manage the audio played on the dogbot, allowing the assistant to interact with the user, allowing users to send text to be played on the dogbot (TTS), and even allowing remote users to speak from an external device through the dogbot's speakers.

On-going developments & integrations

Scene virtualization using data from robot sensors

Different services to get, integrate and pre-process information from the cameras and other sensors of the robot have been developed. To generate a virtual scene, other services are being developed and tested with the objective of minimizing the latency for generating the virtual scene.

A service to launch on a server with proper GPU and generate a virtualization of an environment through video/images is being optimized. Gaussian splatting techniques are being integrated and investigated.

Integration with Robot Husky from SPIRIT partner, T-Systems

The features described in the previous section are currently being adapted to interact with the robot Husky, integrated in the T-Systems testbed of the EU Project SPIRIT, located in Berlin.

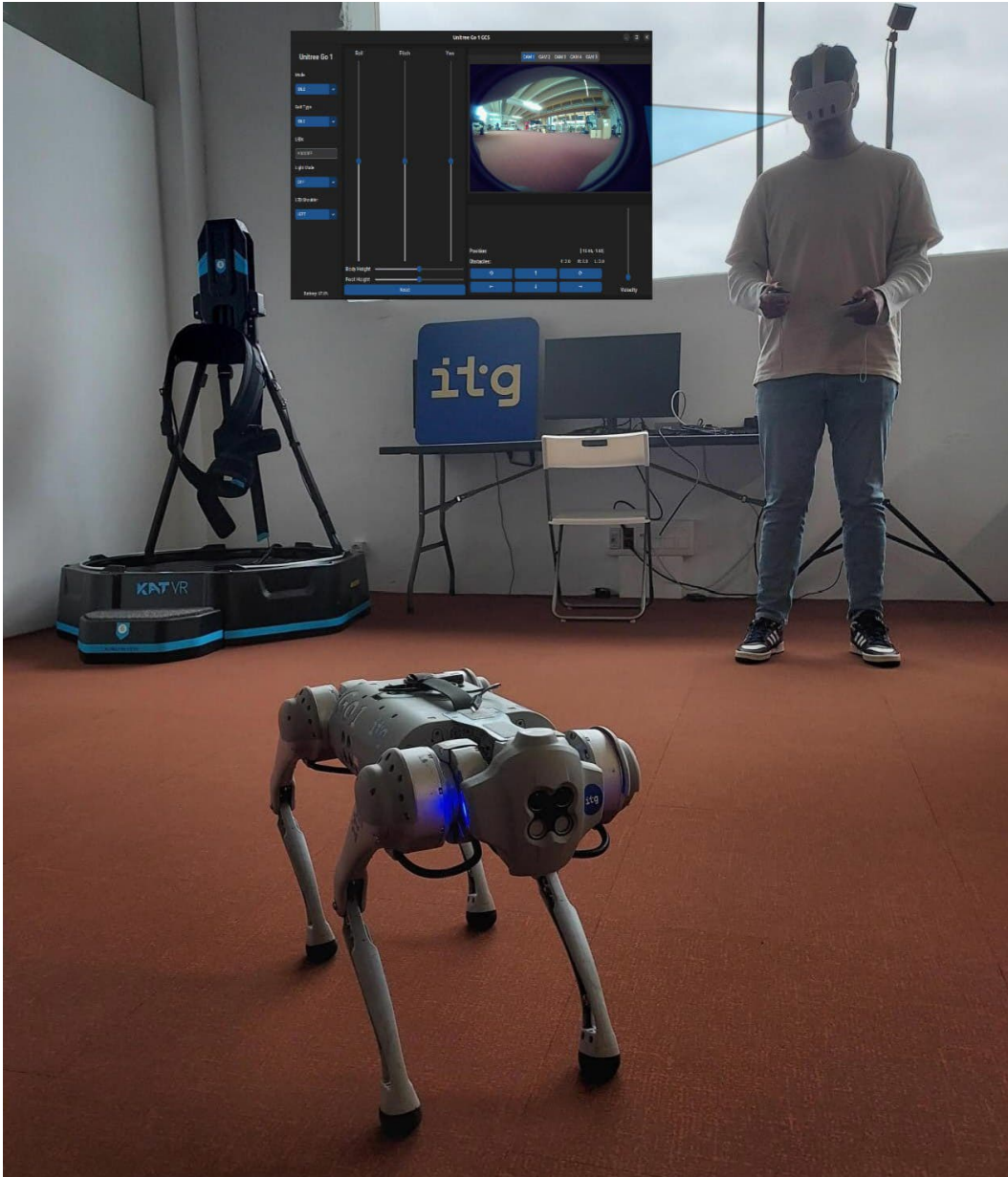


Figure P 9. Testing ROTIE telepresence framework with dogbot Unitree GO 1 and AR glasses Quest 3.

Acknowledgments

This project is funded by the European Union (Scalable Platform for Innovations on Real-time Immersive Telepresence (SPIRIT), 101070672).

This work has also received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) and the UK Research and Innovation (UKRI).

Anabel Díaz Labrador is developing this research with the support of the Xunta de Galicia, through the programme Doutoramento Industrial 2024 (ref. 02_IN606D_2024_3100897)¹⁷

¹⁷ <http://gain.xunta.gal/>

Robust AR Anchoring Using Ultra-Wideband (UWB) Beacons for Spatial Consistency in XR Environments

Alexandre Fresnais, Alex Levigoureux, Éric Heurtier, and Karen Crevecoeur

ARY SAS, France

Corresponding author: ehurtier@ary.eu

Keywords: Augmented-Reality, Spatial Anchoring, Wireless, Ultra-Wideband, Indoor Positioning System, Beacons

Motivation

Accurate spatial anchoring in augmented reality (AR) environments is essential for delivering consistent and immersive experiences, particularly across multiple sessions or for multi-user interactions. Existing AR systems often depend on visual cues such as QR codes or environment scanning, which can be unreliable in dynamic, cluttered, or visually ambiguous settings.

With support from the CORTEX2 consortium and funding from the European Union, ARY introduces a wireless alternative based on Ultra-Wideband (UWB) beacons to improve the accuracy and robustness of spatial anchoring in AR applications.

Proposed System

The system developed by ARY consists of a minimum of three interconnected beacons. These beacons initially broadcast over Bluetooth to establish a connection with the user's smartphone. Once a stable link is formed, the devices transition to Ultra-Wideband (UWB) communication, leveraging its sub-15 cm real-time localization accuracy. Users remain connected for only a few seconds, just long enough to become fully anchored within a shared spatial coordinate system. From that point onward, the device can rely on its native SLAM (Simultaneous Localization and Mapping) capabilities for the remainder of the AR session.

The system integrates seamlessly with Unity-based AR applications through a custom plugin that aligns virtual anchors with UWB-derived world coordinates. Once synchronized, developers can place virtual content confidently, ensuring that the AR environment appears in the same physical location across all connected devices.

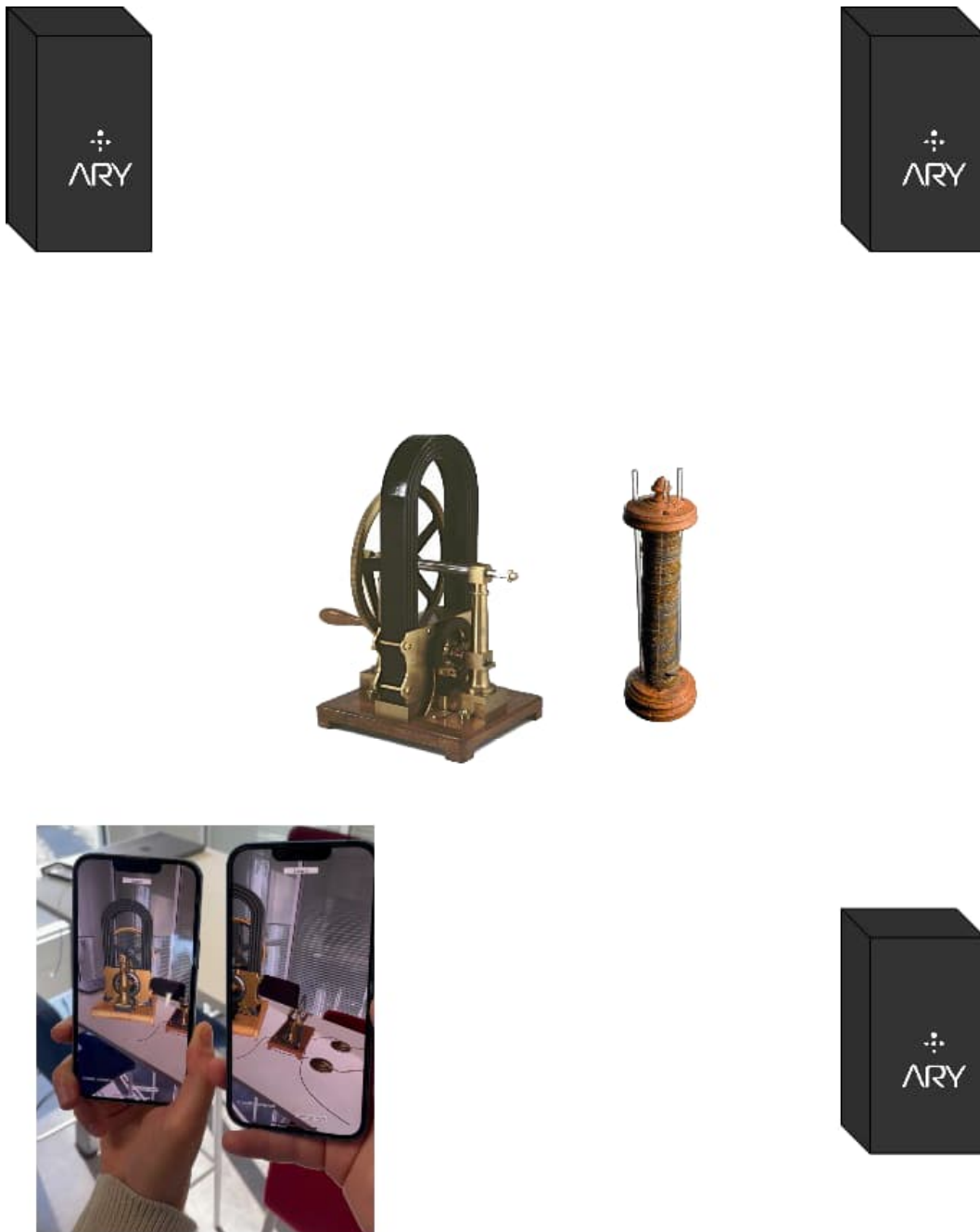


Figure P 10. An AR scene anchored using ARY. Virtual content is positioned within a coordinate space defined by three beacons, allowing user devices to display the content at the same physical location across multiple views.

In addition to the beacons, ARY introduces a compact UWB tag that enables any smartphone to be compatible with the system. As of 2025, not all smartphones are equipped with native Ultra-Wideband capabilities, most iPhones have supported UWB since the iPhone X, but on Android, this feature is generally limited to high-end models. To eliminate this limitation and ensure a consistent user experience, ARY developed a standalone UWB module that extends compatibility to virtually any smartphone.

Key Metrics

ARY has thoroughly tested the solution to determine its performance limits. When used with a compatible smartphone or with the ARY UWB tag, users can expect a localization accuracy of approximately 10 centimeters. In practice, this means that two users connected to the system will perceive the same virtual content with a physical alignment difference of no more than 10 centimeters.

Due to inherent limitations of Ultra-Wideband technology, each beacon can support up to five concurrently connected devices. However, as described above, the anchoring process requires less than 10 seconds per user, after which the connection slot is freed for others. To accommodate more simultaneous users, additional beacons can be deployed within the environment.

Beacons can maintain their accuracy at distances of up to 30 meters. In general, ARY recommends deploying three beacons to cover an area of 100 m². In environments with large crowds, significant obstructions, or potential signal interference, additional beacons may be necessary to ensure reliable and consistent coverage.

Conclusion

This work presents a robust solution for precise and user-friendly spatial anchoring in augmented reality environments, addressing the limitations of existing vision-based methods. ARY's system leverages Ultra-Wideband (UWB) technology to offer real-time localization with sub-15 cm accuracy, enabling consistent placement of virtual content across multiple devices and sessions.

Key contributions include:

- A plug-and-play infrastructure based on a network of UWB beacons requiring minimal setup.
- A cross-device anchoring mechanism that completes in under 10 seconds and supports scalable user access through dynamic connection handling.
- A custom Unity integration that allows developers to align virtual scenes with real-world coordinates seamlessly.
- A UWB tag extension, ensuring compatibility with a wide range of smartphones, including devices without native UWB support.

By removing the dependency on visual markers and enhancing spatial consistency across users, ARY provides a scalable and reliable anchoring solution ideal for persistent AR experiences, multi-user applications, and industrial use cases.



Funded by
the European Union



VISOR: Virtualization Service for Object Reconstruction

Iordanis Evangelou, Anastasios Gkaravelis, Nick Vitsas, and Andreas A. Vasilakis

Phasmatic, Greece

Corresponding author: iordanis@phasmatic.com

Keywords: 3D reconstruction, neural networks, digital twin.

Introduction

Recent advances in deep learning, particularly in 3D geometric learning, combined with the growing availability of specialized hardware, have led to the development of powerful new methodologies, algorithms, and tools for 3D content creation. These innovations enable automatic generation, manipulation, and interpretation of complex spatial data at a computational cost that is manageable with today's commodity hardware. This work focused on a self-supervised training approach that leverages input data such as 2D images or other task-specific modalities captured from the physical world. The output is a digital twin in the form of a 3D triangular mesh, ready for real-time use in any conventional visualization framework (e.g., web, mobile, desktop). This type of 3D geometric learning algorithms has the potential to automate the creation of digital replicas of real-world objects. Consequently, these technologies can greatly accelerate the traditionally labor-intensive process of generating accurate virtual representations, streamlining workflows and enabling more efficient and high-fidelity novel content creation. However, architecting a reliable end-to-end workflow is a non-trivial task since one needs to strike balance between several key factors. These include output quality in terms of surface structure and associate properties (e.g., color), computation time and resources to prepare the input stream and generate the final mesh as well as transform the content of the output mesh such as quality is maximized while size is minimal to be transmitted and loaded to any conventional device.

This work introduces VISOR, a virtualization service for object reconstruction which builds upon the latest developments in the literature Neural Radiance Fields (Mildenhall et al., 2021) (NeRFs) and stands as a complete pipeline for generating 3D content from real-world physical objects. The proposed pipeline integrates and extends state-of-the-art techniques at the intersection of computer vision, computer graphics, and neural optimization and it is structured into three distinct phases: pre-processing, optimization, and post-processing, each of which is discussed in detail below. Finally, we present empirical results on a diverse set of objects to demonstrate the effectiveness of the pipeline and discuss its current limitations along with potential future directions for improvement.

Methodology

VISOR is designed for the reconstruction of single entities from the physical world that can be captured by conventional recording devices such as a smartphone. In this section, we outline the key methodologies and design considerations required to construct such a workflow to transform a sequence of images, or a video stream, of a physical object into a 3D triangular mesh. Briefly, the VISOR pipeline can be described by three main stages: pre-processing, optimization, and post-processing, each designed to efficiently extract and refine spatial and appearance information from unstructured visual input.

Pre-processing. VISOR receives input from a finite stream of images or a video sequence capturing the target physical object. There are no restrictions on the type of images or the video encoding as well as the capturing resolution. To generate a plausible result, the input set of images can be as low as 50 unique and overlapping photos while the video sequence can be at least 30 seconds long. Since VISOR optimizes mesh structure from raw images, if a video stream is supplied instead, we automatically convert it to a finite image set by adaptively sampling the sharpest consecutive frames. To proceed with the NeRF optimization stage, camera calibration to extract both intrinsic and extrinsic camera parameters is required since these per-image parameters are typically unknown but implicitly defined by the user's capture device. VISOR recovers those parameters by applying the GLOMAP (Pan et al., 2024) calibration toolbox which balances between speed and accuracy. Finally, we isolate the object interest from the background for each image by applying a general-purpose dichotomous image segmentation neural network (Yu et al., 2024). Each generated tuple of segmentation mask and viewport transformation is forwarded to the optimization stage.

Optimization. VISOR's reconstruction methodology follows the signed distance field (SDF) formulation for which the reference 3D surface lies at the zero-level set. SDF essentially defines a real function that maps 3D coordinates to real values indicating their closest distance from the reference surface. In the special case of the zero-level set, these distances must be equal to zero. This paradigm is a common modelling approach in object reconstruction due to its continuous, flexible and mathematically grounded form. The task of the VISOR service is to learn and recover this function from the input stream of images. VISOR optimizes the SDF by following the current state-of-the-art paradigm based on NeRF theory (Mildenhall et al., 2021). Briefly, it leverages volume rendering designed for novel view synthesis tasks reparameterized in a way that allows for SDF optimization at the same time. As is typical in this optimization regime, ray batches are iteratively generated from the calibrated viewports of the input images. These rays are then sampled across their span in a predefined granularity and the volume rendering equation is evaluated in a per-pixel basis. VISOR follows practices from the most recent literature (Bailey et al., 2024), which demonstrates how opaque objects can be formulated as volumetric entities by adopting stochastic geometry theory. Then, by reparameterizing the volume rendering equation the novel synthesis task can be reformulated as an SDF optimization problem, effectively retrieving the zero level-set of the reference object. VISOR follows this paradigm as its core optimization methodology. To make this approach tractable in terms of resources and computation time, VISOR is implemented based on state-of-the-art hardware accelerated neural encoding modules (Müller et al., 2022) as well as adopting specific design choices for the underlying neural-based modules for this task (Wang, et al., 2023) to further accelerate the training phase. As an additional optimization and to further accelerate convergence, VISOR incorporates loss-driven adaptive ray generation. This scheme effectively decouples the input image resolution from the actual physical object which is only depicted on a small portion of it.

Post-processing. In the final phase, VISOR processes the signed distance values sampled from the optimized field to generate an initial triangular 3D mesh. An isosurface extraction algorithm (in our case, Marching Cubes (Lorenson & Cline, 1998)) is used to construct the polygonal surface. Notably, the chosen resolution to discretize space directly influences both quality and file size. Since we aim to extract a triangular surface that closely resembles the optimized SDF we prioritize quality over size in the first step. Then, the surface is appropriately simplified to reduce primitive redundancies (Kapulkinė, 2025) to drastically reduce the initial triangle count. To also embed surface properties such as color information we apply UV-unwrapping (Young, 2022), allowing for texture baking. Since the simplified geometry includes new points, we need to resample the radiance in the hemisphere around the local normal vector using the already trained model to estimate the per-point appearance. Surface colors are encoded into the texture map we have already created and enable realistic shading of the mesh. The completed 3D model is then exported in a standard file format (e.g., FBX or GLB), making it ready for real-time rendering in 3D game engines such as Unity or Unreal Engine.

Results

In this section we provide some experimental results demonstrating the efficiency of the VISOR service. The full workflow of VISOR can be executed in under 10 minutes on commodity GPU hardware, such as NVIDIA RTX 4090, for a single object instance. In **Figure P 11**, we depict some indicative outputs from our service. For each example case, we include the size of the reference object in physical units, the total time required to get the output 3D mesh along with the final file size and triangle count of the actual content.

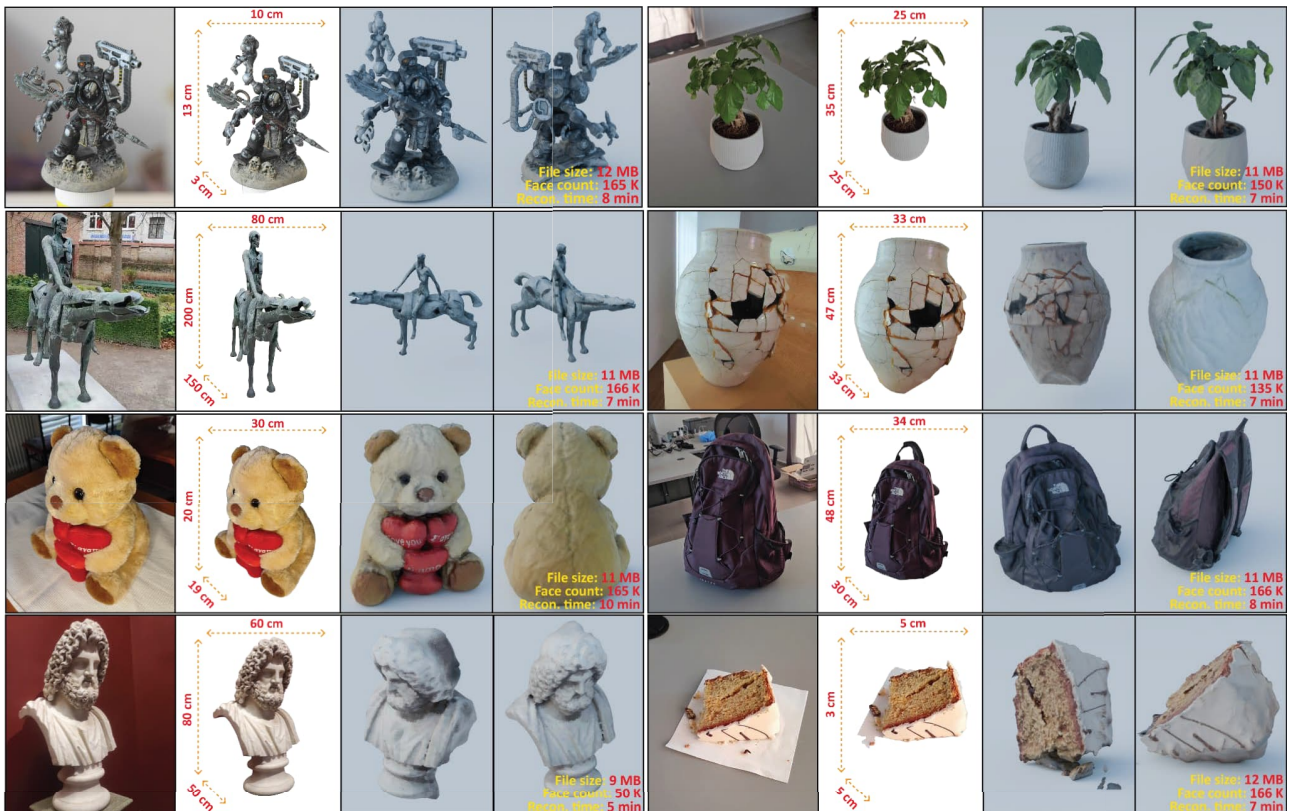


Figure P 11. Indicative outputs from the VISOR service.

Conclusion & Future Work

The VISOR pipeline offers an end-to-end solution for 3D reconstruction from a set of input images or a video stream. While each module is designed with research-driven methodologies, certain limitations remain, primarily due to the quality of input data and implementation-specific assumptions embedded within intermediate stages. For instance, early-stage errors, such as motion blur or insufficient coverage of the target object, can propagate through the pipeline and significantly degrade the fidelity of the final reconstruction. To support accessibility and ease of use, VISOR is deployed as a web-based service, allowing users to upload videos or image collections and receive a reconstructed 3D asset in return. The planned future updates aim to align the pipeline more closely with the latest research. Improvements are anticipated in the calibration stage, enhancing robustness to challenging capture conditions, and in surface appearance modeling to better approximate physically accurate rendering. However, the most critical area of ongoing development is the optimization phase, with a focus on improving the converged SDF quality and reducing the time required to achieve high-quality reconstructions.

Acknowledgements

This work was funded by the European Union under the CORTEX2 project (Grant Agreement no. 01070192).

References

- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99-106. <https://doi.org/10.1145/3503250>
- Pan, L., Baráth, D., Pollefeys, M., & Schönberger, J. L. 2024. Global structure-from-motion revisited. In *European Conference on Computer Vision* (pp. 58-77). Cham: Springer Nature Switzerland.
- Yu, Q., Zhao, X., Pang, Y., Zhang, L., & Lu, H. 2024. Multi-view aggregation network for dichotomous image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3921-3930).
- Miller, B., Chen, H., Lai, A., & Gkioulekas, I. 2024. Objects as volumes: A stochastic geometry view of opaque solids. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 87-97).
- Müller, T., Evans, A., Schied, C., & Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4), 1-15. <https://doi.org/10.1145/3528223.3530127>
- Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., & Liu, L. 2023. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3295-3306).
- Lorensen, W. E., & Cline, H. E. (1998). Marching cubes: A high resolution 3D surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field* (pp. 347-353). <https://doi.org/10.1145/280811.281026>
- Kapoulkine, A. 2025. "Mesh Optimizer", Available: <https://github.com/zeux/meshoptimizer>
- Young, J. 2022. "Xatlas", Available: <https://github.com/jpcy/xatlas>

SAME-XR: A VCS-Integrated Asset Management Pipeline to Bridge the Gap Between 3D Artists and XR Developers

Burak Karaceylan, and Zafer Karadayi

Nara Egitim Teknolojileri A.S., Turkey

Corresponding author: burakkaraceylan@nara.com.tr

Keywords: extended reality, asset management

Introduction

The development of Extended Reality (XR) applications is frequently hindered by workflow friction between 3D artists and software engineers. Recent industry surveys quantify this pain point, revealing that most studios spend more than an hour per week on basic asset management tasks like searching for the latest files (53%) and transforming formats (62%) (Unity Technologies, 2024). This paper introduces SAME-XR (Scalable Asset Management and Conversion Engine for XR), a platform designed to solve these inefficiencies.

The same data reveals a clear demand for specific solutions, with **77% of studios rating version control as a very or extremely important feature** in an asset management system (Unity Technologies, 2024). This insight became the cornerstone of our approach. While an initial proof-of-concept for a centralized platform was successful, we made a strategic pivot to integrate with the tool at the heart of modern development: the **Version Control System (VCS)**. By leveraging GitLab, we created a system that mirrors a familiar Git-based workflow, directly addressing the industry's most-demanded feature.

The Git-Integrated Workflow

Our approach centers on a "Projects microservice" that links a SAME-XR project to a GitLab repository using webhooks. This enables a bi-directional, automated synchronization of "issues," which act as the primary unit for asset requests and tracking. The workflow is as follows:

- 1. Request:** A developer opens an issue in GitLab requesting an asset, applying a designated samexr label.
- 2. Synchronize:** SAME-XR automatically creates a corresponding, linked issue in the design team's project.
- 3. Create & Upload:** An artist creates the model and uploads it directly to the issue using the SAME-XR Blender add-on. The system handles versioning and optimization in the background.
- 4. Resolve & Notify:** When the artist resolves the issue in SAME-XR, the linked GitLab issue is automatically closed, notifying the developer.

- Import:** The developer uses the SAME-XR Unity package to import the final, engine-ready asset directly into their scene with a single click.

This automated process provides a fully traceable audit trail from request to final delivery, managed within the preferred tools of both developers and artists.

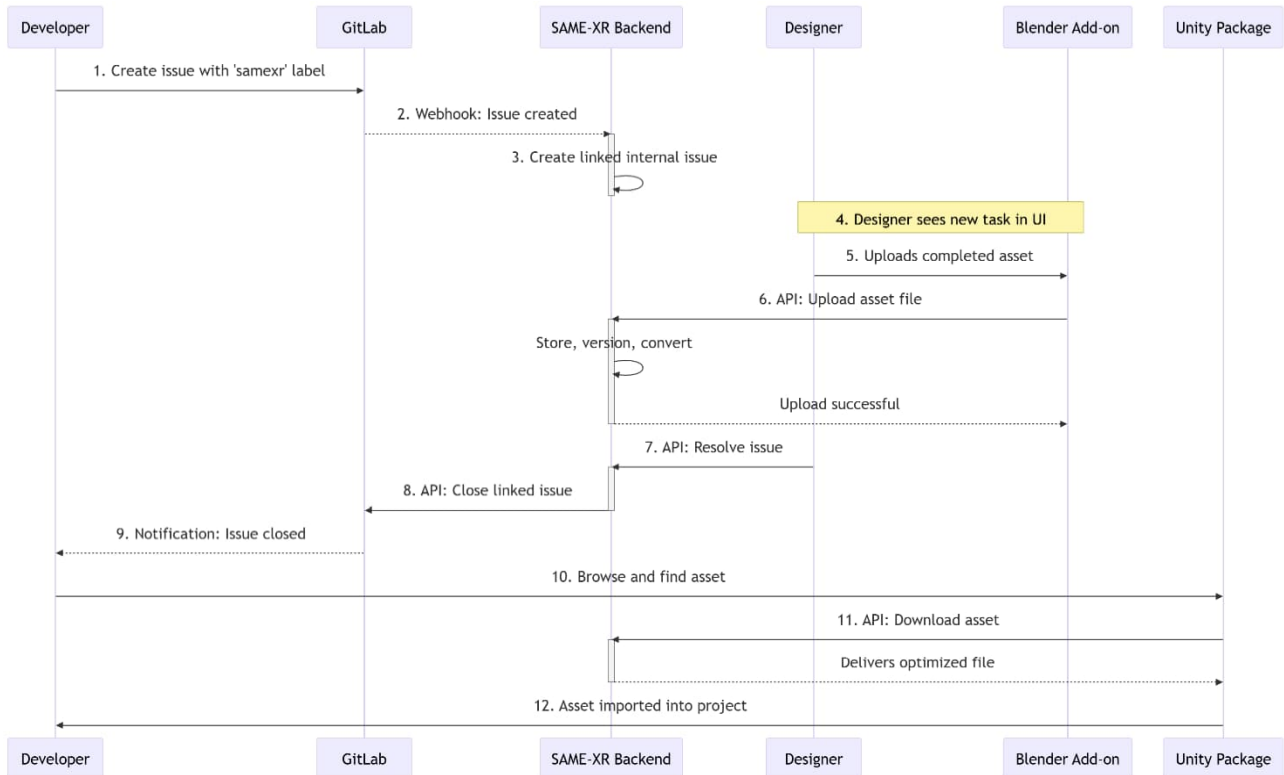


Figure P 12 Sequence diagram of the SAME-XR workflow.

System Architecture and Components

SAME-XR is architected as a set of containerized microservices developed using Golang designed for scalable deployment on Kubernetes. The primary user-facing components are:

- **A Blender Add-on:** Allows 3D artists to manage their assigned issues, view asset details, and upload models directly from their primary design environment.
- **A Unity Package:** Provides an editor window for developers to track asset requests and import completed models directly into the game engine.
- **A Web Dashboard:** Serves as the central hub for project managers to configure GitLab links, oversee all issues, and manage team access.

These components communicate via a suite of RESTful APIs, ensuring consistent and reliable interaction with the backend services.

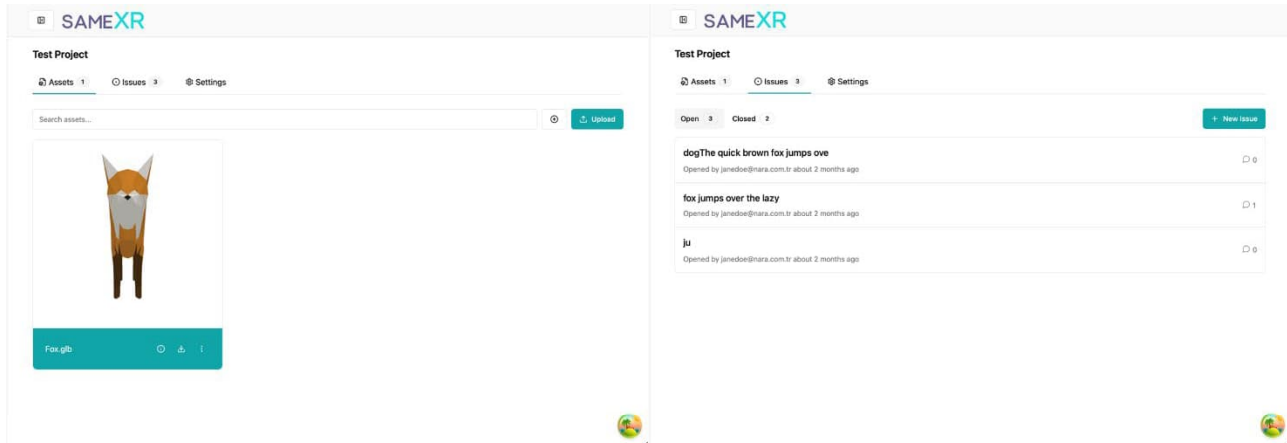


Figure P 13. Screenshots from the web interface.

Results and Future Work

The GitLab-integrated workflow has been successfully implemented. The functional prototypes of the Blender add-on and Unity package enable the core asset upload and import capabilities, proving the viability of our VCS-centric approach. By directly targeting the industry's most-requested feature—version control—and automating time-consuming tasks like format conversion and asset searching, our solution effectively bridges the gap between design and development toolchains.

Future work is focused on three key areas. First, we will refine the component user interfaces based on feedback from testing with CORTEX² partners. Second, we plan to expand VCS integration to include other platforms such as GitHub. Finally, we will continue to enhance the backend file processing capabilities to incorporate more advanced 3D model optimizations

Acknowledgements

This project has been funded by the EU project CORTEX² (GA: 101070192).

References

- Unity Technologies. 2024. 2024 Unity Gaming Report. [online]. Available at: <https://unity.com/resources/gaming-report-2024>, Accessed 21 May 2024
- Karadayı, Z., Özdem, M., Karaceylan, İ.B. & Duman, Z.Y. 2024. Establishing the Future of Work and Collaboration through Institutional Metaverse. 2024 International Wireless Communications and Mobile Computing (IWCMC), 1400–1403. <https://doi.org/10.1109/IWCMC61514.2024.10592490>

MANTRA: Mixed-Reality Augmented Networks for Teleoperated Robotics Applications

Andrés Meseguer Valenzuela¹, Patricia Pons Tomás¹, Samuel Navas Medrano¹, Jose Vera Pérez¹, Jose Luis Soler-Domínguez¹, Marta García Ballesteros¹, David Millán Escrivá¹, Pablo Rocamora Montoro², and Alfredo Giménez Millán²

¹*Instituto Tecnológico de Informática (ITI), Spain*

²*Factor Ingeniería y Decoletaje S.L., Spain*

Corresponding author: ameseguer@iti.es

Keywords: Mixed Reality, Mobile Robot Teleoperation, 5G private networks

Introduction

The MANTRA project aims to develop a Mixed Reality (MR) interface for the teleoperation of industrial mobile robots, integrating multimodal feedback—such as haptic interfaces—to enhance user experience and operational precision. Central to the MANTRA approach is participatory design, involving stakeholders through iterative co-design sessions to ensure the development of a robust and user-friendly system for remote operations.

To enable real-time bi-directional communication, the project also has targeted the implementation of novel software mechanisms that leverage the interoperability of mixed 5G and Wi-Fi networks. Additionally, edge process offloading mechanisms are addressed to reduce computational requirements in mobile robots.

This work presents the main developments achieved in the MANTRA project. It begins by outlining the project's primary objectives, followed by a detailed description of the overall system architecture. Subsequently, the results section highlights key outcomes obtained from real-world validation sessions. Finally, the discussion section offers insights and reflections on the impact and implications of the project's findings.

Objectives

The main objectives of the project are summarized as follows:

1. **Reduce occupational accidentality in the manufacturing industry by integrating novel MR-enabled telepresence tools.** A dedicated MR interface has been designed to ensure an effective and intuitive user experience. Through a co-design process, key visual elements required by

workers teleoperating autonomous mobile robots (AMRs) were identified, along with the necessary sensor inputs to support remote supervision of manufacturing processes.

2. **Enhance 5G and Wi-Fi interoperability.** Reliable communication for mobile robots must be ensured to prevent any degradation in Quality of Experience (QoE). As these devices operate in environments where radio coverage may be inconsistent, maintaining service continuity requires the implementation and configuration of advanced mechanisms that leverage multi-radio connectivity—particularly through the combined use of Wi-Fi and 5G networks.
3. **Enabling XR robot remote control modules into edge platform.** Remote computation of software components is essential to minimize latency in robot control and sensor feedback, given the limited processing capabilities of onboard embedded systems (Meseguer Valenzuela, et al. 2024). To address this, process offloading mechanisms must be defined to enable timely robot operation and support the handling of data from additional sensors, including thermal and depth sensors.

Architecture

The MANTRA project adopts a modular and layered architecture designed to enable safe and efficient remote teleoperation of mobile robots using MR technologies. The solution is structured around three core components: MR Interface, 5G–Wi-Fi Interoperability, and Edge Multimedia Processing.

1. **Mixed Reality User Interface.** The core of MANTRA is a Mixed Reality application for Microsoft HoloLens 2, enabling remote control of Autonomous Mobile Robots (AMRs). Co-designed with stakeholders (Soler-Dominguez, et al. 2025), the interface delivers multimodal feedback—visual, haptic, and auditory—to enhance user awareness and precision. It offers three control modes: an arrow panel for precise navigation, a slider for linear motion, and a joystick for free-range movement.

Users can switch in real time between RGB, thermal, and point cloud visualizations, adapting the view to the task—e.g., detecting overheating with thermal imagery or analyzing spatial constraints with point clouds. The interface integrates with ROS via a ROS# library, subscribing to topics such as /pointcloud and /battery for live sensor and status data.



Figure P 14. Microsoft HoloLens 2 user interface for robot teleoperation.

2. **5G–Wi-Fi Interoperability.** Seamless connectivity is crucial for remote robot teleoperation. MANTRA tackles this through a hybrid 5G/Wi-Fi infrastructure that leverages the ATSSS (Access Traffic Steering, Switching, and Splitting) framework from 3GPP Release 15 to switch between networks and maintain QoE. A dedicated software module monitors connection quality—using RSSI for Wi-Fi and RSRP for 5G—via AT commands or REST APIs, depending on the hardware. To ensure stable behavior near threshold levels, a hysteresis mechanism is implemented. Based

on real-time metrics, the system dynamically selects the most reliable interface, ensuring continuous connectivity in variable radio environments.

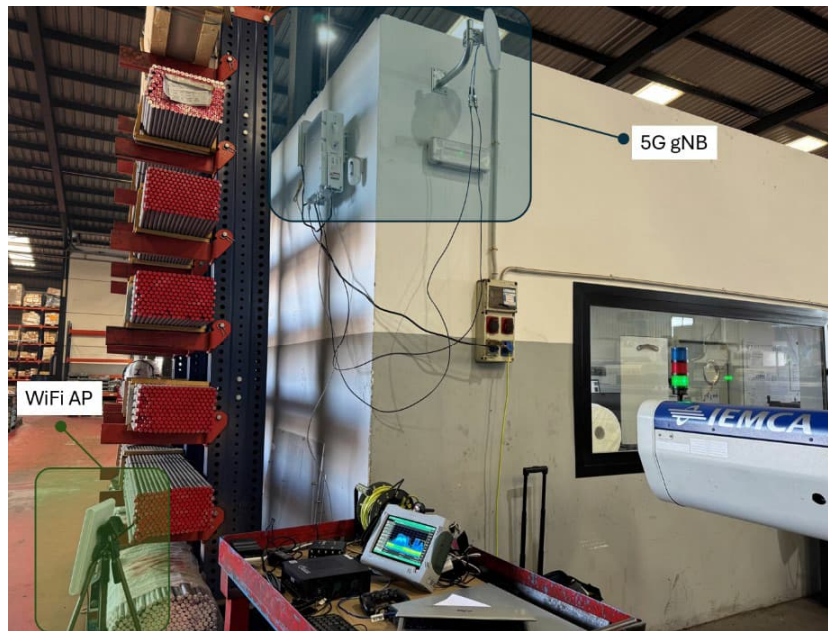


Figure P15. Infrastructure deployment at Factor S.L (end user) premises

- 3. Edge Multimedia Processing and Offloading.** To minimize latency and overcome the limited computing capacity of mobile robots, MANTRA implements edge-based process offloading. Sensor data streams—including RGB, thermal, point cloud, and battery status—are processed in real time and transmitted to external servers or MR devices via ROS and GStreamer. An RTSP-based pipeline, built with MediaMTX¹⁸ and deployed in Docker containers, ensures compatibility and ease of integration. A Python ROS node gathers sensor data, overlays key information on video feeds, and re-publishes them via RTSP. This solution was containerized for deployment in SPIRIT's Trusted Execution Environment.

Results

The MANTRA MR teleoperation system was tested through pilot and formal usability studies involving FACTOR employees and a smaller user group at the Berlin testbed. During the initial MVP testing at ITI's facilities, FACTOR team members explored the MR interface, operated the robot, and provided early feedback. Following this, a usability study was conducted at FACTOR with 10 participants aged 23 to 57, with varying prior XR experience. Users teleoperated the robot via the MR interface and completed three standardized questionnaires assessing satisfaction (USEQ), workload (NASA-TLX), and presence (TPI).

- **User Satisfaction (USEQ)** averaged 24.5/30 (>80%), indicating strong acceptance and positive experience.
- **Workload (NASA-TLX)** showed low frustration, effort, and physical demand. Temporal and mental demands were moderate (around 2.5/5), with performance rated highest (3/5), resulting in an overall moderate to low workload.

¹⁸ n.d. Accessed 07 04, 2025. <https://github.com/bluenvirion/mediamtx>.

- **Presence (TPI)** results reflected high immersion and mental engagement, with visual quality rated strongly. However, sound realism and real-world consistency scored slightly lower, suggesting room for audio and coherence improvements.

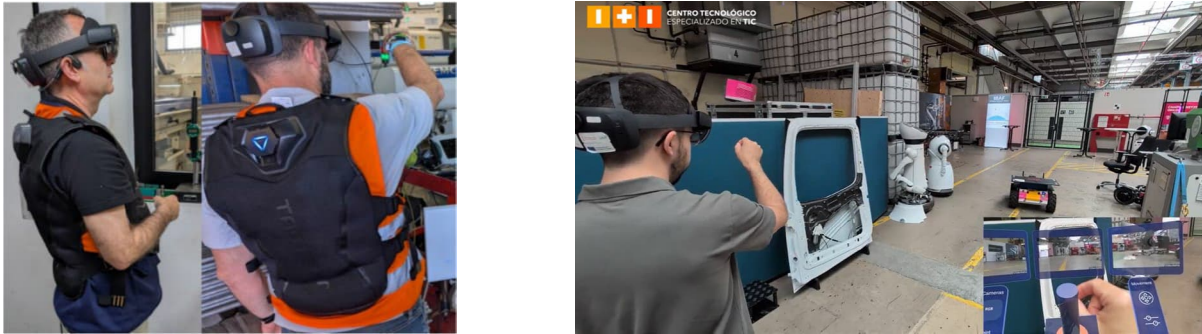


Figure P16. User validation at Factor S.L. premises (left) and DT premises (right)

Overall, results demonstrate that the MR teleoperation system provides compelling and usable immersive experience, with moderate workload and good user acceptance, while indicating key areas for future enhancements in audio realism and comfort. Regarding the results in terms of 5G and Wi-Fi interoperability, in addition to edge process offloading, target values were achieved, demonstrating the correct performance of the solution.

*Table P 2.*Results summary

ID	Description	Target Value	Achievement
KR1	Number of users testing the interface	>10 users	13 users
KR2	Usability and user acceptance questionnaires (3 questionnaires per user)	>80% of user acceptance	81.6%
KR3	Multimodal feedback	3 sources: haptic, visual and audio	3 sources: haptic, visual and audio
KR4	Number of design guidelines	>= 5	5
KR5	Maximum interoperability timeout between 5G and Wi-Fi	40 ms	12ms (average)
KR6	Number of video streams	>=4 x (720p @30fps)	4 x (720p @30fps)
KR7	Number of additional sensor streams	>=3	3 streams: Point cloud, thermal and battery sensor

References

- Meseguer Valenzuela, A., Silvestre-Blanes, J., Sempere-Paya, V.M., & Bartolín-Arnau, L.M. 2024. Multi-Access Edge Computing performance into Non-Public 5G Networks: A robot-based experiment. *IEEE 20th Int Conf Factory Communication Systems (WFCS)*. Toulouse.
- Soler-Dominguez, J.L., Navas Medrano, S., García Ballesteros, M., Pons, P. 2025. *Design guidelines learned from the MANTRA project*. Instituto Tecnológico de Informática, Zenodo. <https://doi.org/10.5281/zenodo.15753852> .

FLYTEX: Merging IoT Data into Videoconferences for A Better Decision-Making Process in Agriculture

Gonzalo Blázquez Gil^{1,2}, Antonio Carreiro Alonso², Pablo M. Fernández Graña²,
and Ana Almécija Pereda²

¹*FlyThing Technologies S.L., Spain*

²*ITG – Instituto Tecnológico de Galicia, Spain*

Corresponding author: apereda@itg.es

Keywords: IoT, augmented reality, extended reality, virtual reality, videoconferences, real-time data

Introduction

Digitalisation is transforming the agriculture sector. In the last few years, innovations based on data, including **advanced IoT systems**, are being deployed in different farms and business, providing **data-driven** insights and enabling farmers to make informed decisions about their crops and farms.

At the same time, extended and augmented reality are increasingly being used in professional and educational environments, offering a wide set of features and better user experience. CORTEX² project aims to bridge the divide between video-conferencing tools and XR-based solutions, to boost the uptake of next-generation extended reality tele-cooperation.

Funded by CORTEX² first open call (co-development track), FLYTEX project goal was to transform video meetings and immersive environments by providing **access to IoT data in videoconferencing** applications, validating this technology in the agricultural sector. This innovative approach contributes to advancing the agrifood sector through the application of IoT and XR technologies, enabling data-driven insights to improve the decision-making process.

The integration of IoT data into an immersive environment provided the agrifood sector into a more technologically advanced, efficient, and sustainable industry. This integration leads to improved decision-making and resource management, supporting long-term goals like widespread adoption of smart farming.

The absence of standardized protocols or adaptors capable of universally interpreting and integrating data from a diverse array of IoT devices is a significant limitation in the agrifood sector. This lack of standardization results in a disjointed process of data utilization, diminishing the efficiency and potential benefits of XR applications in the agrifood industry.

By using Flythings, a **versatile platform** designed for **real-time monitoring** and management in business environments, enables the seamless integration of different types of devices. Flythings can

operate with the most extended communication protocols and is easily integrable with other company systems such as ERPs or SCADAs.

Objectives

- Integration with CORTEX² Framework to enhance XR videoconferencing applications, by displaying real-time environmental data.
- Real-Time Data Integration in XR Environments, enhancing decision-making and operational efficiency through immersive visualization.
- Expand CORTEX² to support a variety of IoT devices and protocols (Zigbee, OPC-UA, Modbus TCP, etc.) used in the agri-food industry.
- Testing and Validation in agri-food contexts, ensuring robustness and applicability in real world agricultural settings

Development

During the project implementation, we designed and implemented an architecture organised in three core functional blocks (Figure P 17).

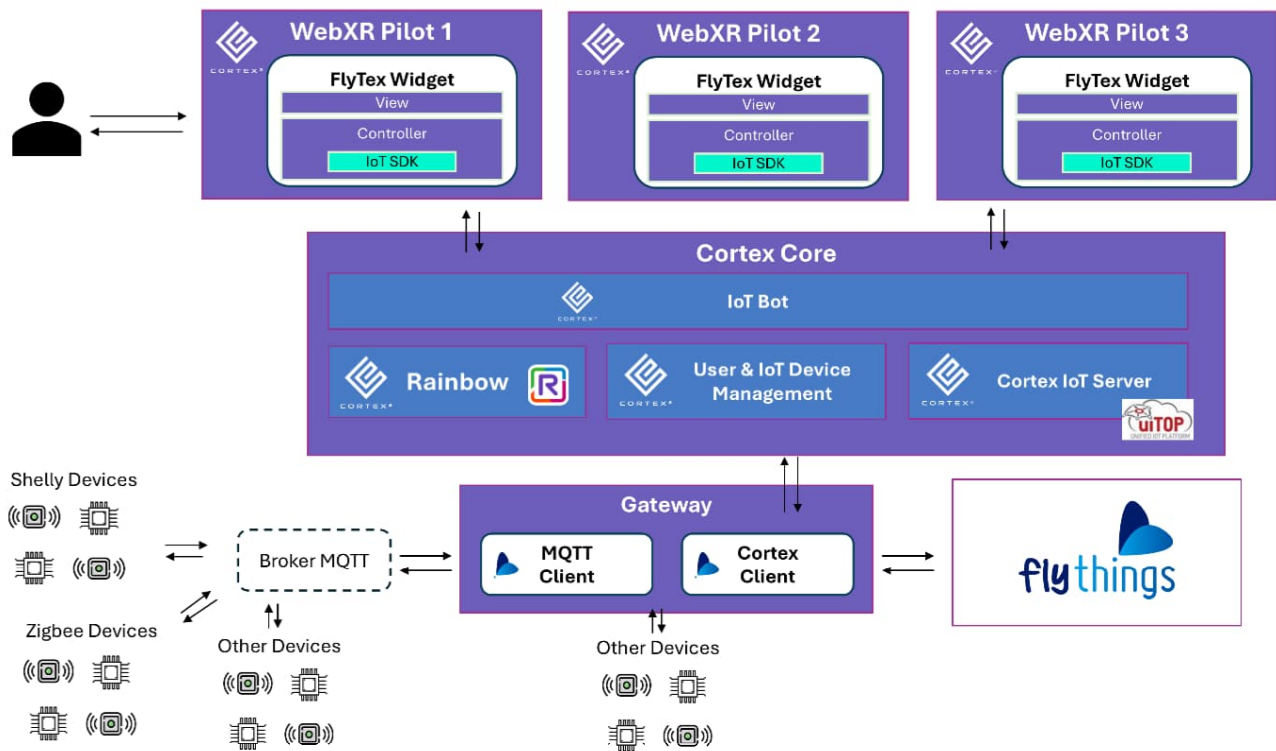


Figure P 17. FLYTEX architecture.

A widget was also developed to monitor and interact with IoT devices in CORTEX² virtual environments through a bubble chat. This widget enables the visualization of real-time and historical data in virtual meeting rooms. These data can be also shared with other meeting rooms by a button in the virtual meeting room.

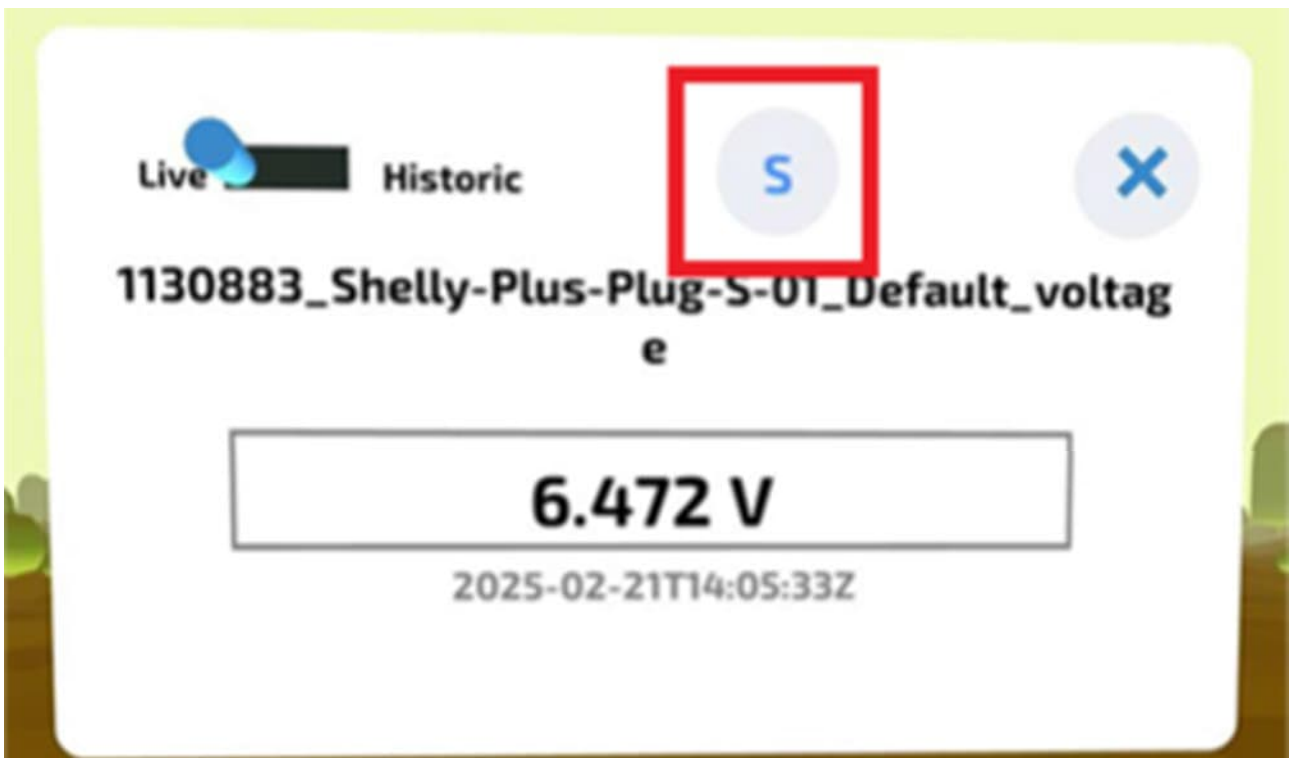
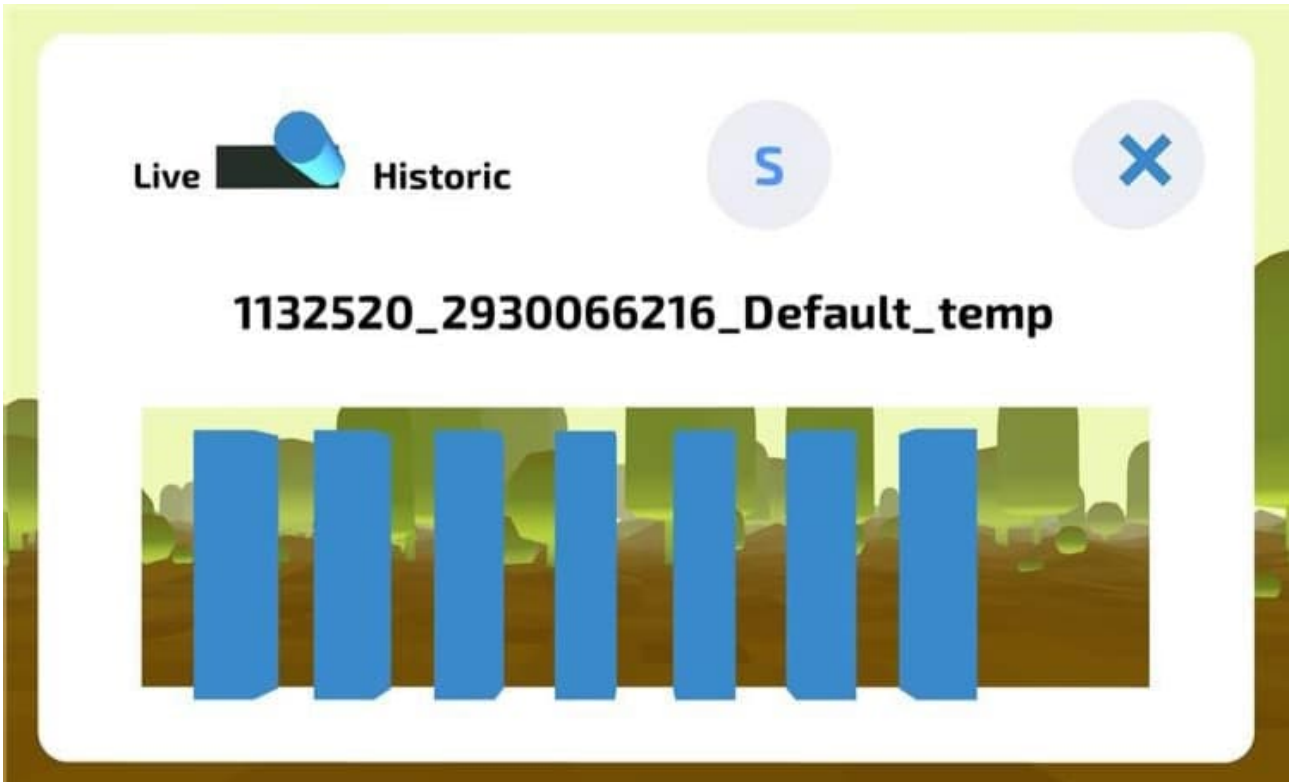


Figure P 18. FLYTEX widget.

The FLYTEX widget was tested using Meta Quest 2 glasses, confirming that both the environment and the widget are fully visible and functional within a real XR setting, beyond a conventional browser experience



Figure P 19. FlyThings Developer testing Flytex in XR glasses

Results and achievements

During the three sprints of FLYTEX, we **successfully deployed and optimised** a working prototype capable of sending real-time data from multiple devices and seamlessly integrating this data into the CORTEX² platform showing the data in a web XR environment.

This prototype not only ensured the accurate and timely transmission of **sensor readings** but also validated the **compatibility of our system with CORTEX²'s** data ingestion pipelines. By utilizing robust communication protocols and ensuring adherence to platform requirements, we laid the groundwork for scalable and efficient device integration.

Figure P 20 shows the KPIs defined in the proposal phase and the results obtained.

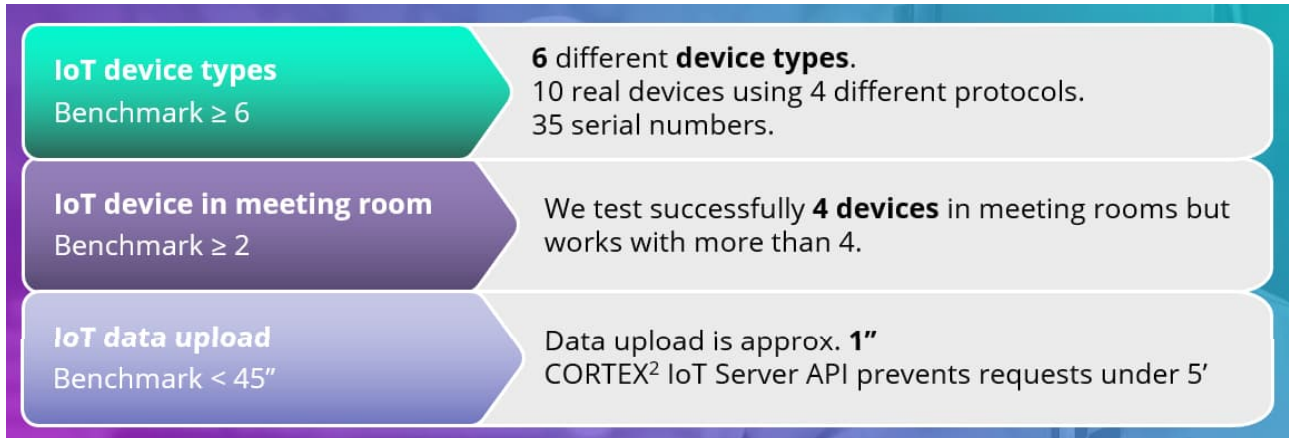


Figure P 20. KPIs

Conclusions

The FLYTEX project has successfully demonstrated the feasibility and **benefits of integrating IoT data into XR videoconferencing** environments, particularly within the agri-food sector. By leveraging the Flythings platform and extending the CORTEX2 framework, the project enabled **real-time data visualization** and interaction during immersive meetings, contributing to more informed and efficient decision-making processes.

Funding

FLYTEX project has indirectly received funding from the European Union's Horizon 2020 research and innovation action programme, via the CORTEX² Open Call 1 issued and executed under the CORTEX2 project (Grant Agreement no. 01070192).

References

- Kosior, K. 2018. Digital transformation in the agri-food sector—opportunities and challenges. *Roczniki (Annals)*, 2018(2). <https://doi.org/10.22004/ag.econ.293647>
- CORTEX2 project. <https://cortex2.eu/>
- Flythings IoT platform. <https://itg.es/en/flythings-industria-4-0/>

Quality of Experience Evaluations for Multi-cloud Streaming

Simone Porcu^{1,2}, Renato Caboni¹, Alessandro Floris^{1,2}, Luigi Atzori^{1,2}

¹*DIEE, University of Cagliari, Italy*

²*CNIT, University of Cagliari, Italy*

Corresponding author: simone.porcu@unica.it

Keywords: Quality of Experience, Point Clouds, Immersive Streaming, eXtended Reality.

Abstract

Point clouds are widely used in immersive multimedia systems because they can be acquired in real time using sensors like LiDAR (Light Detection and Ranging) or Time-of-Flight (ToF) cameras, which makes them ideal for building realistic and interactive 3D experiences. However, their large size introduces challenges for compression and streaming, especially in complex scenes with multiple objects viewed through head-mounted displays (HMDs). This paper presents two experimental applications developed for two recent underinvestigated HMDs: the Apple Vision Pro and Meta Quest 3. The first application allows users to view compressed dynamic point clouds and rate their perceived visual Quality of Experience (QoE) using a simple hand-based interface. The second application focuses on complex 3D scenes with multiple point clouds and supports the development of an adaptive streaming algorithm that optimizes the QoE based on the user's view and system constraints. These tools aim to support future research on immersive content delivery in extended reality (XR), with possible applications in tourism and cultural heritage.

Introduction

Point clouds (PCs) are a key content format in immersive applications thanks to their ability to represent 3D scenes with high realism and freedom of exploration. Each point in a PC includes geometric and visual attributes that enable a full six degrees of freedom (6DoF) experience. However, the size of PC data is typically very large, which makes compression essential. Two standard methods have been defined by MPEG: Video-based Point Cloud Compression (V-PCC) and Geometry-based Point Cloud Compression (G-PCC) (Graziosi et al., 2020).

Although these techniques reduce data volume, they also introduce visual artifacts that can affect the user's Quality of Experience (QoE). Several studies have investigated this trade-off, but mostly by showing single compressed point clouds on 2D screens or on a limited set of head-mounted displays (HMDs) such as Oculus Rift (Subramanyam et al., 2020; Viola et al., 2022), HoloLens (Nguyen et al., 2024; Fan et al., 2024; Nguyen et al., 2023), HTC Vive (Wu et al., 2021; Gutierrez et al., 2023; Alexiou et al., 2020), and Meta Quest 2 (Van Damme et al., 2023). There is still a lack of studies that consider new-generation devices like the Apple Vision Pro and Meta Quest 3. At the same time, current solutions do

not consider scenes composed of multiple PCs, which are common in real applications such as virtual tourism, cultural heritage, or remote collaboration. These multi-object scenes require more advanced streaming strategies. While some recent works proposed adaptive streaming methods that prioritize parts of a PC based on the user's attention (Li et al., 2020 & 2023; Wang et al., 2023), they typically assume the presence of a single PC and do not address the problem of dynamically managing multiple PCs with different levels of detail (LoD).

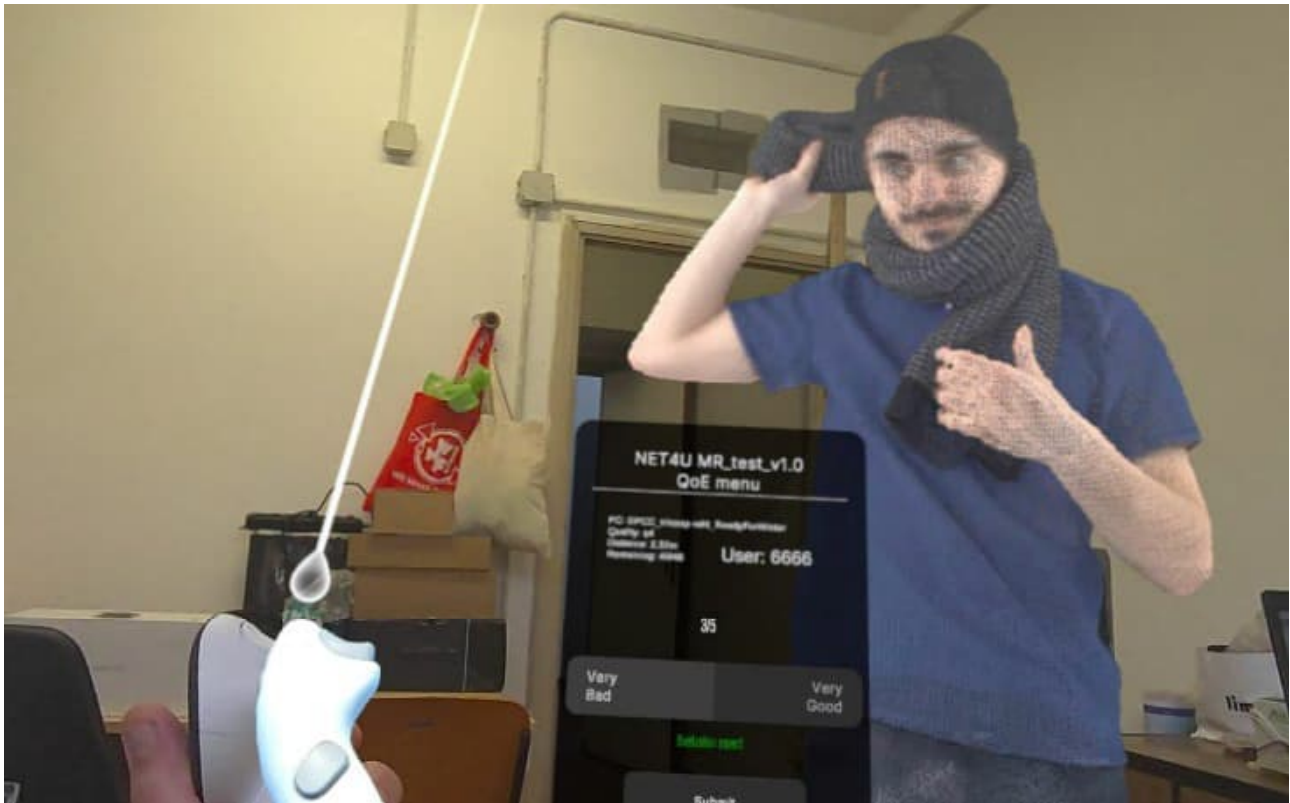


Figure P 21. A point cloud displayed in XR mode using the Meta Quest 3.

In this work, we propose a twofold contribution. First, we developed an application to display and evaluate compressed PCs on the Apple Vision Pro and Meta Quest 3 HMDs. This app allows for subjective assessments on devices that have not yet been considered in previous studies. Second, we developed an additional application, compatible with the two HMDs, designed to render scenes that include multiple PCs compressed at different LoDs and placed at various distances from the user's point of view. This app supports the development of a new adaptive streaming algorithm that adjusts the quality of each object based on user's attention, network conditions, HMD's rendering capabilities, and object distance from the user's point of view. Together, these tools provide a framework for studying QoE in complex immersive environments and improving the efficiency of streaming in extended reality (XR) applications.

Application 1: QoE Evaluation of Compressed Point Clouds on Modern HMDs

The first application allows to understand how users perceive the quality of 3D dynamic PCs. The PC is shown on two of the most advanced headsets currently available: the Apple Vision Pro and the

Meta Quest 3. To this aim, we used the public dataset ComPEQ-MR¹⁹, which includes a variety of PCs previously used for quality assessment in mixed reality scenarios (Nguyen et al., 2024). The PCs selected from the dataset were compressed using the two MPEG-standard methods: G-PCC and V-PCC (Graziosi et al., 2020). To present this content in an immersive environment, we developed a custom cross-platform application in Unity (showed in Figure P 21), compatible with both Apple Vision Pro and Meta Quest 3.

The application was designed to support natural hand interaction without requiring physical controllers. Users can move freely within the XR scene, observe the dynamic PC content, and provide a subjective quality rating using a 5-point Absolute Category Rating (ACR) scale. The interface allows intuitive and direct selection of scores using hand gestures. This setup enables the collection of QoE scores in realistic XR conditions on under-investigated HMDs, contributing to a more comprehensive understanding of how compression impacts user experience in next-generation immersive systems.

Application 2: Streaming of Multiple Point Clouds with Adaptive Quality

The second application focuses on scenes that contain multiple PCs shown at the same time. This type of setup reflects real use cases, such as virtual tours or remote visits to museums and cities. Thus, we created a new dataset that includes 3D PCs of selected artworks (sound stones) by the Sardinian artist Pinuccio Sciola, as illustrated in Figure P 22. These artworks were acquired using high-resolution photogrammetry to preserve both texture and geometric details. After the acquisition, the PCs were compressed using the V-PCC and G-PCC methods, each at different LoD. The compressed PCs were then placed in virtual scenes, with some PCs positioned in the foreground and others in the background. These scenes were used to build new applications for Apple Vision Pro and Meta Quest 3, allowing users to explore complex environments.



Figure P 22. Two PCs representing the selected sound stones created by the artist Pinuccio Sciola.

¹⁹ <https://ftp.itec.aau.at/datasets/ComPEQ-MR/>

The goal of this application is to support the development of a new adaptive streaming algorithm. This algorithm will adjust the quality of each point cloud based on the user's field of view, the number of visible objects, the available network bandwidth, and the rendering capacity of the HMD. The idea is to show high quality PCS only where needed, reducing bandwidth usage while keeping the experience smooth and immersive.

Conclusions

This work presents two experimental applications designed to explore how PC compression and scene complexity affect QoE in XR environments. In both cases, we developed apps for Apple Vision Pro and Meta Quest 3, which are currently among the most widespread HMDs. The first application focuses on the evaluation of single dynamic PCs, allowing users to rate visual QoE using hand gestures. The second application introduces more complex scenes with multiple PCs and supports the design of a new adaptive streaming algorithm. These tools create a solid base for future subjective studies on new-generation headsets and for improving immersive streaming in real use cases such as virtual tourism.

Acknowledgment

This work has been funded by the European Union (SPIRIT, 101070672). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. The SPIRIT project has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

References

- Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., & Tabatabai, A. 2020. An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC). *APSIPA Transactions on Signal and Information Processing*, 9, e13. <https://doi.org/10.1017/ATSIP.2020.12>.
- Li, J., Zhang, C., Liu, Z., Sun, W., Hu, W., & Li, Q. 2020. Demo abstract: Narwhal: A DASH-based point cloud video streaming system over wireless networks. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 1326-1327). IEEE. <https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162937>.
- Li, J., Zhang, C., Liu, Z., Hong, R., & Hu, H. 2022. Optimal volumetric video streaming with hybrid saliency based tiling. *IEEE Transactions on Multimedia*, 25, 2939-2953., <https://doi.org/10.1109/TMM.2022.3153208>.
- Nguyen, M., Vats, S., Zhou, X., Viola, I., Cesar, P., Timmerer, C., & Hellwagner, H. 2024. Compeq-mr: Compressed point cloud dataset with eye tracking and quality assessment in mixed reality. In *Proceedings of the 15th ACM Multimedia systems conference* (pp. 367-373)., <https://doi.org/10.1145/3625468.3652182>.
- Subramanyam, S., Viola, I., Hanjalic, A., & Cesar, P. 2020. User centered adaptive streaming of dynamic point clouds with low complexity tiling. In *Proceedings of the 28th ACM international conference on multimedia* (pp. 3669-3677)., <https://doi.org/10.1145/3394171.3413535>.
- Viola, I., Subramanyam, S., Li, J., & Cesar, P. 2022. On the impact of vr assessment on the quality of experience of highly realistic digital humans: A volumetric video case study. *Quality and User Experience*, 7(1), 3. <https://doi.org/10.1007/s41233-022-00050-3>.
- Fan, C., Zhang, Y., Zhu, L., & Wu, X. 2024. PCQD - AR: Subjective quality assessment of compressed point clouds with head - mounted augmented reality. *Electronics Letters*, 60(5), e13134. <https://doi.org/10.1049/ell2.13134>.

- Nguyen, M., Vats, S., Van Damme, S., van der Hooft, J., Vega, M. T., Wauters, T., ... & Hellwagner, H. (2023). Characterization of the quality of experience and immersion of point cloud videos in augmented reality through a subjective study. *Ieee Access*, 11, 128898-128910. <https://doi.org/10.1109/ACCESS.2023.3326374>.
- Wu, X., Zhang, Y., Fan, C., Hou, J., & Kwong, S. 2021. Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(12), 4630-4644. <https://doi.org/10.1109/TCSVT.2021.3101484>.
- Gutiérrez, J., Dandyeyeva, G., Dal Magro, M., Cortés, C., Brizzi, M., Carli, M., & Battisti, F. 2023. Subjective evaluation of dynamic point clouds: Impact of compression and exploration behavior. In *2023 31st European signal processing conference (EUSIPCO)* (pp. 675-679). IEEE. <https://doi-org/10.13039/501100011033>.
- Alexiou, E., Yang, N., & Ebrahimi, T. 2020. PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-6). IEEE. <https://doi.org/10.1109/QoMEX48832.2020.9123121>.
- Wang, L., Li, C., Dai, W., Li, S., Zou, J., & Xiong, H. 2022. QoE-driven adaptive streaming for point clouds. *IEEE Transactions on Multimedia*, 25, 2543-2558., <https://doi.org/10.1109/TMM.2022.3148585>.
- Van Damme, S., Mahdi, I., Ravuri, H. K., van der Hooft, J., De Turck, F., & Vega, M. T. 2023. Immersive and interactive subjective quality assessment of dynamic volumetric meshes. In *2023 15th international conference on quality of Multimedia experience (QoMEX)* (pp. 141-146). IEEE., <https://doi.org/10.1109/QoMEX58391.2023.10178610>.

HOPE: Holographic Optimized Processing Engine

Vasileios-Ektoor Kotsis-Panakakis², Iordanis Evangelou², Fotios Bistas²,
Anastasios Gkaravelis¹, Nikolaos Vitsas¹, Andreas-Alexandros Vasilakis¹,
and Georgios Papaioannou²

¹Phasmatic, Greece

²Athens University of Economics & Business, Greece²

Corresponding author: iordanise@aueb.gr

Keywords: adaptive sampling, 3D compression, holographic communication, telepresence

Introduction

Real-time holographic communication is rapidly transforming the way people interact remotely by enabling immersive, lifelike 3D representations of individuals in motion. This technology goes beyond conventional video calls by capturing and streaming dynamic human features in three dimensions, delivering a strong sense of physical presence. These capabilities open new possibilities for a wide range of applications, from remote collaboration to social interaction, education, and virtual training experiences. As volumetric video technologies continue to evolve, several key challenges have emerged, driving ongoing research and innovation. A central requirement is the efficient delivery of high-bandwidth, data-rich content suitable for real-time telepresence applications based on fluctuating network conditions and the processing capabilities of end-user devices. These challenges lie at the core of making holographic communication scalable, accessible, and seamless across platforms and use cases. In this work, we present HOPE (Holographic Optimized Processing Engine), an ongoing research effort in advancing real-time holographic communication through a suite of context-aware technologies. HOPE employs importance-driven, semantic-aware filtering to preserve critical details while degrading less relevant data. In this scenario, gesture recognition is utilized to dynamically highlight points of interest enabling more intuitive interaction and enhancing the telepresence realism. By integrating smart tuning of efficient mesh compression methods, HOPE aims to offer high-quality 3D streaming under bandwidth constraints. A dynamic adaptation mechanism will fine-tune processing parameters in real time based on device capabilities and network conditions. Together, these components position HOPE as a promising enabler of scalable, bandwidth-efficient holographic experiences with high visual fidelity and responsive performance, particularly in interactive use cases. This poster/demo presents ongoing work and early insights into the system's design and potential.

Methodology

The HOPE architecture is structured around a producer-consumer pipeline, where the end-to-end process begins with the capture of a human subject's 3D shape and motion using a depth camera. The choice of camera setup depends on the positional requirements of the application. The captured raw data is then processed into a point cloud-based volumetric video. The data generated undergoes preprocessing steps, including background removal, tracking, segmentation and compression, to reduce data size while retaining visual fidelity. Rather than relying on uniform data strategies, HOPE project focuses on context-aware sampling and adaptive compression, preserving crucial details (such as user suggestion areas) while discarding redundant or less important data. The processed 3D content is then packaged and transmitted in real time to the consumer side. Upon reception, the data are decoded, reconstructed into 3D form, and rendered for visualization on the user's XR device. Currently, the pipeline supports unidirectional (asymmetric) communication, where one participant is viewed by many, enabling real-time, one-to-many holographic streaming conferencing.

Depth Camera Capturing. Our system captures 3D data using a single Intel RealSense Depth Camera D435i, positioned to provide a frontal view of the user. However, the generation of a fused point cloud from multiple depth cameras is inherently supported. The depth stream is used to generate a point cloud, where the spatial position of each point is calculated based on depth measurements. Simultaneously, the color stream is utilized to map RGB values to the corresponding points in the cloud, resulting in a colorized 3D representation of the subject.

Preprocessing. To manage the large volume of 3D data produced by depth sensing, HOPE incorporates a multi-step preprocessing pipeline designed to adapt to varying client conditions and prioritize meaningful visual content. The first step applies a distance-based filtering to remove points located beyond a depth threshold from the camera (De Fré et al., 2024). This early filtering significantly reduces the raw point cloud size by excluding irrelevant background data, allowing subsequent processes to focus on the subject. The main approach to reducing data complexity is a flexible, context-aware segmentation method. The point cloud is divided into N independent descriptions corresponding to objects and regions detected by neural image analysis (Wang et al., 2025, Nikhila et al., 2024). This segmentation is further refined by incorporating gesture recognition, enabling users to intuitively select or highlight important objects and areas (see Figure P 23, second column). These segmented descriptions can then be processed into different quality presets and selectively streamed, prioritizing regions of interest for enhanced visual fidelity. Optionally, when required due to bandwidth constraints, uniform sampling may be applied as a simple downsampling step (De Fré et al., 2024). This combination of distance filtering, semantic segmentation, tracking and gesture-driven selection enables adaptive, bandwidth-efficient streaming that preserves critical details while responding to user intent.

Compression. After segmentation, each cluster is smartly tuned and compressed individually using advanced mesh compression techniques such as Draco and MeshOpt. Different compression profiles (low, medium, and high priority) have been identified, with selected parameters optimized for each case. This targeted approach allows the system to preserve essential details in critical regions while minimizing data size for less important areas. By applying this compression strategy selectively, HOPE aims to achieve efficient bandwidth usage without sacrificing visual quality.

Communication. To achieve low-latency, real-time transmission between the capturing and rendering components, the system employs WebSockets. In the current implementation, WebSockets is used for one-way communication, where 3D mesh data flows from the producer (capture side) to the consumer (rendering side) without feedback. In future iterations, the client will be able to transmit

feedback messages to the server, enabling adaptive quality control based on the estimated available bandwidth.

VR Rendering. The 3D data are finally transmitted to the consumer-side, where it is decoded and reconstructed into a full 3D mesh representation of the subject. This reconstructed point cloud is then rendered in a virtual environment for immersive visualization on the user's XR device. The rendering has been implemented using WebXR in combination with Three.js, enabling cross-platform compatibility and seamless deployment within web-based XR experiences. The solution has been successfully tested on a Meta Quest 3.

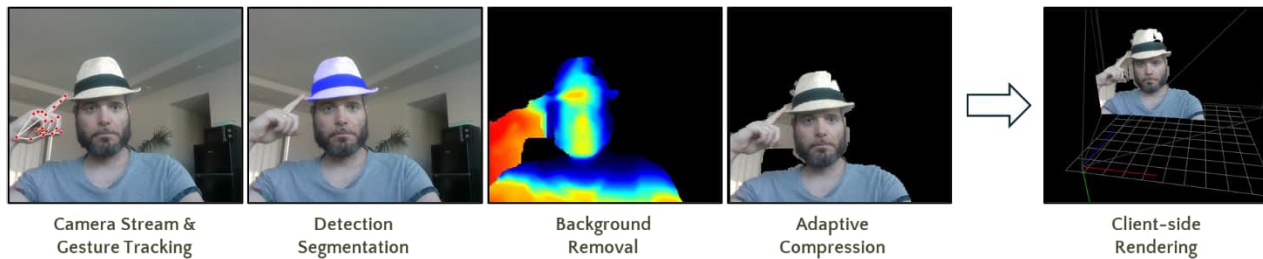


Figure P 23. HOPE pipeline in detail

Conclusion & Future Work

HOPE represents a step toward realizing the vision of the "holographic internet," where real-time 3D communication becomes a natural part of digital interaction. This ongoing work lays the foundation for a technical framework that supports immersive, scalable, and bandwidth-efficient holographic experiences. Looking ahead, we aim to enhance HOPE's capabilities through several key development areas: integrating multi-camera capture setups to improve 3D reconstruction accuracy and depth perception, optimizing client-side rendering for resource-constrained devices by reducing latency (for example supporting adaptive rendering strategies (Gourlay & Held, 2017), and expanding device compatibility across a wider range of VR headsets. These directions will further strengthen HOPE's role as a versatile solution for next-generation interactive holographic communication.

Acknowledgements

This work was funded by the European Union under the SPIRIT project (Grant Agreement no. 101070672).

References

- Wang, A., Liu, L., Chen, H., Lin, Z., Han, J., & Ding, G. 2025. Yoloe: Real-time seeing anything. arXiv preprint arXiv:2503.07465. <https://doi.org/10.48550/arXiv.2503.07465>
- Ravi, N., Gabeur, V., Hu, Y. T., Hu, R., Ryali, C., Ma, T., ... & Feichtenhofer, C. (2024). Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714. <https://doi.org/10.48550/arXiv.2408.00714>
- De Fré, M., van der Hooft, J., Wauters, T., & De Turck, F. (2024, April). Scalable mdc-based volumetric video delivery for real-time one-to-many webrtc conferencing. In Proceedings of the 15th ACM multimedia systems conference (pp. 121-131). <https://doi.org/10.1145/3625468.3647617>
- Gourlay, M. J., & Held, R. T. (2017). Head - Mounted - Display tracking for augmented and virtual reality. Information Display, 33(1), 6-10. <https://doi.org/10.1002/j.2637-496X.2017.tb00962.x>

On the Skinning of Gaussian Avatars

Nikolaos Zioulis, Nikolaos Kotarelas, Georgios Albanis, Spyridon Thermos,
and Anargyros Chatzitofis

Moverse, Greece

Corresponding author: nick@moverse.ai

Keywords: Gaussian Avatars, Quaternion Averaging, Linear Blend Skinning, Reconstruction

Introduction

Gaussian splatting (Kerbl et al., 2023) based methods have recently been used to reconstruct human avatars, showing that we can significantly downscale the systems needed for creating animated human avatars. They are typically implemented using forward skinning from the canonical to the observation space. However, the linear blend skinning (LBS) required for the deformation of the Gaussians does not provide valid results for their non-linear rotation properties.

To overcome this, earlier works heavily relied on neural networks to predict the Gaussian rotation (Hu et al., 2024; Kocabas et al., 2024; Li et al., 2023; Moreau et al., 2024; Pang et al., 2024), or predict rotation offsets to correct the LBS rotations (Li et al., 2023; Qian et al., 2024; Zielonka et al., 2023). More recent works employ the rotation information [Paudel et al., 2024; Shao et al., 2024] embedded in the mesh structure via the face normal direction and rotate the Gaussian's accordingly. But these result in either the loss of rotation degrees of freedom, suffer from interaction with the position updates or introducing complexity when animating and rendering.

Regarding view dependent effects, as enabled by the use of spherical harmonics (SH), most works on LBS-based Gaussian avatar works only predict the constant zero-order coefficients or color directly (Hu et al., 2024; Moreau et al., 2024; Pang et al., 2024; Shao et al., 2024; Zielonka et al., 2023), or use neural networks to predict higher order coefficients (Kocabas et al., 2024). Still, the models would need to be conditioned on pose to be able to learn the necessary animation invariance required to properly model the SH consistently across poses. To address this, a group of prior works have been transforming the view direction for each Gaussian using the rotation of the LBS transform (Jung et al., 2023; Liu et al., 2024; Qian et al., 2024). Still, as aforementioned, the LBS rotation is not proper, leading to inconsistently transformed directions.

Instead, we employ quaternion averaging to perform weighted rotation blending using the Gaussians' skinning weights and then use that blended rotation to rotate the SH coefficients. The resulting method essentially employs a vertex-based Gaussian animation technique, one that is highly parallel, agnostic to the Gaussian rasterizer used, and does not require modifications to the rasterizer

itself. This leads to a simple Gaussian avatar representation that can be more efficiently used in XR contexts.

Approach

Gaussian Splatting (GS) represents scenes with a set of N unorganized 3D Gaussians $\{G_i \mid i \in \llbracket 1, \dots, N \rrbracket\}$, where each one is defined as: $G_i = \{\mu_i, \mathbf{q}_i, \sigma_i, \alpha_i, \mathbf{s}_i\}$, where $\mu \in \mathbb{R}^3$ is the position, $\mathbf{q} \in \mathcal{H}$ is its rotation represented as a quaternion, $\sigma \in \mathbb{R}^3$ is its scale, $\alpha \in (0, 1)$ is its opacity, and $\mathbf{s} \in \mathbb{R}^{3 \times 3}$ are third order SH coefficients for each color channel. The rotation and scaling properties define the Gaussian's covariance: $\Sigma = \mathbf{R} \mathbf{S} \mathbf{S}^T \mathbf{R}^T$, when given in their matrix forms $\mathbf{S} := \text{diag}(\sigma)$ and $\mathbf{R} := \mathcal{R}(\mathbf{q}) \in \mathbb{S}\mathbb{O}^3$, with \mathcal{R} representing the quaternion to rotation matrix conversion. Using the covariance, its Gaussian's spatial extent is defined using its spatially varying opacity: $\varrho(x) = \alpha e^{\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)}$. The Gaussians are splatted using the Elliptical Weighted Average (Zwicker et al., 2002) after alpha blending the depth sorted splats in a volume-like integration manner.

However, this representation is for static scenes, so when considering animated avatars, the Gaussians need to be likewise animated. Using a parametric body model as a canonical representation [38, 50], each Gaussian is initially defined in the canonical frame, and accompanied by a skinning weights vector $w_i \in \mathbb{R}^J$, s.t. $w_i^j \geq 0$ and $\sum_{j=0}^J w_i^j = 1$, where J is the number of joints describing the body's articulation. Using a specific pose at time t , and via a forward kinematics, we derive the body's bones transformations set $\{\mathcal{T}_j^t := \begin{bmatrix} \mathbf{R}_j^t & \mathbf{t}_j^t \\ \mathbf{0} & 1 \end{bmatrix} \in \mathbb{S}\mathbb{E}^3, j \in \llbracket 1, \dots, J \rrbracket\}$. The weights w_i provide an association of each Gaussian with the bones, and its positions are deformed using linear transform blending: $\mu_i^t = \mathcal{T}_i^t \mu_i$, $\mathcal{T}_i^t = \sum_{j=0}^J \mathcal{T}_j^t w_i^j$, with \mathcal{T}_i^t being the linearly blended transforms using the skinning weights w_i .

The LBS transform \mathcal{T} is a standard technique for animating positions, although there exists prior work that explores optimizing the skinning weights to improve realism. It is also typically used to rotate the Gaussian's rotation component and, in some cases, the first order SH coefficients, and/or transform the viewing direction to the canonical coordinate system. But the problem is that linearly blending rotations does not produce a valid rotation for a large portion of typical skinning weights. While this results in volume loss in certain configurations for the positions, the effect is more pronounced when only considering the rotation component, making it inadequate for applying it to the rotation-aware Gaussian components, \mathbf{q} and \mathbf{s} .

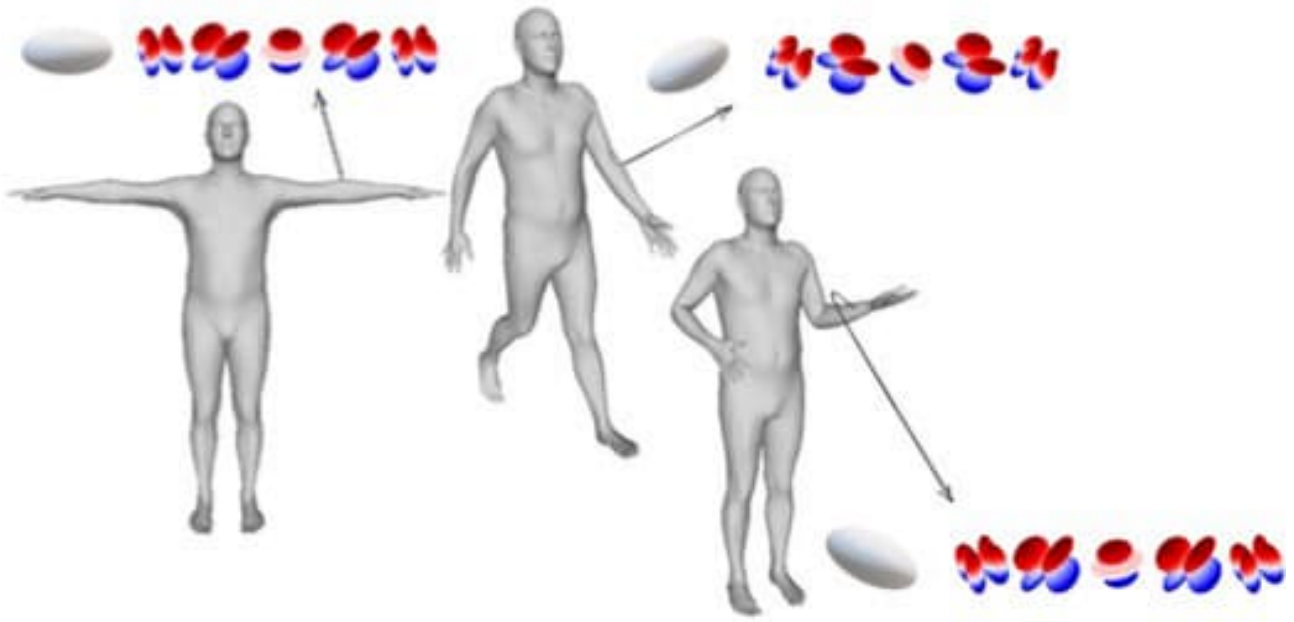


Figure P 24. Our approach properly skinning the non-linear Gaussian components, the rotation and SH coefficients.

One solution to this problem is weighted rotation averaging, and specifically quaternion averaging (Markley et al., 2007). With \mathcal{R}^{-1} being the rotation matrix to quaternion conversion, we extract the rotational component of the bone transforms \mathbf{T}_j^t as a quaternion $b_j^t = \mathcal{R}^{-1}(\mathbf{R}_j^t)$. To calculate the weighted rotation average for each Gaussian G_i we need to construct the matrix $\mathbf{A}_i^t = \sum_{j=0}^J w_i^j (b_j^t)^T b_j^t$. The weighted average quaternion \bar{q}_i^t is the eigenvector corresponding to the maximum eigenvalue of \mathbf{A} . Then for each pose t we can properly skin and rotate each Gaussians G_i rotation component $q_i^t = \bar{q}_i^t q_i$. Having obtained a properly skinned rotation, it can also be used to rotate the SH coefficients. This can be achieved with the Wigner D-matrices $\mathcal{W}^l(\mathbf{R})$ (Gilmore et al., 2008) which map elements \mathbf{R} of $\mathbb{S}\mathbb{O}^3$ to $(2l + 1) \times (2l + 1)$ matrices. These can be used to rotate each SH order's l vector of coefficients. Considering the scale σ and opacity α are rotation invariant properties, the remaining Gaussian components, position μ , rotation q , and SH coefficients s , are now properly skinned, with LBS for the former and quaternion averaging for the non-linear latter two components.

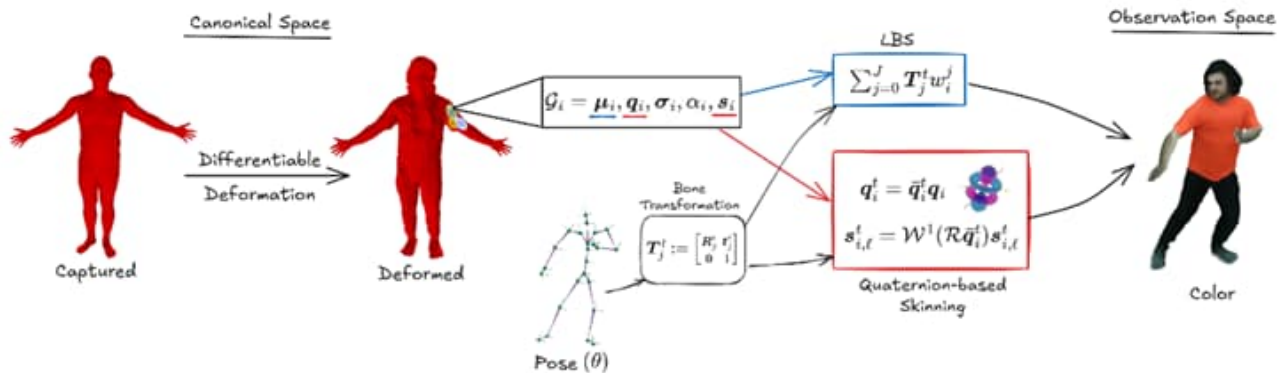


Figure P 25. An efficient Gaussian Avatar pipeline with no learnable parameters apart from the Gaussian components.

Results

We integrate our skinning formulation into existing works and present the results in terms of visual quality improvements in the following tables; each trained on the corresponding datasets the original work was trained on. We also implement a custom Gaussian Avatar reconstruction method (Figure P 25) that relies on per-vertex skinning that renders efficiently using compute shaders in Unity3D and can be used in constrained XR environments, achieving 120 FPS rendering for 4 simultaneous avatars on an Oculus and over 200 FPS for 6 avatars on laptop (GTX 3080 Ti mobile). To achieve consistent parallel shader performance, we use the iterative power method for eigenvalue decomposition.

Table P 3 Comparison using HUGS (Kocabas et al., 2024) on the NeuMan dataset.

Seq.	Lab		Bike		Seattle		Citron		Jogging	
	Original	Ours	Original	Ours	Original	Ours	Original	Ours	Original	Ours
PSNR ↑	22.07	22.38	21.51	21.79	19.70	19.81	21.66	22.12	18.48	18.63
SSIM ↑	0.7952	0.7969	0.7696	0.7701	0.7052	0.7088	0.7706	0.7778	0.6800	0.6816
LPIPS ↓	0.1769	0.1670	0.1841	0.1773	0.1648	0.1503	0.1249	0.1199	0.2218	0.2099

Table P 4. Comparison using iHuman (Paudel et al., 2024) on the PeopleSnapshot dataset.

Seq.	male-3		female-3		male-4		female-4	
	Original	Ours	Original	Ours	Original	Ours	Original	Ours
PSNR ↑	27.82	27.89	21.99	22.26	25.30	25.46	26.89	26.92
SSIM ↑	0.9658	0.9660	0.9392	0.9405	0.9511	0.9529	0.9622	0.9624
LPIPS ↓	0.0198	0.0206	0.0505	0.0466	0.0369	0.0361	0.0244	0.0230

Table P 5 Comparison using SplattingAvatar (Shao et al., 2024) on the PeopleSnapshot dataset.

Seq.	male-3		female-3		male-4		female-4	
	Original	Ours	Original	Ours	Original	Ours	Original	Ours
PSNR ↑	29.83	29.97	28.28	28.44	28.26	28.36	29.26	29.46
SSIM ↑	0.9729	0.9746	0.9714	0.9721	0.8712	0.9724	0.9705	0.9719
LPIPS ↓	0.0710	0.0704	0.0819	0.0825	0.0900	0.0896	0.0711	0.0711

Table P 6 Comparison using Animatable 3D Gaussians (Liu et al., 2024) on the GalaBasketball dataset

Seq.	Dribble		Idle		Shot		Turn	
	Original	Ours	Original	Ours	Original	Ours	Original	Ours
PSNR ↑	37.56	37.84	38.72	38.60	38.95	39.08	37.57	37.16
SSIM ↑	0.9926	0.9929	0.9958	0.9958	0.9945	0.9947	0.9952	0.9952
LPIPS ↓	0.0046	0.0045	0.0034	0.0032	0.0043	0.0045	0.0037	0.0036

Acknowledgements

Funded by the European Union (SPIRIT, 101070672). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the EU. Neither the EU nor the granting authority can be held responsible for them. The SPIRIT project has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

References

- Kerbl, B., Kopanas, G., Leimkühler, T. and Drettakis, G., 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4), pp.139-1.
- Hu, L., Zhang, H., Zhang, Y., Zhou, B., Liu, B., Zhang, S. and Nie, L., 2024. Gaussianavatar: Towards realistic human avatar modeling from a single video via animatable 3d gaussians. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 634-644).
- Kocabas, M., Chang, J.H.R., Gabriel, J., Tuzel, O. and Ranjan, A., 2024. Hugs: Human gaussian splats. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 505-515).
- Li, Z., Sun, Y., Zheng, Z., Wang, L., Zhang, S., & Liu, Y. (2023). Animatable and relightable gaussians for high-fidelity human avatar modeling. *arXiv preprint arXiv:2311.16096*. <https://doi.org/10.48550/arXiv.2311.16096>
- Moreau, A., Song, J., Dharmo, H., Shaw, R., Zhou, Y. and Pérez-Pellitero, E., 2024. Human gaussian splatting: Real-time rendering of animatable avatars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 788-798).
- Pang, H., Zhu, H., Kortylewski, A., Theobalt, C. and Habermann, M., 2024. Ash: Animatable gaussian splats for efficient and photoreal human rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1165-1175).
- Li, M., Tao, J., Yang, Z., & Yang, Y. (2023). Human101: Training 100+ fps human gaussians in 100s from 1 view. *arXiv preprint arXiv:2312.15258*. <https://doi.org/10.48550/arXiv.2312.15258>
- Qian, Z., Wang, S., Mihajlovic, M., Geiger, A. and Tang, S., 2024. 3dgs-avatar: Animatable avatars via deformable 3d gaussian splatting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5020-5030).
- Zielonka, W., Bagautdinov, T., Saito, S., Zollhöfer, M., Thies, J. and Romero, J., 2023. Drivable 3d gaussian avatars. *arXiv preprint arXiv:2311.08581*. <https://doi.org/10.48550/arXiv.2311.08581>
- Paudel, P., Khanal, A., Paudel, D.P., Tandukar, J. and Chhatkuli, A., 2024, September. ihuman: Instant animatable digital humans from monocular videos. In *European Conference on Computer Vision* (pp. 304-323). Cham: Springer Nature Switzerland.
- Shao, Z., Wang, Z., Li, Z., Wang, D., Lin, X., Zhang, Y., Fan, M. and Wang, Z., 2024. Splattingavatar: Realistic real-time human avatars with mesh-embedded gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1606-1616).
- Jung, H., Brasch, N., Song, J., Perez-Pellitero, E., Zhou, Y., Li, Z., Navab, N. and Busam, B., 2023. Deformable 3d gaussian splatting for animatable human avatars. *arXiv preprint arXiv:2312.15059*. <https://doi.org/10.48550/arXiv.2312.15059>
- Liu, Y., Huang, X., Qin, M., Lin, Q. and Wang, H., 2024, October. Animatable 3d gaussian: Fast and high-quality reconstruction of multiple human avatars. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 1120-1129).
- Zwicker, M., Pfister, H., Van Baar, J. and Gross, M., 2002. EWA splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3), pp.223-238.
- Markley, F.L., Cheng, Y., Crassidis, J.L. and Oshman, Y., 2007. Averaging quaternions. *Journal of Guidance, Control, and Dynamics*, 30(4), pp.1193-1197. <https://doi.org/10.2514/1.28949>
- Gilmore, R., 2008. *Lie groups, physics, and geometry: an introduction for physicists, engineers and chemists*. Cambridge University Press.

Promoting Pro-Environmental Behaviour through Immersive Animal Perspective-Taking

Yue Li¹, Marcello A. Gómez-Maureira², Stéphanie M. van den Berg¹,
and Funda Yildirim¹

¹University of Twente/ Faculty of Behavioural, Management and Social Sciences (BMS), The Netherlands

² University of Twente/ Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS),
The Netherlands

Corresponding author: Yue.li@utwente.nl

Keywords: Perspective-taking; Virtual Reality; Pro-environmental Behaviour; Donation Behaviour; Animal; Empathy

Abstract

Despite increasing environmental awareness, translating this concern into consistent sustainable behavior remains limited, a challenge often referred to as the environmental attitude–behavior gap (Borges-Tiago et al., 2024). Visual perspective-taking within immersive environments presents a promising intervention strategy for reducing psychological distance between individuals and environmental circumstances by enhancing feelings of belongingness and emotional arousal, thereby potentially facilitating the translation from understanding to action (Arendt & Matthes, 2016; Van Loon et al., 2018). Virtual Reality (VR) technology, with its inherent capacity for high immersion and role adoption, presents natural advantages for implementing such interventions (Ahn et al., 2016; Akerlof et al., 2013; Erisen et al., 2024).

This study explores how animal perspective-taking experiences in VR influence pro-environmental attitudes and behavioural changes. To identify the specific effects of VR technology, a video-watching group served as the control condition. Participants in the intervention adopted a deer's perspective, experiencing its daily life through activities such as wandering, foraging, and social interaction while witnessing environmental challenges including forest fires and human rescue efforts (**Figure P 26**).

Our experiment (N = 69) employed a between-subjects design with 35 participants in the VR experimental group and 34 in the video-watching control group. We measured self-reported animal and ecological behaviour attitudes at pre-intervention and one-week post-intervention. Immediately following the intervention, empathy and feelings of presence were assessed (Healey & Grossman, 2018; Melo et al., 2023). After the session, participants received €5.00 reward (ten 50-cent coins) and were given the opportunity to donate any portion to local or global environmental protection organizations, with donation behaviour recorded as a key behavioural measurement.

Results showed that both VR and video groups reported increased environmental attitudes one-week post-intervention, indicating lasting behavioural influence ($F(1, 122) = 2313.90, p < .001$). However, the immersive VR experience produced significantly stronger behavioural outcomes compared to traditional video format. Specifically, the VR group demonstrated significantly higher donation amounts ($\beta = 0.24, p = .048$) and greater role adoption levels ($\beta = 0.56, p = .02$) than the control group. Meanwhile, VR increased feeling of Inclusion through presence - when accounting for presence, the direct VR effect disappeared ($p = .714$), with presence explaining 79.6% of the total effect ($p = .014$).

These preliminary findings suggest that while both stimuli effectively produced pro-environmental attitude changes, immersive ecological perspective-taking more effectively promotes concrete pro-environmental behavioural outcomes. Future research should explore how varying levels of interactivity and enhanced sensory feedback mechanisms within VR environments might further amplify these prosocial behavioural effects.

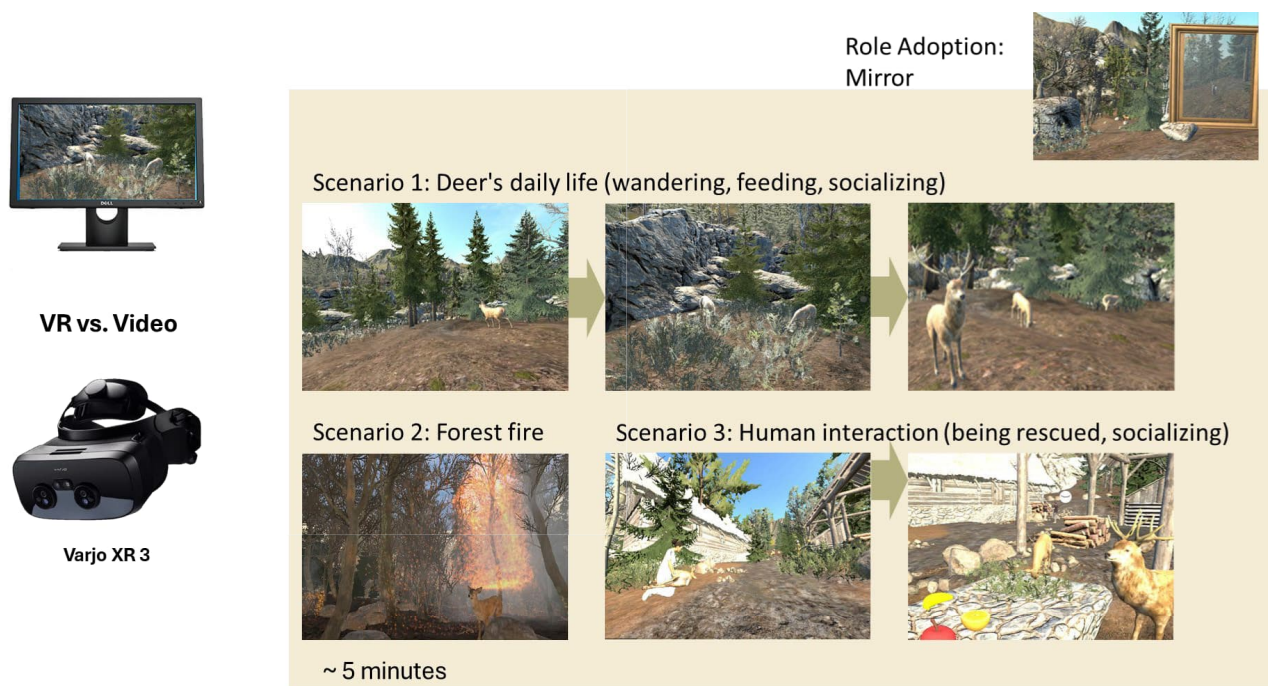


Figure P 26. Experimental Design for Animal Perspective-Taking Experience

References

- Ahn, S. J., Bostick, J., Ogle, E., Nowak, K. L., McGillicuddy, K. T., & Bailenson, J. N. 2016. Experiencing nature: Embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *Journal of Computer-Mediated Communication*, 21(6), 399-419. <https://doi.org/10.1111/jcc4.12173>
- Akerlof, K., Maibach, E. W., Fitzgerald, D., Cedeno, A. Y., & Neuman, A. 2013. Do people "personally experience" global warming, and if so how, and does it matter?. *Global environmental change*, 23(1), 81-91. <https://doi.org/10.1016/j.gloenvcha.2012.07.006>
- Arendt, F., & Matthes, J. (2016). Nature documentaries, connectedness to nature, and pro-environmental behavior. *Environmental Communication*, 10(4), 453-472. <https://doi.org/10.1080/17524032.2014.993415>
- Borges-Tiago, M. T., Almeida, A., Tiago, F. G. B., & Avelar, S. M. M. (2024). Bridging the innovative Attitude-Behavior Gap: A dual-level analysis. *Journal of Innovation & Knowledge*, 9(4), 100561. <https://doi.org/10.1016/j.jik.2024.100561>

- Erisen, E., Yildirim, F., Duran, E., Şar, B., & Kalkan, I. (2024). Exploring the effectiveness of virtual reality in combating misinformation on climate change. *Political Psychology*. <https://doi.org/10.1111/pops.13057>
- Healey, M. L., & Grossman, M. (2018). Cognitive and Affective Perspective-Taking: Evidence for Shared and Dissociable Anatomical Substrates. *Frontiers in Neurology*, 9, 491. <https://doi.org/10.3389/fneur.2018.00491>
- Melo, M., Gonçalves, G., Vasconcelos-Raposo, J., & Bessa, M. (2023). How Much Presence is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score. *IEEE Access*, 11, 24675–24685. Van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PLOS ONE*, 13(8), e0202442.

XRehab: Extended Reality for Neurological Rehabilitation

Cecilia Vona¹, Vera Colombo², Silvia Bolognini², Flavio Tonetto¹,
Stefano Regondi², Roberto Pierdicca^{1,3}, and Raffaele Pugliese²

¹Deep Reality, Pesaro, Italy;

²NLab Research Center Milan, Italy

³Dept. Civil Engineering Building and Architecture (DICEA), University Politecnica delle Marche, Italy.

Corresponding author: raffaele.pugliese@nemolab.it

Keywords: virtual reality, neurorehabilitation, AI text-to speech, emotional engagement, user experience

Abstract

Virtual Reality (VR) has great potential for supporting rehabilitation through personalized, engaging, and well-controlled experiences. The XRehab VR platform is designed for patients with neuromuscular and neurodegenerative diseases. The use of CORTEX² services allows the patients to access it remotely. Preliminary tests validated the approach and produced promising results in terms of usability, engagement and emotional impact.

Introduction

XRehab is an AI-powered virtual reality (VR) platform designed to support and enhance the rehabilitation of patients with neuromuscular and neurodegenerative diseases. It makes rehabilitation more user-friendly, accessible, personalized, and remotely controlled and monitored. While traditional rehabilitation programs are proven effective, they often face challenges such as reduced patient engagement due to repetitive exercises and poor adherence. This requires a paradigm shift. VR has emerged as an innovative solution to these limitations by providing personalized, controlled, and engaging interventions (Lasaponara, Ceradini, & Calderone). VR environments can recreate realistic scenarios that challenge motor control and cognitive abilities while promoting positive emotions [Voinescu, Bernardo, & Somarathna]. Rehabilitation can occur in clinical and home settings. Integrating VR with artificial intelligence (AI) further enhances therapeutic care by enabling it to adapt dynamically to patients' capabilities and progress. Furthermore, ongoing performance tracking allows for detailed monitoring and analysis over time, enabling adjustments to therapy based on data to improve long-term functional recovery [Słyk].

Methodology

XRehab application: The XRehab solution consists of immersive experiences developed in Unity. There is outdoor exploration of natural environments to reinforce positive emotional states, as well as activities of daily living to stimulate upper limb movements. Both are available in English and Italian, which can be selected from a user-friendly menu at the beginning of the experience. Examples of the scenarios are displayed in Figure P 27. Each outdoor environment has been chosen to evoke a specific emotion. The beach setting enhances calm and relaxation, the tourist city enhances energy, and the mountain setting enhances joy. Each exploration lasts five minutes and is guided by an AI-generated voice that helps patients focus on bodily sensations and emotional responses during the experience. The second set of scenarios consists of two challenges. The first, the Range-Separated Challenge, involves reordering objects placed at different distances from the patient. The second, the Sorting Challenge, involves sorting matryoshkas from smallest to largest. Both experiences monitor errors and completion time. Instructions are provided via graphic panels. Designed for long-term home-based rehabilitation, XRehab integrates CORTEX²'s videoconference service, which uses Rainbow SDKs for C#. The application has been optimized for use with Meta Quest 3. **Validation Protocol:** Validation follows the protocol outlined in Figure P 28. It begins with a familiarization phase to ensure comfort with the VR equipment. Anxiety levels will be assessed using the State-Trait Anxiety Inventory (STAI-Y1) before and after the VR sessions. The participants will complete two VR sessions: the first involving the three outdoor exploration scenes and the second concerning the activities of daily living. After each VR scenario, participants will complete the Self-Assessment Manikin (SAM) in VR to evaluate valence, arousal, and dominance. For outdoor exploration scenarios, they will also execute the Numeric Rating Scale (NRS) to evaluate emotional change induced by each scenario by responding to the question, "How do you feel?" directly to the clinician. At the conclusion, participants will complete questionnaires evaluating their user experience: The ITC Sense of Presence Inventory will assess spatial presence, engagement, ecological validity, and negative effects. The System Usability Scale will assess usability, complexity, and learnability. The User Experience Questionnaire will assess attractiveness, efficiency, and stimulation.



Figure P 27. XRehab simulated 3D environments. The three outdoor landscapes are on the left. On the right there is the simulated home location where the range-separated and the sorting challenges are performed.

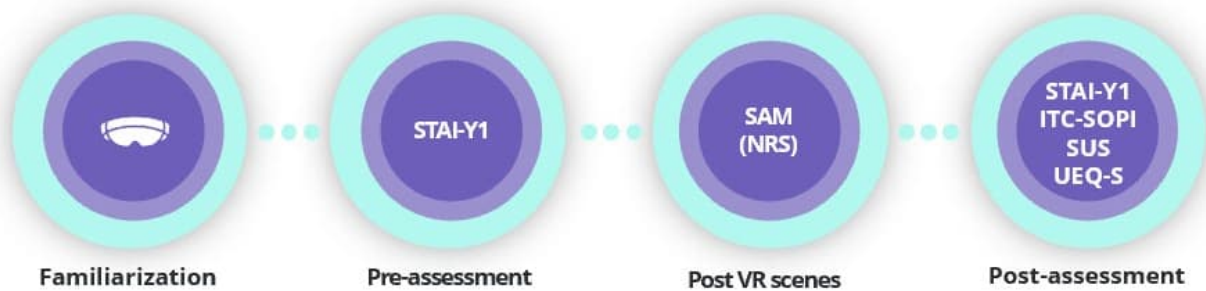


Figure P 28. Structure of the user validation protocol (NRS for emotional scenes only).

Results

User experience validation was carried out at NLAB with a group of voluntary participants (three female and two male). The group's median age was 24, and the median number of years of education was 15.8. Three participants reported no previous experience with virtual reality. Two participants are shown in Figure P 29a. **Anxiety level evaluation:** No significant changes in anxiety levels were observed before and after the VR experiences, as assessed using the STAI-Y1 questionnaire. Average scores ranged from 28 to 29 (Figure P 29b), which is below the clinical threshold of 40. Given the short duration and single-session format, significant variations in anxiety were not anticipated. Future studies should investigate whether repeated exposure to XRehab scenarios across multiple sessions might produce measurable changes. **Emotional evaluation of the nature-based environments:** Number Rating Scale (NRS) scores varied across scenarios (Figure P 29c). Participants rated the city scenario the highest, describing it as the most emotionally positive environment. This was also confirmed through feedback afterward. Conversely, the beach scenario received the lowest ratings, suggesting minimal emotional impact. The Self-Assessment Manikin (SAM) results showed variations in valence, arousal, and dominance across scenarios (Figure P 29c). Valence was lowest in the beach scenario, reflecting limited perceived pleasantness. The city scenario received higher valence ratings, consistent with the NRS findings. The mountain scenario scored intermediately. The lower rating for the beach scenario may be due to unmet participant expectations for relaxation and a perceived lack of graphical realism. High arousal scores across all scenarios suggest overall activation irrespective of environment type. Dominance ratings remained low to moderate, possibly due to limited interactivity since no specific tasks were assigned. **Overall User eXperience:** According to the ITC-SOPI results (Figure P 29D), no adverse side effects were reported. Participants experienced a moderate to low level of physical presence, yet reported a high level of engagement. The perceived naturalness of the environment was also moderate. Several participants noted that joystick movement felt too slow, which limited realism and freedom of exploration. Notably, the beach scenario was frequently cited as detracting from the overall experience and contributing to lower scores. Future improvements should aim to enhance visual fidelity and navigational responsiveness. UEQ-S results, rated on a scale from -3 to 3, showed positive scores across all three dimensions. The experience was rated highly for attractiveness and moderately for originality and stimulation. These outcomes reflect the virtual environments' general pleasantness and novelty, though limited interactivity may have slightly reduced perceived engagement. The SUS results indicate that the scenarios were generally perceived as usable and acceptable (average = 82.5, interquartile range = 7.5). Participants found the system easy to use, suggesting an intuitive interface. Although improvements in interactivity are needed, the overall usability was positively rated.

Conclusion

The validation of XRehab yielded promising results, suggesting its effectiveness as a neurological rehabilitation tool. No side effects or cybersickness were observed. To strengthen immersion, nature-based scenarios should have enhanced visual realism and environmental detail. Activity-oriented environments should be developed further by introducing structured tasks with progressive difficulty and cognitive engagement. These improvements could increase the therapeutic impact and user experience in future implementations. Once patients are well familiarized with the technology, the integration of the application with the videoconference service could allow for the investigation of home-based rehabilitation.

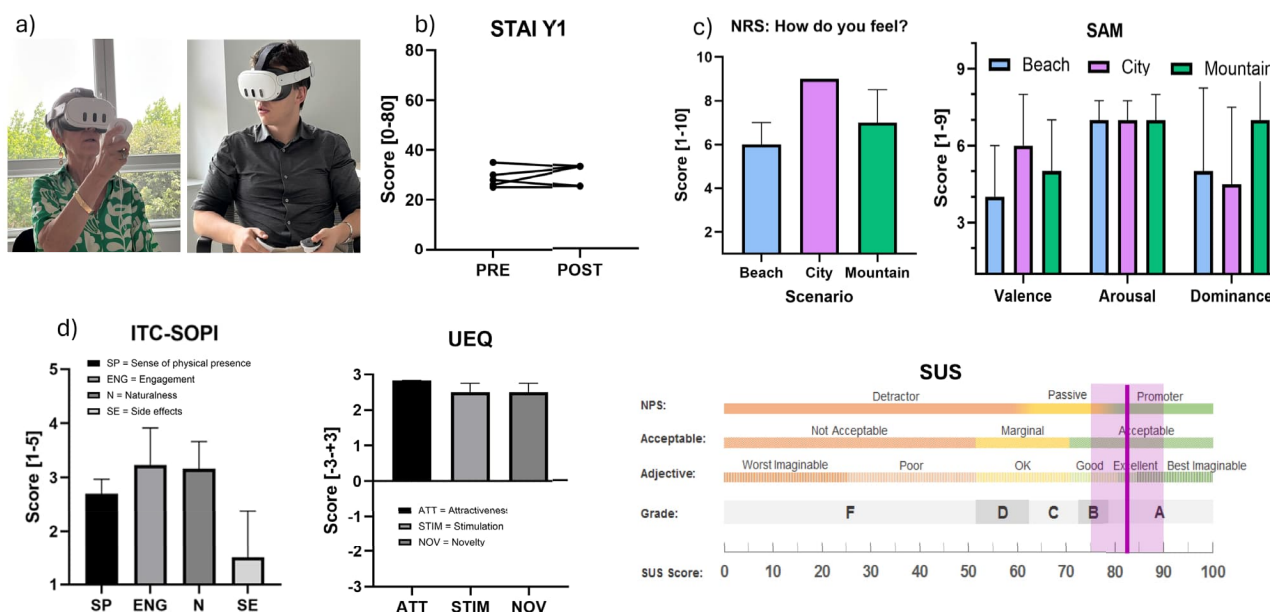


Figure P 29. a) Pilot participants, b) Pre-Post results of the STAI-Y1 for evaluating the anxiety level, c) Emotional evaluation of the outdoor scenarios (NRS and SAM), d) User experience evaluation (ITC-SOPI, UEQ-S, and SUS).

Acknowledgement

Work funded by CORTEX2 Open Call 1 - the European Union's Horizon Europe Research and Innovation Programme. Grant no. 01070192. We thank Tommaso Pantaloni, Sina Rahimzadeh, Laura Pistagnesi, and Michele Sasso for their fundamental technical help in implementing the XRehab application.

References

- Lasaponara, S., Marson, F., Doricchi, F., & Cavallo, M. (2021). A scoping review of cognitive training in neurodegenerative diseases via computerized and virtual reality tools: What we know so far. *Brain sciences*, 11(5), 528. <https://doi.org/10.3390/brainsci11050528>
- Ceradini, M., Losanno, E., Micera, S., Bandini, A., & Orlandi, S. 2024. Immersive VR for upper-extremity rehabilitation in patients with neurological disorders: A scoping review. *J. NeuroEngineering Rehabil.*, vol. 21, no. 1, p. 75, 2024
- Calderone, A., Latella, D., De Luca, R., Gangemi, A., Impellizzeri, F., de Pasquale, P., ... & Calabrò, R. S. (2025). Virtual horizons: Enhancing rehabilitation of neuromuscular diseases through virtual reality and gamification. *Journal of Neuromuscular Diseases*, 22143602241311194. <https://doi.org/10.1177/22143602241311194>

- Voinescu, A., Sui, J., & Stanton Fraser, D. 2021. Virtual reality in neurorehabilitation: An umbrella review of meta analyses. *J. Clin. Med.*, vol. 10, no. 7, p. 1478. <https://doi.org/10.3390/jcm10071478>
- Bernardo, P. D., Bains, A., Westwood, S., & Mograbi, D. C. 2021. Mood induction using virtual reality: A systematic review of recent findings. *J. Technol. Behav. Sci.*, vol. 6, no. 1, pp. 3-24.
- Somarathna, R., Bednarz, T. & Mohammadi, G. 2022. Virtual reality for emotion elicitationDa review. *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 2626-2645.
- Słyk, S., Zarzycki, M. Z., Kocwa-Karna, A. & Domitrz, I. 2019. Virtual reality in the diagnostics and therapy of neurological diseases. *Expert Rev. Med. Devices*, vol. 16, no. 12, pp. 1035-1040. <https://doi.org/10.1080/17434440.2019.1693892>

Effects of Musical Pleasantness on Intrinsic Motivation During Exercise in Non-Immersive Virtual Reality: A Pilot Study

Marta Mondellini^{1,2}, Vera Colombo³, Stefano Regondi^{3,4}, Marco Sacco¹,
and Alessandro Antonietti²

¹STIIMA-CNR, Italy,

²Catholic University of the Sacred Heart, Italy,

³NLAB Research Center, ASST GOM Niguarda, Italy,

⁴GNOSIS Research Institute, Italy

Corresponding author: marta.mondellini@cnr.it

Keywords: hedonic music; non-immersive virtual reality; engagement; rehabilitation; user experience

Abstract

Physical activity is a crucial element for physical and mental health, both in healthy people and, even more so, in people with pathologies. In some cases, such as chronic or degenerative diseases, physical activity is mainly aimed at slowing down functional decline rather than generating tangible improvements, and this makes the motivation for therapeutic adherence particularly difficult to maintain. Effective motivational strategies can offer significant support in sustaining long-term commitment to exercise programs. In this scenario, the present pilot study aims to compare the effects of pleasant music compared to a neutral rhythmic stimulus (drums) on the level of perceived involvement during physical activity. Fifty-two healthy participants will perform two exercise sessions on a cycle ergometer in non-immersive virtual reality, listening to both personalized musical accompaniment and the beat of a drum. Selfreport data, performance scores, and physiological measures are collected with the aim of investigating whether and how preferred music promotes greater motivation and a better overall experience during exercise. The results could also offer useful indications for rehabilitation contexts, where maintaining motivation is crucial.

Background

Physical activity is a fundamental aspect of maintaining good physical and mental health, and the cycle ergometer is one of the most widely used tools for practicing aerobic exercise. However, the difficulty in maintaining adherence to exercise programs over time is a well-documented criticality. Factors such as boredom, lack of variety, or lack of motivational support can negatively influence the frequency and duration of physical activity.

In this context, music proves to be a powerful resource. It constitutes an aesthetic stimulus capable of activating numerous neural circuits linked to pleasure and motivation. Recent studies in the field of

neuroaesthetics, which investigates the neural bases of aesthetic experiences, have highlighted how listening to music can promote brain plasticity, improving motor and cognitive recovery (Oliva et al., 2023). In particular, exposure to favorite musical stimuli is associated with positive effects on memory, attention, and implicit learning, increasing cognitive performance and engagement during physical activity (Sarasso et al., 2022). Interventions targeting rhythm and timing have demonstrated improvements in motor skills (Dalla Bella, 2018) and in mood and quality of life (Raglio et al., 2015). The emotional component linked to musical familiarity could therefore modulate the psychophysiological response to exercise, supporting adherence and optimizing motor and cognitive results.

With the advancement of digital technologies, virtual reality (VR) offers new possibilities to integrate multisensory experiences into physical exercise, creating motivating and personalized environments capable of supporting user engagement.

This pilot study is at the intersection of these lines of research, exploring the effectiveness of using personalized music in combination with non-immersive virtual reality to promote more engaging physical exercise, with the aim of preliminarily evaluating the validity and efficacy of this approach.

Study protocol

The evaluation will be carried out on 52 healthy volunteers over 18 years of age, without neurological, cognitive, and sensorial problems, previous history of epileptic seizures or motion sickness. At T0, about a week before the first cycling session, the participant will be explained how the study will be conducted, and they will be asked to make a list of 15 songs or music they love; in this phase, two dates for the experiment will be scheduled. These two days are separated by a period of at least 2 weeks. Both times, the participant is welcomed at the laboratory where the experiment is carried out. The participant will wear the heart rate chest band with Polar H10 sensor.

In the two sessions, before starting the exercise, the participant is asked to stay seated, relaxed, and in silence for 5 minutes; this phase is necessary to have resting heart rate data. After this moment, the participant is asked to play two exergames displayed on a large monitor (40") placed in front of them. The two exergames are part of a non-immersive VR application, called *Velody*, described in the following section. The app is installed on a Windows OS PC (ROG Zephyrus 14) and works in connection with a cycle-ergometer, the *Thera Trainer Tigo*. On one of the two days, the participant will pedal while hearing a drum rhythm in the background (rhythm condition), consistent with the RPM to be maintained on the cycle ergometer. On the other day, the participant will listen to the music previously selected by them (music condition), appropriately adapted via free online software (<https://29a.ch/timestretch/>) to have the appropriate RPM for the exercise. The order of exposure to one or the other experimental condition is proposed randomly.

At the end of each cycling session (T1-music and T1-rhythm) the participant will be asked to fill out the following questionnaires: User Engagement Scale (O'Brien et al., 2018), Situational

Intrinsic Motivation Scale (Guay et al., 2000), Intrinsic Motivation Inventory (enjoyment/interest scale) (McAuley et al., 1989), Game Experience Questionnaire (IJsselstein et al., 2013), Self-Assessment Manikin (Bradley et al., 1994), and Borg scale (Borg, 2007). During each session, objective data relating to gaming performance (pedaling speed and score in exergames) and heart rate are also collected.

The Application: Velody

The Velody application is a non-immersive VR app working in combination with a cycleergometer for navigation. The Thera Trainer Tigo is a powered ergometer for upper and lower limbs allowing passive, active, or assisted cycling.

The application - Velody - has been developed with Unity version 2022.3.24f1, and it is an updated version of previous applications developed by STIIMA research group (Colombo et al., 2019; Mondellini et al., 2023; Colombo et al., 2024). It includes a User Interface (UI) for creating the player profile, selecting the preferred exercise, and setting the duration and target speed of the exercise. A Bluetooth communication protocol has been developed based on the information provided by Thera Trainer to enable the communication of the physical exercise data.

The virtual environment represents a park in which the player can navigate along a circular pathway on a virtual bicycle. The cycling speed measured by the ergometer is sent via Bluetooth to the Unity application and is used to move the virtual bicycle along the path. The training time is also sent and displayed on the UI. Velody includes two exercises: "Keep it up" and "Overtakes" (Figure P 30).

The aim of the first exercise is to maintain a predefined constant pedaling speed. Along the path, a series of stimuli (e.g., butterfly, flower) will appear. At any given time, three stimuli are visible on the path. As the user approaches a stimulus, a score from 0 to 3 is assigned based on how closely the average speed over the most recent segment (from the previous stimulus to the current one) matches the predefined target speed set by the researcher. In particular, the maximum score is assigned if the average speed differs from the target speed by no more than 3 RPM; 2 points are given if the difference is between 3 and 7; 1 point is assigned if the difference is from 7 to 15 RPM; 0 points are given if the difference is higher than 15 RPM. Depending on the score obtained, the player collects from one to three stimuli. Collected stimuli turn green, and a sound is played for each collected item to provide immediate feedback. An example is shown in Figure P 30. During the pedaling exercise, the rhythm of the selected auditory feedback (music/drums) is synchronized to the desired pedaling frequency (RPM – revolutions per minute).

In the second exercise, the aim is to modulate the speed by accelerating and decelerating at predefined time points. The user progresses at a target cruising speed. Along the path, beside the predefined trajectory, animals appear one at a time. As the user approaches, the animal begins to move at a speed faster than the user's current pace, prompting the user to increase their pedaling speed in order to overtake it. The interface displays the required speed needed to surpass each moving stimulus. The required speed is set to 20 RPM above the target cruising speed with a tolerance of 0.05 RPM. The player has to maintain the required speed for at least 10 seconds and the maximum time allowed to complete the task is 100 seconds. If the player overtakes the animal, a feedback sound is played and the score is increased by one point. In this exercise too, the rhythm of the auditory feedback is synchronized to the desired pedaling frequency, both during standard pedaling and during acceleration phases.

Exercise parameters are set before starting the exercise. Exercise duration is set through the Thera Trainer panel and sent to the Velody application by Bluetooth; remaining time is updated in real time. Target pedaling speed is set through the Velody UI. Both exercises can be played by both using the arms and the legs, depending on the therapy goal.

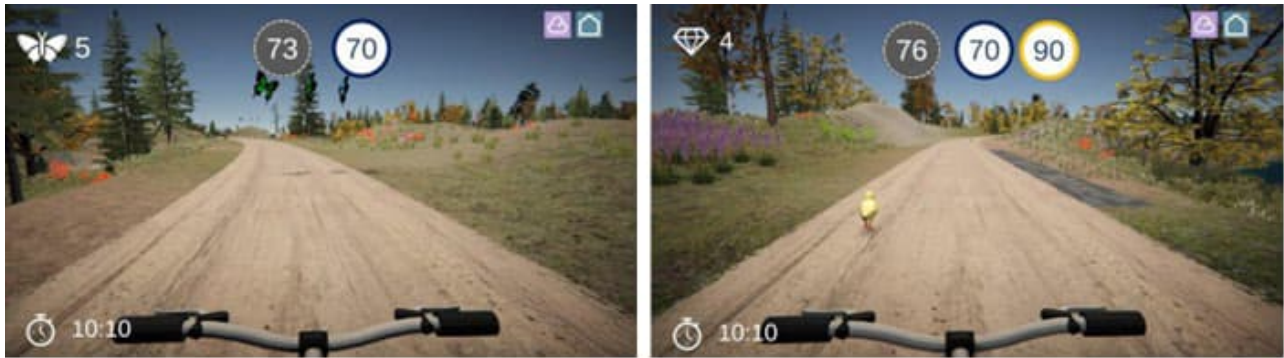


Figure P 30. Task 1 "Keep it up" on the left; task 2 "Overtakes" on the right

Future work

In the present study it is hypothesized that the use of favorite music, compared to a neutral rhythmic stimulus, is associated with greater engagement and higher perceived motivation, with an increase in emotional valence and a reduction in subjectively perceived effort. We are interested in observing any variation in performance in the task and in the physiological response as well. The aim is to verify whether the use of familiar and pleasant musical stimuli, integrated into an interactive virtual context, can effectively enhance the user's experience during physical activity, and, in particular, the intrinsic motivation. The results of this pilot study may offer valuable insights for the design of personalized interventions to promote physical activity.

References

- Borg E., 2007. On perceived exertion and its measurement, US-AB, Stockholm
- Bradley, M. M., & Lang, P. J. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1), 49-59, [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Colombo, V., Mondellini, M., Gandolfo, A., Fumagalli, A., & Sacco, M. 2019. Usability and acceptability of a virtual reality-based system for endurance training in elderly with chronic respiratory diseases. In *Virtual Reality and Augmented Reality: 16th EuroVR International Conference, EuroVR 2019, Tallinn, Estonia, October 23–25, 2019, Proceedings 16* (pp. 87-96). Springer International Publishing.
- Colombo, V., Mondellini, M., Fumagalli, A., Aliverti, A., & Sacco, M. 2024. A virtual reality-based endurance training program for COPD patients: acceptability and user experience. *Disability and Rehabilitation: Assistive Technology*, 19(4), 1590-1599. <https://doi.org/10.1080/17483107.2023.2219699>
- Dalla Bella, S. 2018. Music and movement: Towards a translational approach. *Neurophysiologie Clinique*, 48(6), 377-386. <https://doi.org/10.1016/j.neucli.2018.10.067>
- Guay, F., Vallerand, R. J., & Blanchard, C. 2000. On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and emotion*, 24, 175-213
- IJsselstein, W. A., De Kort, Y. A., & Poels, K. 2013. The game experience questionnaire
- McAuley, E., Duncan, T., & Tammen, V. V. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, 60(1), 48-58. <https://doi.org/10.1080/02701367.1989.10607413>
- Mondellini, M., Rutkowski, S., & Colombo, V. 2023. Cycling in Immersive VR: Motivation and Affects in Post-COVID Patients. In *International Conference on Extended Reality* (pp. 353-366). Cham: Springer Nature Switzerland.

- O'Brien, H. L., Cairns, P., & Hall, M. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112, 28-39. <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- Oliva, A., Iosa, M., Antonucci, G., & De Bartolo, D. 2023. Are neuroaesthetic principles applied in art therapy protocols for neurorehabilitation? A systematic mini-review. *Frontiers in Psychology*, 14, 1158304. <https://doi.org/10.3389/fpsyg.2023.1158304>
- Raglio, A., Attardo, L., Gontero, G., Rollino, S., Groppo, E., & Granieri, E. 2015. Effects of music and music therapy on mood in neurological patients. *World Journal of Psychiatry*, 5(1), 68. <https://doi.org/10.5498/wjp.v5.i1.68>
- Sarasso, P., Barbieri, P., Del Fante, E., Bechis, L., Neppi-Modona, M., Sacco, K., & Ronga, I. 2022. Preferred music listening is associated with perceptual learning enhancement at the expense of self-focused attention. *Psychonomic Bulletin & Review*, 29(6), 2108-2121

VR Journeys for Older Adults: Storytelling-Enhanced Virtual City Tours and Their Impact on Senior Customer Experience

Giulia Maria Egglı, and Wolfgang Schäfer

ZHAW Zurich University of Applied Sciences, Switzerland

Corresponding author: wolfgang.schaefer@zhaw.ch

Keywords: Virtual Reality, Storytelling, Older Adults, Customer Experience, Tourism, Inclusive Design

Background and Motivation

Europe will count more citizens aged 65 + than under 15 by 2030 (Eurostat, 2024), and in Switzerland this cohort already controls ~37 % of leisure expenditure (Berg & Liljedal, 2022). Ageing is accompanied by reduced physical mobility and, in many cases, social isolation. Recent studies show that immobile tourism via desktop videos or live webcams generates only modest engagement because sensory richness and embodied agency are missing (Liu & Lehto, 2024). Virtual Reality (VR) is uniquely positioned to close this gap by providing situated, first-person access to place experiences while the user remains safely seated. In controlled lab trials VR exposure has improved mood, reduced perceived loneliness and temporarily eased chronic pain in older adults (Muslu et al., 2025). Yet most VR tourism content is produced for younger thrill-seekers and focuses on fast-paced, game-like interactions that can overwhelm seniors (Aleti et al., 2024).

Storytelling can function as a cognitive scaffold: a well-paced narrative gives structure, fosters anticipation and helps users integrate what they see with prior memories (Fog et al., 2010). Nevertheless, little is known about how specific storytelling elements—plot arcs, linguistic style, cultural references—shape senior customer experience (CX) in VR. Our study addresses this gap by exploring a narrated 360° city tour of a city designed explicitly for seniors. The research questions were:

- RQ1. What everyday benefits do seniors derive from story-based VR travel?
- RQ2. Which storytelling strategies enhance presence, enjoyment and social connection for this audience?

Theoretical Foundations

We integrate three streams of literature: **(i) presence & embodiment**, **(ii) narrative transportation**, and **(iii) service-dominant logic of CX**. Presence theory posits that spatial (i.e., feeling there) and social (i.e., feeling with others) presence underpin meaningful VR experiences (Slater & Wilbur, 1997). Narrative transportation theory argues that emotionally laden stories prompt users to suspend disbelief, allocate mental resources and adopt protagonists' perspectives (Van Laer et al., 2014). Finally, consumer-dominant logic views tourists not as passive consumers but as active co-creators of

experience value through cognitive, affective and social processes of self-location and copresence (Cummings et al., 2022).

Synthesizing these perspectives, we propose a conceptual model (Figure P 31) in which **story form** (plot clarity, pacing, language vividness) and **story delivery** (narrator persona, prompts, onboarding) influence transportation, which together with presence drives four senior benefits: **mobility substitution, emotional uplift, social connectedness and cognitive stimulation**.

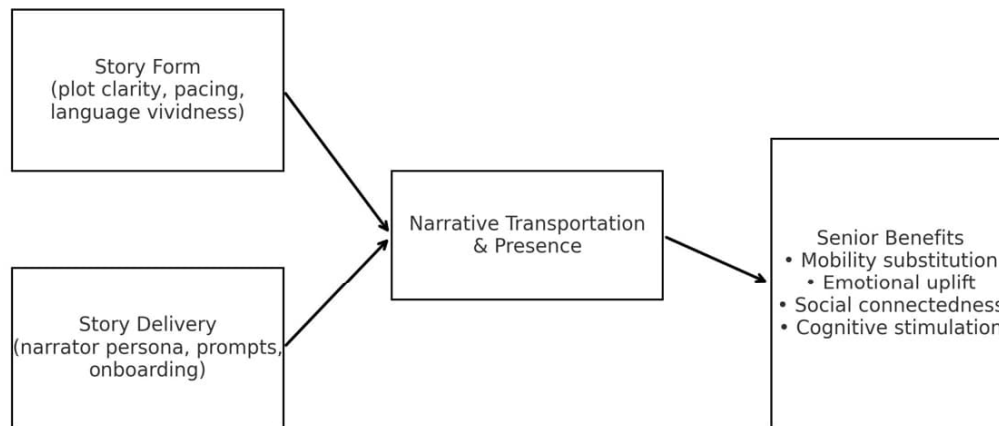


Figure P 31. Conceptual model linking storytelling elements → narrative transportation & presence → four senior benefits. (own illustration)

Methodology

Study Design

An exploratory, qualitative design was chosen to gain deep insight into seniors' perceptions. We produced a 15-minute, 5-stop 360° tour of city's historic core using an 360° camera, scripted narration in Standard German (female voice actor) and spatial audio ambience. The tour opens with a short onboarding sequence that demonstrates how to adjust straps and recenter view.

Participants & Recruitment

Nine older adults (age range 66–94, M = 78.4; 5 female, 4 male) were recruited via two community centers and a retirement residence. Inclusion criteria were normal or corrected-to-normal vision and hearing, and no prior adverse reactions to VR. To capture heterogeneity, we intentionally balanced:

- **Mobility level:** 5 could walk unaided, 4 used walking aids or wheelchairs.
- **Technology familiarity:** 4 had never used VR; 5 had tried 360° video on mobile.
- **Residential context:** 3 urban, 6 suburban/rural dwellers.

Procedure

Participants attended individual 45-minute sessions. After consent, basic demographic and technology attitudes were recorded. The VR experience was delivered on a Meta Quest 2 (90 Hz, visual field ≈ 97°) in swivel chairs to allow free head rotation. Post-experience, semi-structured interviews probed affective reactions, perceived benefits, comprehension of narrative and usability issues. Each interview lasted 18–28 min and was audio-video recorded.

Data Analysis

Recordings were transcribed verbatim and analyzed with the application f4. We followed Froschauer & Lueger's thematic analysis, coding deductively for the four CX benefits and inductively for emerging storytelling cues. Discrepancies were discussed until consensus. Themes were mapped onto the conceptual model to identify influence pathways.

Results

Everyday Benefits of VR Travel

1. **Mobility Substitution.** Most mobility-restricted seniors framed VR as "doors back into the city" (P7) and likened headset time to an outing: "I cannot manage the tram anymore, but this felt like being on the [name] square again—without the aches" (P4).
2. **Emotional Uplift.** All participants reported positive affect; words such as *joyful*, *liberating* and *nostalgic* dominated.
3. **Social Connectedness.** Even solo viewing fostered a form of mediated companionship: "The speaker guided me like a friend" (P2). Two participants suggested watching simultaneously with grandchildren via headsets.
4. **Cognitive Stimulation.** Several seniors referenced new knowledge: Historical facts, legend details and orientation cues spurred later conversation; one participant (P6) googled the city's towers after the session.

Storytelling Mechanisms

- **Narrative Spine & Signposting.** A clear beginning (a central bridge), middle (a cathedral) and end (view over a lake) provided temporal anchors.
- **Pacing & Pause Points.** Slower camera pans (max 8°/s) and scheduled narration silences allowed seniors to visually explore. Fast pans in a pilot cut triggered mild nausea in 3/5 testers, underlining the importance of pacing.
- **Cultural Hooks & Legends.** Embedding local myths cultivated curiosity and personal anecdotes: "My father told me the same legend; I could see it again" (P3).
- **Linguistic Style.** Concrete sensory language ("river breeze") evoked multisensory imagery.
- **Agency Prompts.** Second-person invitations ("look left to spot the guild houses") fostered active search behavior; gaze-tracking logs showed head turns within 2 s of prompts for 78 % of occurrences.
- **Story-Integrated Onboarding.** The narrator introduced controls while 'standing' on a square, framing the technical tutorial as part of the tour, which reduced onboarding time to under 90 s for all participants.

Persona Differences

Mobility-restricted seniors (Persona 2) valued emotional uplift and mobility substitution most, rating overall usefulness 5/5 versus 4/5 for Persona 1. Mobile travelers, in contrast, focused on cognitive stimulation and saw VR as *complementary pre-visit inspiration*.

Discussion and Implications

Our findings extend narrative transportation theory by showing that **agency prompts** and **story-integrated onboarding** act as *antecedents* of transportation in VR tourism for older adults. Practically, designers should:

1. **Script VR tours as guided walks.** Use a narrator persona, clear plot points and sensory metaphors to lower cognitive load.
2. **Optimize pacing for vestibular comfort.** Maintain slow camera motion and provide static pause scenes.
3. **Embed cultural hooks** that reference users' autobiographical memory to heighten self-relevance.
4. **Facilitate shared viewing.** Include multi-user modes or simple "call-and-join" links to leverage social uplift.

Limitations and Future Work

A qualitative sample of nine limits statistical generalisation; however, rich narratives yielded transferable design insights. Future work will conduct a 2 × 2 experiment (high vs. low narrative structure × seated vs. standing setup) with 60 + participants and include psychophysiological presence measures (heart-rate variability). Moreover, integrating real-time voice interaction with live guides could merge story-doing with social telepresence.

References

- Aleti, T., Figueiredo, B., Martin, D. M., & Reid, M. 2024. Digital inclusion in later life: Older adults' socialisation processes in learning and using technology. *Australasian Marketing J.* 32(4), 295–307. <https://doi.org/10.1177/14413582231187652>
- Berg, H., & Liljedal, K. T. 2022. Elderly consumers in marketing research: A systematic literature review and directions for future research. *Int. J. Consumer Studies*, 46(5), 1640–1664. <https://doi.org/10.1111/ijcs.12830>
- Cummings, J. J., Tsay-Vogel, M., Cahill, T. J., & Zhang, L. 2022. Effects of immersive storytelling on affective, cognitive, and associative empathy. *New Media & Society*, 24(9), 2003–2026. <https://doi.org/10.1177/1461444820986816>
- Eurostat. 2024. Ageing Europe — 2024 edition. Publications Office of the European Union.
- Fog, K., Budtz, C., Munch, P., & Blanchette, S. 2010. *Storytelling: Branding in practice* (2nd ed.). Springer-Verlag.
- Froschauer, U., & Lueger, M. 2020. *Das qualitative Interview: Zur Praxis interpretativer Analyse sozialer Systeme* (2. Ed., 2020). Facultas.
- Liu, Y., & Lehto, X. 2024. Sense of confinement and preferences of virtually delivered tourism offerings: A tale of two stories. *J. Destination Marketing & Management*, 34, 100964. <https://doi.org/10.1016/j.jdmm.2024.100964>
- Muslu, L., Karakuş, Z., Ası, E., Bayındır, R., & Özer, Z. 2025. Time travel of older people through virtual reality: A qualitative study. *BMC Geriatrics*, 25(1), 42.
- Slater, M., & Wilbur, S.(1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence*, 6(6), 603–616.
- Van Laer, T., De Ruyter, K., Visconti, L. M., & Wetzels, M. 2014. The Extended Transportation-Imagery Model: A meta-analysis of the antecedents and consequences of consumers' narrative transportation. *J. Consumer Research*, 40(5), 797–817. <https://doi.org/10.1086/673383>
- Witmer, B. G., & Singer, M. J. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3), 225–240. <https://doi.org/10.1162/105474698565686>

Dynamic, Room-Adapted Spatial Audio for Augmented Reality Telepresence

Nils Meyer-Kahlen, Ivan Benc, Peter Lindahl, Jaakko Heikkinen,
Johannes M. Arend, and Tapio Lokki

Acoustics Lab, Department of Information and Communications Engineering, Aalto University, Finland

Corresponding author: nils.meyer-kahlen@aalto.fi

Keywords: Spatial Audio, Room Acoustics, AR Telepresence, Holographic Calling

Introduction

Augmented Reality (AR) telepresence—also referred to as holographic calling—could be a promising future communication technology. It allows users to see their interlocutors as three-dimensional holograms, displayed through semi-transparent glasses with built-in displays. Ideally, this should facilitate more natural interactions across distances, leading to a stronger sense of co-presence than present-day video calling technology. Achieving such visual rendering requires depth cameras, efficient point-cloud compression and transmission, and high-contrast, large-field-of-view AR displays. While a lot of important work is directed towards improving these technologies, we focus here on the acoustic challenges involved in AR telepresence. Ideally, a holographically rendered person should not only look as if they were present in a user's room, but they should also sound as if they were. This requires dynamic, room-adapted binaural headphone rendering. The poster will present the work conducted towards this goal within the AURA project, a part of the SPIRIT Open Call 2. It will outline the technical system necessary for such audio rendering and summarize the results of experiments regarding blind room acoustic identification, perceptual requirements for creating renderings of virtual sound sources that are perceived as originating from the user's own room, and the overall quality of experience obtained with the proposed room-adaptive rendering system.



Figure P 32 Concept of a holographic call.

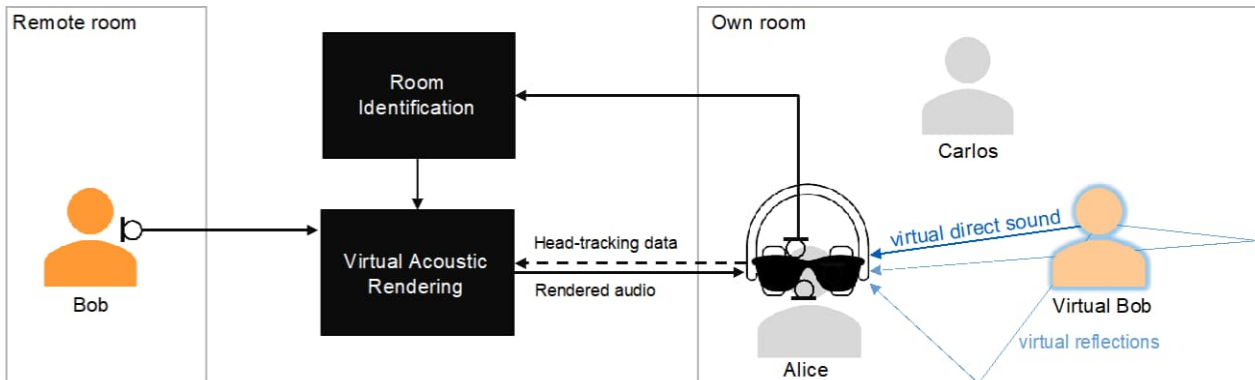


Figure P 33 Sketch of a technical audio rendering system capable of dynamic, room-adapted spatial audio.

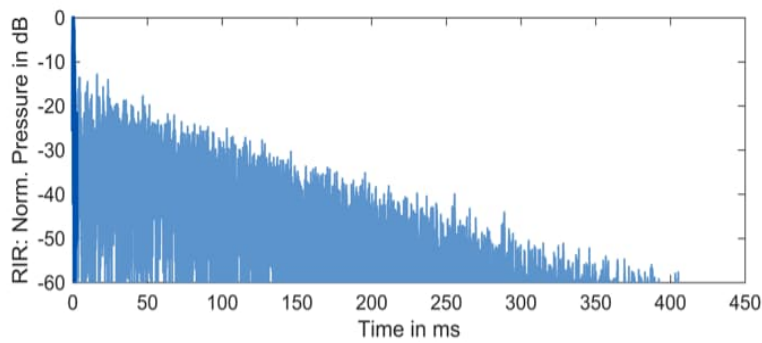


Figure P 34. Example of a room impulse response, consisting of direct sound (dark blue) and reflections (light blue).

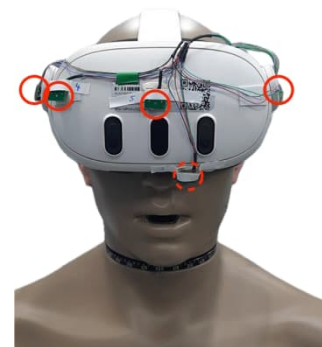


Figure P 35. Head-worn microphone array used for testing blind room identification algorithms.

Room-Adapted Spatial Audio Rendering System

Figure P 33 illustrates an exemplary AR telepresence audio rendering system comprising its essential components. The system aims to make a remote person's (Bob's) voice sound like it originates from a specific spot in the user's (Alice's) own room. Such a system requires several specialized components, including head-tracked, transparent headphones, and an efficient virtual acoustic rendering engine that is adapted to the user's own room acoustics. As performing dedicated acoustic measurements in the user's room would not be feasible in practice, blind room acoustic identification is required for such adaptation. In general, the room acoustics of a space is characterized by a room impulse response (RIR), as seen in Figure P 34. Often, the RIR is characterized using at least two important, frequency-dependent acoustic quantities: reverberation time (RT) and direct-to-reverberant energy ratio (DRR). The RT describes how quickly the reflected energy decays, i.e., the slope of the light blue decay in Figure P 34; the DRR is the ratio of the direct sound energy (dark blue) and the reflected energy.

In our project, we employ methods by Deppisch et al. (2024) to perform room acoustic identification. To achieve this, we utilize a microphone array attached to a head-mounted display. We then performed recordings and measurements with the array in 20 different spaces, allowing for systematic quantification of errors across a wide range of rooms. In this context, two scenarios are considered: one in which the user's own speech is used for room acoustic identification, and another in which the speech of other people present in the environment (e.g., Carlos in Figure P 33) is used. A head-worn array like the one shown in Figure P 35 will be used. Identification results will be presented in terms of frequency-dependent RT and DRR.

Perceptual Requirements for Audiovisual Congruence

The second part of our project focuses on the perceptual requirements for the room-adapted rendering such that users perceive a virtual sound as originating from their own room. So far, room mismatch between the real room and the virtual rendering has been studied by Meyer-Kahlen et al. (2024), where the acoustics of the real room was changed physically, impacting both RT and DRR. Here, we systematically vary the DRR of the rendered acoustics, as perceptual thresholds for DRR are, in general, less well understood than those for RT. One of the few studies on just noticeable differences (JNDs) of DRR, i.e., differences in DRR that can be detected under optimal conditions, was conducted by Larsen et al (2008). Our initial data collected with 12 participants, shown in Figure P 36, confirm the result by Larsen that JND for DRR depends on the absolute DRR. At very low DRR, i.e., where the reverberant sound dominates, larger differences remain unnoticed. As opposed to Larsen et al., we do not see an increased threshold for high DRRs.

In the upcoming work, we will relate these new results on JNDs for DRR to results on audiovisual matching. We will perform an experiment in which a hologram is shown in the experimental room while it is acoustically rendered using different, mismatched DRRs. Subjects are then asked whether the presented auditory rendering matches the visual presentation.

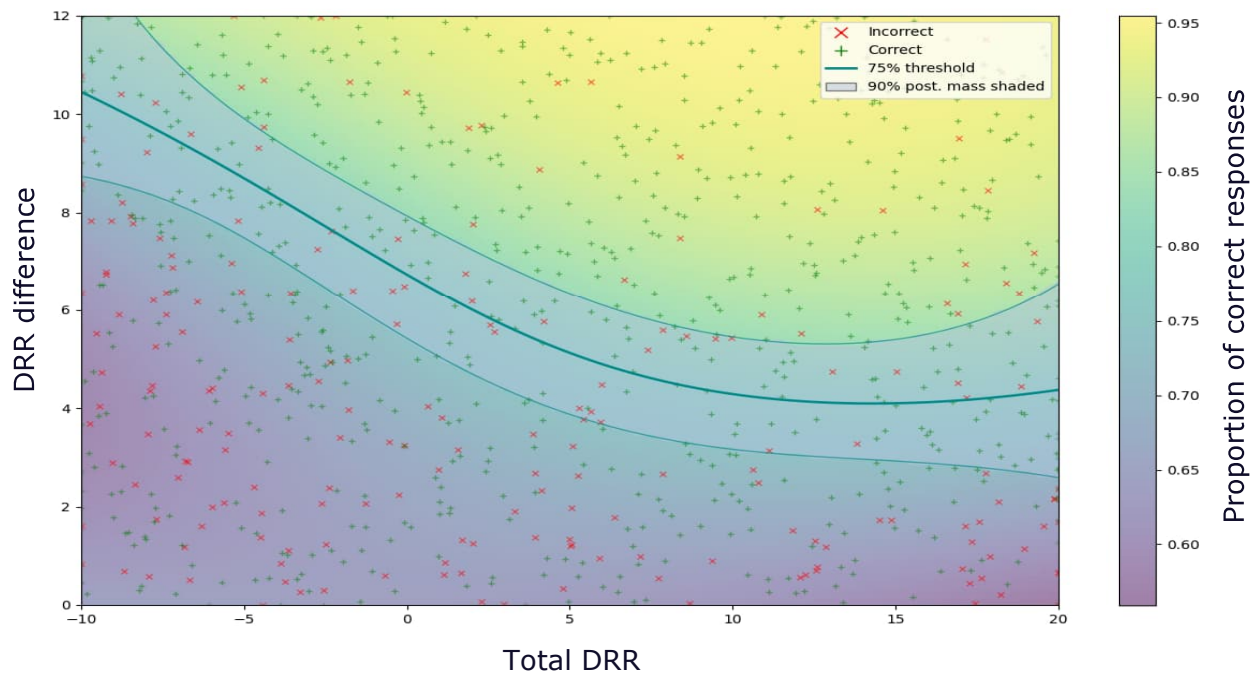


Figure P 36. Initial results for JND of changes in DRR. The threshold is the lowest for medium to high DRRs.

User Testing

Finally, we present results of an experiment that tests the quality of a holographic calling as done by Rendle et al. (2025). We either use no room acoustic rendering, rendering using a generic room and room-adapted rendering. Moreover, these options are tested in a silent environment, or while simulating a noisy space using additional loudspeakers in the room.

The holographic calling setup shown in Figure P 37 consists of an Intel RealSense depth camera and Oculus Quest 3 with passthrough capabilities. For audio rendering, a special, acoustically transparent headphone (Mülleder and Zotter, 2023) and a measurement based, custom renderer is used. Participants engage in structured conversations, as done by Abehsera-Morell et al. (2025) and subsequently rate various aspects regarding overall quality, listening effort, and co-presence.



Figure P 37. Holographic calling setup with HMD and transparent headphones.

Conclusion

AR Telepresence is a promising communication technology of the future. Here, we present an audio rendering approach for it. The aim is to render a remote interlocutor using a virtual acoustic rendering system that is adapted to a user's own room. We present the results of blind acoustic identification, required for such rendering; we also report on limits on noticeability of mismatches in room acoustic parameters and relate them to requirements for virtual sound sources to be perceived as matching to the user's own environment. Finally, the overall quality of experience of such a system compared to rendering with generic room acoustics is tested in a silent and noisy environment.

Acknowledgement

This work has partly been funded by the European Union (SPIRIT, grant no. 101070672)

References

- Meyer-Kahlen, N., Amengual Garí, S. V., Schlecht, S. J. & Lokki, T. 2024. Testing Auditory Illusions in Augmented Reality: Plausibility, Transfer-Plausibility and Authenticity. *J. Audio Eng. Soc.* 72 (11), 797–812
- Deppisch, T., Meyer-Kahlen, N., & Garí, S. V. A. 2024. Blind identification of binaural room impulse responses from smart glasses. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.
<https://doi.org/10.1109/TASLP.2024.3454964>
- Rendle, G., Immohr, F., Kehling, C., Lammert, A., Kreskowski, A., Brandenburg, K., ... & Froehlich, B. (2025, March). Influence of audiovisual realism on communication behaviour in group-to-group telepresence. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)* (pp. 569-579). IEEE.
- Larsen, E., Iyer, N., Lansing, C. R. & Feng, A. S. 2008. On the minimum audible difference in direct-to-reverberant energy ratio. *J. Acoust. Soc. Am.* 124, 450–461. <https://doi.org/10.1121/1.2936368>
- Mülleder A, & Zotter F. 2025. Ultralight circumaural open headphones. 154th Audio Engineering Society Convention. Espoo, Finland
- Abehsera-Morell P., Poirier-Quinot D., Katz B.F.G. 2025. Impact of Room Acoustic Mismatch on Conversational Dynamics in Virtual One-on-One Conferencing: Preliminary Results. *Forum Acusticum. Málaga, Spain*

Illustrating AR Patterns: A Visual Language for Event-Driven Augmented Reality

Steven Häsler¹, Alessandro Holler², Martina Kessler¹, Stefan Schmidlin²,
Yanick X. Lukic¹, and Philipp Ackermann¹

¹ZHAW University of Applied Sciences,

²ZHdK Zurich University of the Arts

Corresponding author: hasv@zhaw.ch

Keywords: Augmented Reality, Design Patterns, Illustrations

Introduction

Augmented Reality (AR) experiences are inherently spatial, dynamic, context-aware, and event-driven. AR storyboards must explain complex interactions that are difficult to specify and communicate. Traditional 2D design methodologies, such as 2D wireframes, collapse under these 3D, event-driven demands (Rajaram, Roesner, & Nebeling, 2023). Recent pattern catalogs describe how to build AR applications (Ackermann, 2023). However, development often fails due to the lack of a standard visual grammar between developers and designers for sketching those patterns (Krauss, Boden, Oppermann, & Reiners, 2021). We address this gap by introducing a compact, open-source, and accessible illustration toolbox. Our system provides a visual language, comprising three canonical canvases, a color-coded vocabulary, and glyphs for anchors, augmented elements, detections, and user traces. The language is grounded in the Event–Condition–Action (ECA) formalism (ARPatterns, 2025), a familiar concept to AR developers, and aligns with anchor-stability guidelines from Microsoft’s Spatial-Anchors Documentation (Microsoft, 2025) and Apple’s RealityKit AnchorEntity API (Apple, 2025). Through three iterative sketch-feedback loops, we produced a CC-BY 4.0 SVG toolkit that can improve cross-disciplinary communication and set the stage for pattern-driven authoring plugins. The toolkit is available at www.arpatterns.dev and on github.com/ARpatterns.

Method

We developed the toolbox using formal principles to maximize cross-disciplinary communication: **A) Precision:** The visuals map unambiguously to the underlying Event-Condition-Action logic and AR Patterns concepts. **B) Distinctiveness:** Core concepts (anchors, augmentations, detections, actions) are clearly distinguished through color and form. **C) Openness:** The toolbox is provided as open-source SVGs, encouraging unrestricted use and extension. We analyzed the existing AR Patterns catalogue (ARPatterns, 2025) to identify the key structural elements that require visual representations. We then

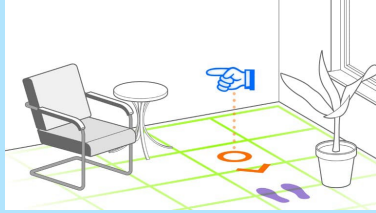
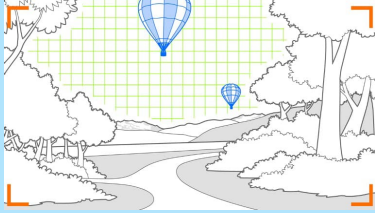
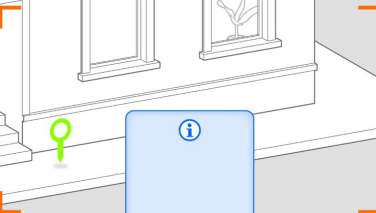
defined a consistent visual grammar with a minimal expressive palette, refined iteratively through reviews and prototyping.

Results

Canvases and Backdrops

Three canonical canvases (room, park, street corner) represent common AR contexts as structural templates, drawn in parallel projection to avoid perspective distortion and ensure illustrations are comparable, coherent, and composable (see Table P 7).





Table P 7: The Core Visual Encoding System

Room	Park	Street Corner
		
Room (only floor and walls): This backdrop can be extended with visual content enhancing the semantics.	Park: This backdrop can be used for composing various outdoor scenes.	Street corner: This backdrop can be used for geo-anchored or map-based AR.

Visual Grammar for Event-Driven AR

The toolkit uses a limited but expressive and accessible visual vocabulary (see Table P 8). Anchors, represented by orange icons, denote the origin reference of spatial content. These can be enhanced with augmentations, indicated in blue. Detected surfaces (usually detected by an AR device), depicted in green, represent objects and areas. Props, such as chairs, tables, and plants, can be added to enrich the semantic context. Additionally, user-related assets illustrate various hand gestures and body poses.





Table P 8: The Core Visual Encoding System

Element	Category	Color	Visual Rationale	Example
Anchor/marker	Spatial	Orange	Represents origin point, marker, or reference for AR content.	
Augmentation	Content	Blue	Represent virtual content being displayed.	
Detection	System event	Bright green	Represents a surface or object recognized by the system.	
User	User extent	Violet	Represents the users's body, gaze, or direct interactions.	

AR Markers

To bind virtual content to the real world, the system uses a vocabulary of AR markers. These markers define the origin, orientation, and boundaries of augmentations. They are the visual representations of the spatial anchors that are fundamental to any AR experience (see Table P 9).


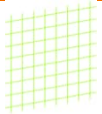

Table P 9. The Visual Vocabulary for AR Markers

Marker Type	Dimension	Description	Example
Origin (anchor within)	2D/3D	Represents a point of origin where the anchor is located directly within the marker.	
Origin (anchor offset)	2D/3D	Represents a point of origin where the anchor is located at an offset from the marker.	
Orientation	2D/3D	An arrow-shaped icon used to indicate the orientation or direction of an object or augmentation.	
Anchor area on floor	3D	An outline that defines a specific area on a horizontal surface, such as the floor, which serves as a larger anchor region for content.	

Detected Instances

Bright green elements depict what an AR system detects in the environment (see Table P 10).




Table P 10. The Visual Vocabulary for detected Instances

Detection Type	Visual Representation	Description	Example
Geolocation	A green map pin icon	Indicates a specific geographical coordinate detected by the system used for location-based AR.	
Planar surface	Green grid overlay on a flat plane	Represents detected flat surfaces, such as floors and walls, which could be used as anchors.	
3D scanned object	A dense green point cloud or mesh representing the full volume of an object	Indicates a 3D scan of an object, capturing its form and geometry.	

Augmented Elements

The augmented instances are visual representations of the digital content within the AR experience. These elements bring the world to life and can range from simple panels to portals, indicating a complete virtual world (see Table P 11).




Table P 11. The Visual Vocabulary for Augmented instances.

Augmentation Type	Visual Representation	Description	Example
Information panels	Blue, semi-transparent 2D pop-ups or 3D panels	Used to display contextual information, such as text, images, or UI elements.	
3D virtual objects	Blue 3D models of abstract or realistic objects	Represents any virtual object placed into the real world.	
Portals	A blue, door-like 3D structure	Functions as a gateway to a virtual environment, creating an immersive transition.	

User Elements

The toolbox includes a vocabulary for representing the user, whose presence, movements, and actions are often the primary drivers of events in AR scenarios (see Table P 12).

Table P 12 The Visual Vocabulary for User Elements

User	Visual Representation	Description	Example
User presence	Violet silhouettes representing the user's face, bust, or full body	Used to illustrate the user's physical position, orientation, and scale within AR. The face and bust can also serve as specific targets for augmentations, such as virtual try-ons or filters.	
Hand gestures and Interaction	A comprehensive set of violet hand poses, including open palms, pointing, pinching/grabbing, and making a fist or V-sign	Represents specific user inputs, commands, or interactions. A hand overlaid with a green detection grid signifies that the system is actively tracking the hand, linking a physical gesture to a system event.	
User movement	Stylized violet footprints, depicted as pairs for a standing position or as a sequence to indicate a walking path	Represents the user's movement and location over time. This is used to visualize the user's trajectory through the environment, or their static position when an interaction occurs.	

Limitations and Future Work

We did not yet validate this toolbox through empirical user studies. Therefore, its impact on communication efficiency remains a hypothesis. Future work includes a controlled user study to evaluate the toolbox's effectiveness empirically. Furthermore, we aim to integrate the system into design tools to bridge visual design and implementation more directly.

Conclusion

This work addresses the challenge of communicating complex, event-driven AR interactions, where traditional 2D tools fall short. The AR Patterns Illustration Toolbox offers an open-source, ECA-based

visual language for communicable and executable design artifacts. All visual elements are assembled in a Miro board to support the interactive creation of AR illustrations.

References

- Ackermann, P. 2023. AR patterns: event-driven design patterns in creating augmented reality experiences. International Conference on Virtual Reality and Mixed Reality, 93-114.
- Apple. (2025, 07 04). Anchor Entity. Retrieved from <https://developer.apple.com/documentation/realitykit/anchorentity>
- ARPatterns. (2025, 07 04). Retrieved from www.ARpatterns.dev
- Krauss, V., Boden, A., Oppermann, L., & Reiners, R. 2021. Current practices, challenges, and design implications for collaborative ar/vr application development. Conference on Human Factors in Computing Systems, 1-15.
- Lee, B., Sedlmair, M., & Schmalstieg, D. 2023. Design patterns for situated visualization in augmented reality. IEEE Transactions on Visualization and Computer Graphics, 1324-1335.
- Microsoft. 2025. Soatuak Anchors. Retrieved from <https://learn.microsoft.com/en-us/windows/mixed-reality/design/spatial-anchors>
- Rajaram, S., Roesner, F., & Nebeling, M. 2023. An augmented reality storyboarding tool for character driven analysis of security & privacy concerns. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 1-15.

Telepresence-Enhanced Network Music Performance

George Xylomenos, Konstantimos Tsioutas, Yannis Thomas, Iakovos Pittaras,
Chalima Dimitra Nassar-Kyriakidou, Athanasia Maria Papathanasiou,
Fotios Bistas, Ioannis Barous, and Anna Kefala

Mobile Multimedia Laboratory, Athens University of Economics and Business, Greece

Corresponding author: xgeorge@aueb.gr

Keywords: XR, 5G, NMP

Introduction

5G can become an enabler for **eXtended Reality (XR)**, especially considering the dropping cost of XR hardware. **Network Music Performance (NMP)** can greatly benefit from 5G, as its latency requirements (up to 30-40 ms) cannot be met by 4G. While audio can be sent uncompressed to reduce delays, 2D video is harder, as it is too bandwidth-heavy to send uncompressed, and compression introduces latency. Volumetric video can transform the NMP experience, if its bandwidth requirements can be met by 5G networks. The **Telepresence-Enhanced Network Music Performance (TENE MP)** project carried out a series of experiments in the Berlin 5G testbed provided by the SPIRIT project, to assess whether the integration of NMP with telepresence over 5G can make NMP a reality.

Testbed setup

In NMP there are two basic communication scenarios, the **Peer-to-Peer (P2P)** scenario where each musician directly communicates with all others, and the **Client-SFU (CSFU)** scenario where musicians communicate indirectly, via an SFU, acting as a relay between them. The SFU must ideally be located close to the endpoints, e.g., at a **Mobile Edge Cloud (MEC)** server.

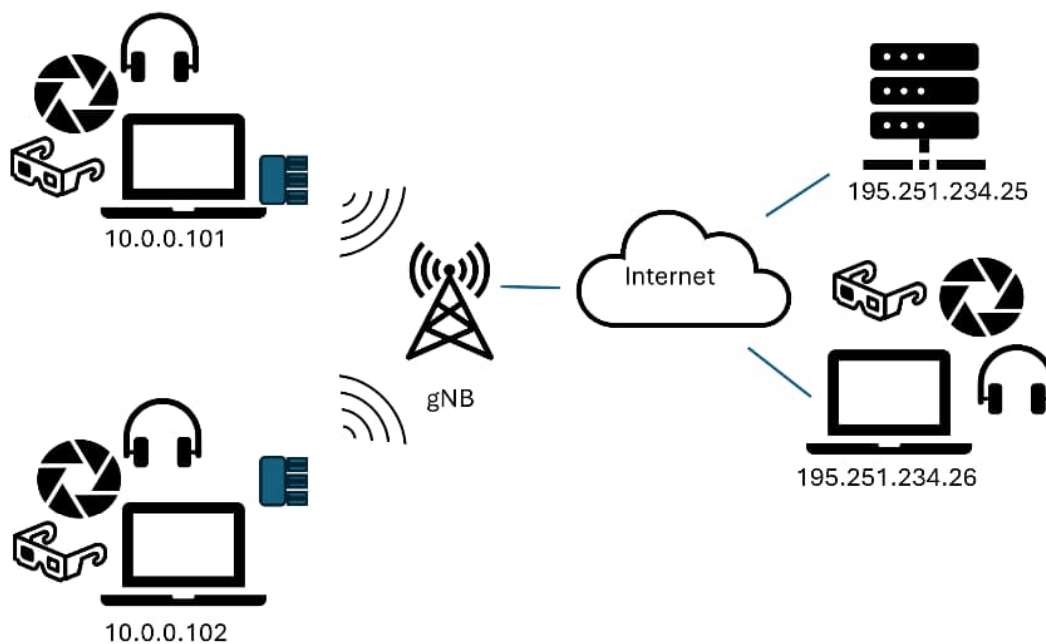


Figure P 38. Athens testbed.

Although our primary goal was to test NMP in the SPIRIT **5G Standalone (5G-SA)** testbed in Berlin, for development and testing we created a testbed in Athens, replicating as far as possible the setup of the Berlin testbed, with Intel RealSense D435 depth cameras, XREAL Air 2 Pro glasses and Teltonica RUTX-50 5G routers. As shown in Figure P 38, the Athens testbed has two 5G endpoints, as well as a server in the MMLab. The endpoints used for the measurements were Asus TUF Gaming A15 laptops, running Ubuntu 24.04.2 LTS, with embedded Realtek sound cards and 720P HD Logitech USB web cameras (for 2D video).

The Athens testbed had three main differences with the Berlin testbed. First, in Athens we only had a **5G Non-Standalone (5G-NonSA)** network (COSMOTE/TELEKOM), while in Berlin the network was 5G-SA. Second, the Berlin testbed included MEC servers in the 5G cell, while in Athens servers were deployed in our LAN. Third, the Berlin testbed was isolated from the Internet, using private IP addresses. The Athens testbed used a public 5G network, which used NAT. Initially, we relied on STUN and TURN servers for NAT traversal, deploying a signaling server (Tsioutas et al, 2025). However, we eventually adopted **UDP hole punching** to establish communication between the endpoints in the Athens testbed.

Measurement procedures

For audio, we measured the **Mouth to Ear (M2E)** latency, which is the time between a user producing a sound and the sound reaching the ears of another user. We used the **reflected pulse** method, shown in Figure P 39: we produce audio pulses, which are sent to a receiver, played back there, captured again and returned to the sender. We record both the original and the returned signal as stereo channels and inspect the recorded waveforms to calculate the round trip delay. With a symmetric connection, half of that is the M2E latency.

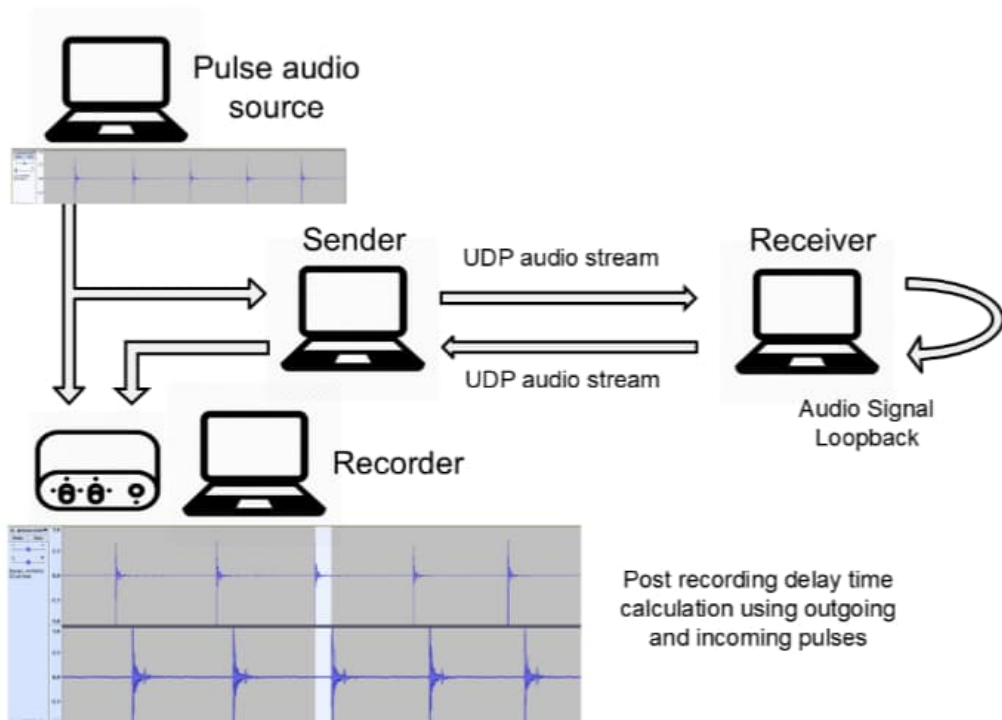


Figure P 39. Audio latency measurement setup

For 2D video, the **Glass to Glass (G2G)** latency is to the time it takes for a frame to be captured by a camera and the corresponding frame to be presented at a screen. We adapted the **flashing LED** method (Bachhuber and Steinbach, 2016), shown in Figure P 40, where a LED is pointed at a web camera, the captured video is displayed at the other end and a light detector captures the LED flash. The same microcontroller lit the LED and monitored the light detector, thus measuring G2G latency. For volumetric video, we tested various options to reduce the video stream size, including dropping color information and reducing resolution; this prevented us from using the flashing LED method. We thus measured volumetric latency by jointly analyzing the logs of the consumer and producer applications (Thomas et al, 2025).

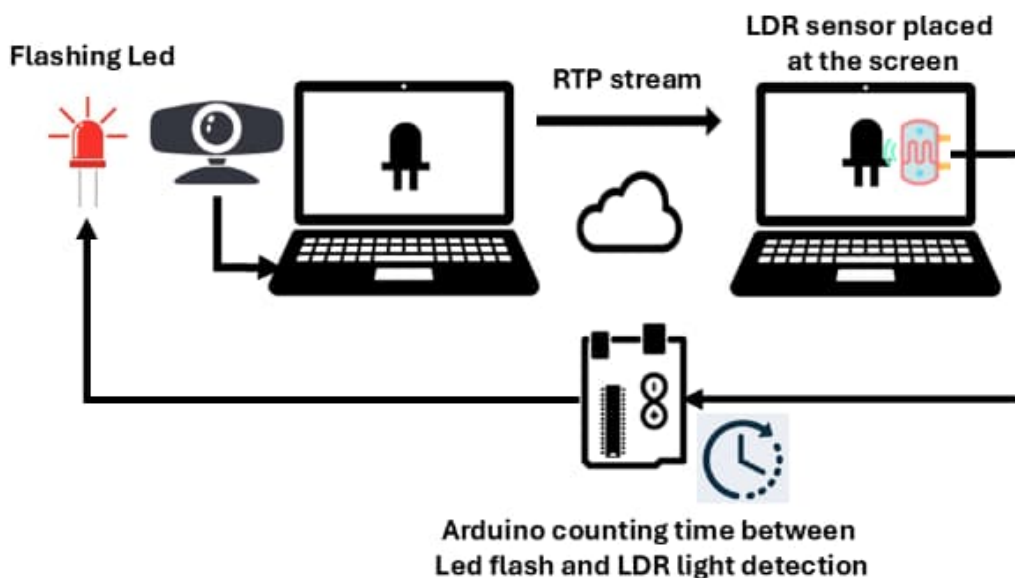


Figure P 40. Video latency measurement setup.

Results

For audio and 2D video, the measurements at both Berlin and Athens used Gstreamer in Linux, in either P2P or CSFU mode. For audio, we used PulseAudio to capture an audio channel at a 44.1 KHz with 16 bits per sample. The buffer size was set to 132 samples, or 3 ms of audio. The RTP packets for the uncompressed audio stream had a size of 276 bytes, including the 12-byte RTP header. In CSFU mode, we used another Gstreamer pipeline at the MEC server; in the Berlin testbed, this was inside the cell, in the Athens testbed it was in our lab. The pipeline simply relayed incoming UDP packets from one client to the other.

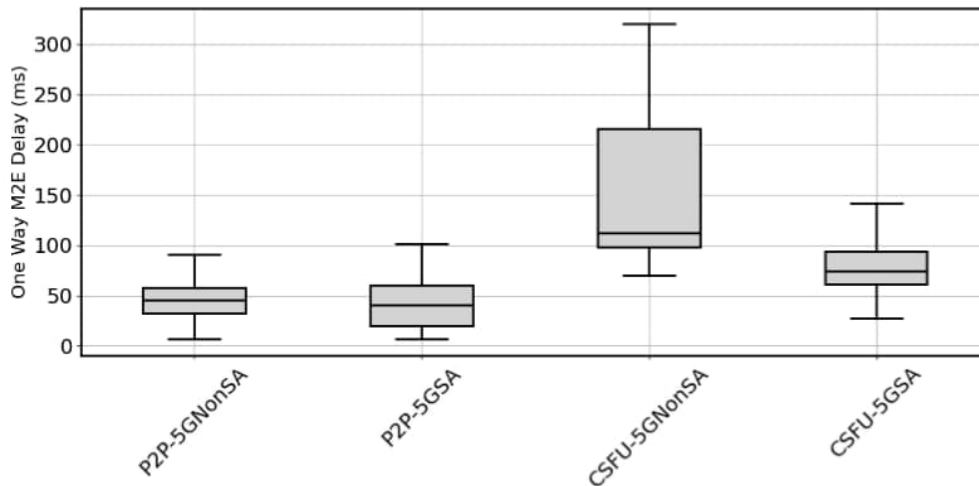


Figure P 41. Audio latency

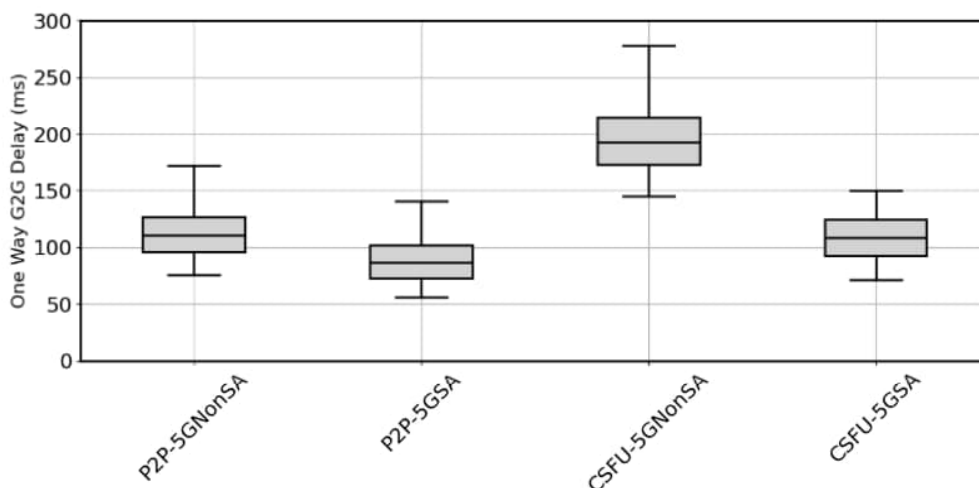


Figure P 42 2D video latency.

Figure P 41 shows boxplots for the M2E audio delay. The median in P2P mode was 40.5 ms for the 5G-SA network, lower than the 45.1 ms for the 5G-NonSA network. In CSFU mode, there was a much larger difference between 5G-SA and 5G-NonSA (74.4 ms vs. 112.2 ms), since in the 5G-NonSA case the SFU was outside the 5G cell.

For 2D video we used the Video4Linux plugins to capture video and H.264 to compress it. In the Berlin testbed we used 1400-byte UDP packets with an RTP payload. For the CSFU topology, we again used a Gstreamer pipeline. In the Athens testbed, that packet size led to losses, therefore we reduced it to 300 bytes; this meant higher overhead due to headers, but led to good video quality. Figure P 42

shows boxplots for the G2G video delay. The median latency in the Berlin 5G-SA testbed (87 ms) was lower than in the Athens 5G-NonSA testbed (111 ms), far more than in for audio. In CSFU mode, we also had a large difference between the two testbeds, since in the Athens testbed the SFU was located outside the 5G network.

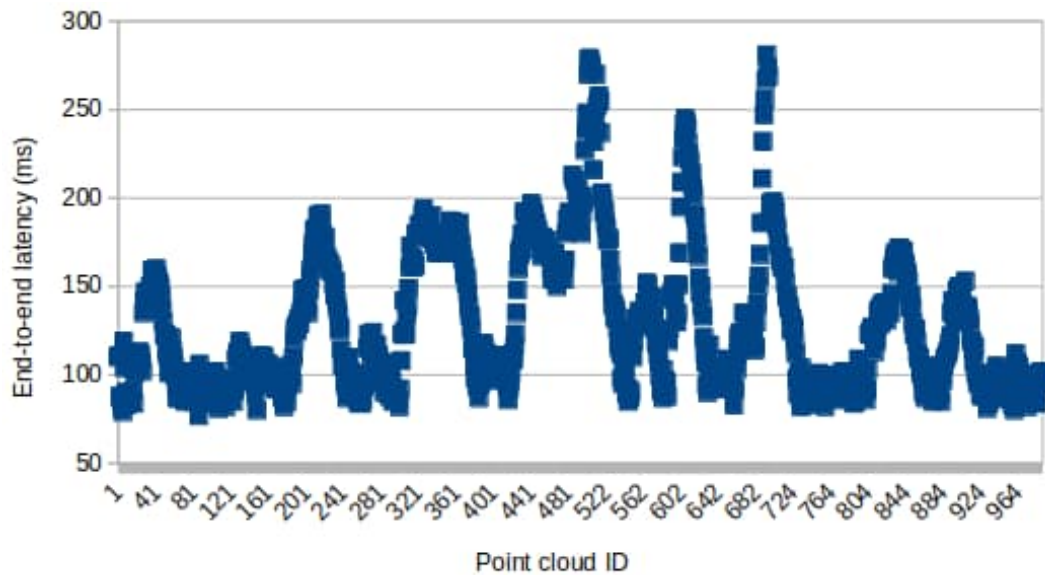


Figure P 43 Volumetric video latency.

For volumetric video, we developed our own **Point Cloud (PC)** streaming tool, using the Intel RealSense SDK, the Google Draco encoder and OpenGL for rendering. The PC frame was captured at 848x480, using a relatively complex scene with 1-2 m of depth, producing around 70% of the maximum PC size. In the experiments, we controlled four parameters: (a) Draco compression levels (1-10) (b) different PC sizes (full frame, 25% and 50% fewer points), (c) 1 or more compression threads and (d) color or no color information; details can be found in (Thomas et al, 2025). Figure P 43 depicts the end-to-end latency of 1000 PCs when using 2 threads for compression with 50% dropped points. The processing latency was roughly 30 ms, supporting 30 FPS, with 97.2% of frames received correctly. The graph suggests that the minimum possible end-to-end latency was roughly 80 ms, albeit with a considerable variance.

References

- Bachhuber, C., & Steinbach, E. 2016. A system for high precision glass-to-glass delay measurements in video communication. Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 2132-2136.
- Tsioutas, K., Thomas, Y., Bistas, F., Barous, I., Xylomenos, G., & Polyzos, G. C. 2025. Network Music Performance Beyond 4G. In 2025 International Wireless Communications and Mobile Computing (IWCMC) (pp. 497-502). IEEE.
- Thomas, Y., & Xylomenos, G. 2025. Ultra-low Latency Point Cloud Streaming in 5G. Proceedings of EuroXR International Conference.
- Tsioutas, K., & Xylomenos, G. 2025. Polyzos. Network Music Performance in 5G Networks. Submitted to IEEE International Symposium on the Internet of Sounds (IS2).

AgriVision: Extended Reality for Efficient and Sustainable Farming

Christina Volioti¹, Vagelis Rossidis¹, Dimitra Perperidou², Nikolaos Nikolaidis¹,
Maria-Theodora Folina¹, Gregory Mygdakos², Apostolos Ampatzoglou¹,
and Alexandros Chatzigeorgiou¹

¹University of Macedonia, Macedonia

²bSpoke Solutions, Greece

Corresponding author: mygdakos@bspoke.gr

Keywords: Extended Reality, Farm Management Information System, Agriculture, User Experience, Decision Support

Introduction

Modern agriculture faces multifaceted challenges, such as climate variability (Khatri et al., 2024), inefficient resource management (Mondal & Palit, 2022), and limited access to real-time decision-making tools (Thilakarathne et al., 2022). The primary agricultural sector is also confronting a series of challenges such as climate change, overpopulation, freshwater scarcity, energy poverty, and urbanization, requiring innovative technological solutions to maintain productivity and sustainability. Despite advancements, current Farm Management Information Systems (FMIS) often lack the capacity to provide real-time, easily understandable and actionable insights (Fountas et al., 2015) through intuitive interaction, especially for young and less experienced farmers, leading to suboptimal decision-making. Companies seeking to leverage FMIS to improve farmers' access to finance, encounter challenges due to a lack of interoperable data systems or additional data collection costs (Souza Filho et al., 2023). FMIS traditionally designed for and used by large-scale operations may not adequately address the specific needs and constraints of small-to-medium farms, creating a gap in accessible agricultural technology (Fountas et al., 2015). The analysis revealed a significant increase in XR technology adoption over the past two years, highlighting its potential within the agricultural sector (Papadimitriou et al., 2025). The findings suggest that XR technologies can enhance productivity, improve training methods, and facilitate remote collaboration in agricultural settings (Tzounis et al., 2022). This emerging trend demonstrates the growing recognition of Extended Reality's transformative potential in addressing traditional agricultural challenges.

In this context, AgriVision addresses this issue aiming to redefine how farmers interact with agricultural data through an innovative XR solution that supports intuitive and immersive visualizations, thereby enhancing their operational efficiency. Through extended reality, farmers can visualize real-time data about their fields directly in their field of view, enabling more informed decision-making processes

(Tzounis et al., 2022). Developed by bSpoke Solutions and the University of Macedonia, and aligned with the EU-funded CORTEX2²⁰ initiative, AgriVision integrates Extended Reality (XR) with FMIS, enhanced by core CORTEX2 components. The solution is available in two versions: a *Pro version* for XR headsets and a mobile-friendly *Light version*, empowering farmers and agri-consultants to make informed and timely, data-driven decisions directly in the field.

AgriVision Overview

AgriVision solution combines XR, FMIS, and real-time collaboration capabilities to enhance productivity and sustainability in agriculture. Combined under AgriVision we are revolutionizing farm operations by streamlining decision-making, optimizing resource management, and enhancing productivity. Through, the FMIS component of AgriVision we deliver five core services to our end-users:

- **Weather Forecasting:** Parcel-level forecasts downscaled based on meteorological and satellite data, enabling proactive planning and climate adaptation.
- **Pest and Disease Forecasting:** Weekly predictive insights based on microclimatic and crop-specific factors, helping to reduce crop losses.
- **Crop Growth Monitoring:** Satellite-based vegetation indices overlaid on field views via AR, allowing detection of stress zones and comparative analysis.
- **Irrigation:** Real-time soil moisture calculations and evapotranspiration tracking, supporting efficient water usage and sustainability.
- **Fertilization:** Vegetation zone-based maps derived from multi-year imagery, enabling precision fertilization with variable rates.



Figure P 44. AgriVision with (a) Microsoft HoloLens 2 and (b) Vizux M400.

To further enhance field collaboration between farmers and agri-consultants and reduce the need for physical consultations, AgriVision incorporates multiple CORTEX2 components, including:

- **Video Conferencing** through the Rainbow TM bubble for real-time remote expert support,

²⁰ <https://cortex2.eu/>

- voice interaction with the **Cortex Virtual Assistant (CoVA)**,
- automated **Meeting Summarization** and
- **User Role Management** to efficiently handle user roles and ensure secure authentication and authorization.

The *Pro version*, deployed on both Microsoft HoloLens 2 (Figure P 44a) and Vizux M400 (Figure P 44b), allows for spatially contextualized, real-time data visualization. For example, crop health maps, and fertilization layers are directly projected onto the physical parcel using AR, significantly enhancing comprehension and response time (Figure P 45a). Meanwhile, the *Light version* (Figure P 45b), built as an Android application, offers full access to FMIS functionalities for mobile users in a cost-effective manner.

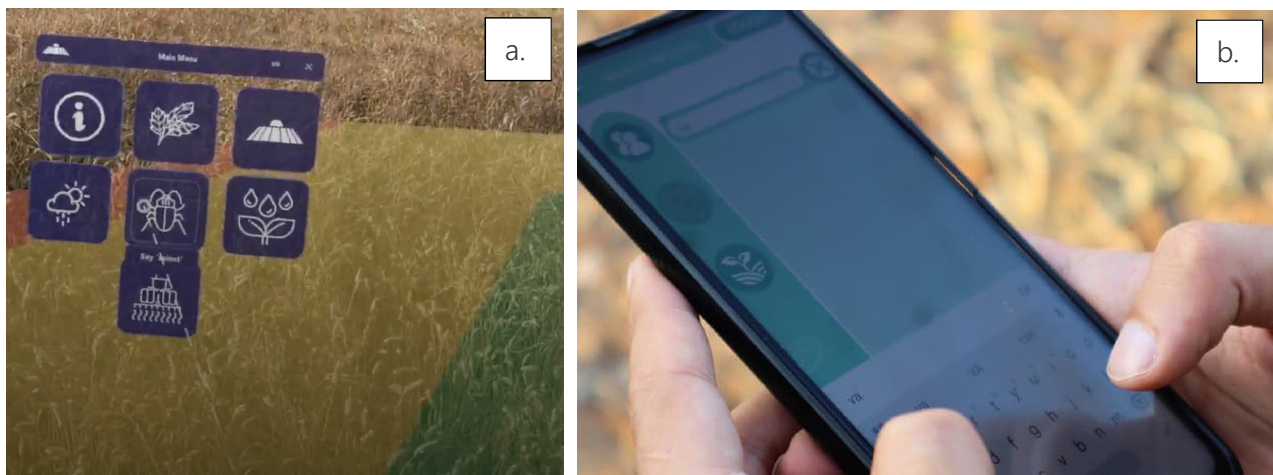


Figure P 45 AgriVision's 2 versions: (a) Pro Version deployed on HoloLens offering real-time enhanced visualization, and (b) Light Version designed for mobile device.

Validation and Evaluation

The evaluation followed a three-phase structure: (a) co-creation with users, (b) limited rollout for field feedback, and (c) final rollout to users. In total, **28 participants** (20 farmers, 6 agri-consultants, and 2 representatives from 2 agri-food companies) tested AgriVision in real-world settings across cotton, wheat, pea, and bean crops. Participants used both Light and Pro versions during real parcel operations, testing key services such as weather information, irrigation schedules, and pest risk assessments. Tasks were monitored for usability, accuracy, and impact. Qualitative feedback was collected via interviews and focus groups, while technical metrics such as response time (e.g., data loading speed) and data accuracy (e.g., weather forecast precision against ground sensors) were recorded.

Key findings from the evaluation highlighted several strengths and areas for improvement in AgriVision. More specifically, in terms of **usability**, farmers appreciated the intuitive design of the Light version, which simplified decision-making processes, while agricultural consultants emphasized the value of the Pro version's immersive XR capabilities for in-situ evaluations. The **impact** of the AgriVision solution was evident across services: crop monitoring and fertilization features enabled instant anomaly detection and reduced resource waste, while pest forecasting improved the timing of treatments, and irrigation scheduling supported more efficient water usage. Regarding **XR feedback**, users preferred the HoloLens over the Vuzix due to its superior visualization quality, despite noting a need for better user training. The Vuzix, although more portable, was limited in display resolution and processing power, leading to its planned discontinuation in future deployments. **End-user suggestions** included the need

for multilingual support, expanded pest and disease databases, and fertilization input customization. Overall user satisfaction was high, particularly for the Light version's ease of use, while the Pro version received praise for its innovation and immersive experience, though it was noted to have a higher learning curve.

Conclusions

AgriVision has demonstrated that XR-enhanced decision-making can transform farm operations across scales and crop types. The Light version offers a reliable, accessible entry point for digital agriculture, while the Pro version, integrated the CORTEX2 components, brings immersive decision support into the field. Real-world validation confirmed technical robustness, and practical relevance. By aligning agricultural practices with next-generation digital tools, AgriVision not only enhances productivity and environmental responsibility but also empowers a new generation of farmers through real-time insight, collaboration, and data-driven action.

References

- Fountas, S., Carli, G., Sørensen, C. G., Tsiropoulos, Z., Cavalaris, C., Vatsanidou, A., Liakos, B., Canavari, M., Wiebensohn, J. & Tisserye, B. A. 2015. Farm management information systems: Current situation and future perspectives. *Computers and electronics in Agriculture*, 115, 40-50. <https://doi.org/10.1016/j.compag.2015.05.011>
- Khatri, P., Kumar, P., Shakya, K. S., Kirlas, M. C., & Tiwari, K. K. 2024. Understanding the intertwined nature of rising multiple risks in modern agriculture and food system. *Environment, Development and Sustainability*, 26(9), 24107-24150.
- Mondal, S., & Palit, D. 2022. Challenges in natural resource management for ecological sustainability. In *Natural resources conservation and advances for sustainability* pp. 29-59, Elsevier. <https://doi.org/10.1016/B978-0-12-822976-7.00004-1>
- Papadimitriou, A., Benos, L., Tsaopoulos, D., & Bochtis, D. 2025. A systematic review of extended reality (XR) technologies in agriculture and related sectors (2022-2024). *IEEE Access*.
- Souza Filho, H. M., Vinholis, M. M. B., Carrer, M. J., & Mozambani, C. I. 2023. Adoption of farm management information systems (FMIS): The case of Brazilian sugarcane farmers. *Outlook on Agriculture*, 52(2), 167-180. <https://doi.org/10.1177/02666669231177864>
- Thilakarathne, N. N., Bakar, M. S.A., Abas, P. E., & Yassin, H. 2022. A cloud enabled crop recommendation platform for machine learning-driven precision farming. *Sensors*, 22(16). <https://doi.org/10.3390/s22166299>
- Tzounis, A., Katsoulas, N., Bartzanas, T., & Kittas, C. 2022. Applications of extended reality (XR) in agriculture, livestock farming, and aquaculture: A review. *Smart Agricultural Technology*, 2, 100070.

Immersive Fundraising: Motivational Mechanisms Converting VR Call-to-Action into Actual Donations

Isabell Aichele, and Wolfgang Schäfer

ZHAW Zurich University of Applied Sciences, Switzerland

Corresponding author: wolfgang.schaefer@zhaw.ch

Keywords: Virtual Reality, Fundraising, Call-to-Action, Donation Motivation, Empathy, Non-profit Marketing

Introduction

Private donation volumes in Switzerland and across Europe have declined for three consecutive years, pressured by wars, crises and donor “compassion fatigue” (Swissfundraising & Stiftung Zewo, 2024; Bircher-Suits, 2025). Humanitarian organizations therefore compete fiercely for scarcer contributions while needing to cultivate long-term trust through transparent, authentic communication (Dethier et al., 2021).

Conventional fundraising—printed mailings, 2-D video spots—often fails to spark the emotional salience necessary to mobilise *first-time* donors, whose initial conversion underpins loyalty programmes (Harken & Avanzanto, 2024). Virtual Reality (VR) is touted as a high-empathy medium capable of putting supporters “on the ground” and triggering powerful perspective-taking (Kandaurova & Lee, 2019; Sooter & Ugazio, 2023). While immersive charity narratives can double donation *intent* relative to flat video (Kandaurova & Lee, 2019), the behavioral funnel from an in-headset call-to-action (CTA) to a *completed* donation remains poorly understood.

RQ — *Which donation-relevant motivational mechanisms convert an immersive VR CTA into a completed donation?*

This study is still a work in progress. At the time of submission of this abstract, data collection is still ongoing. We therefore hope to receive valuable feedback from other participants at the EuroXR 2025 conference.

Theoretical Foundations

We connect three strands of scholarship:

1. **Affective–cognitive donor motivation.** Bekkers & Wiepking’s (2011) eight mechanisms—from warm-glow to perceived impact—explain why people give.
2. **Presence & empathy theory.** Spatial presence, embodiment and perspective-taking intensify empathic concern and prosocial action in immersive media (Cummings et al., 2022; Leveau & Camus, 2023).
3. **Technology-acceptance perspectives.** Perceived ease of use and seamless interactivity moderate the leap from motivation to actual behaviour (Wu et al., 2022).

Figure P 46 synthesizes these ideas into an *immersive donation funnel* with three stages: **Experiential immersion → Motivational appraisal → CTA conversion**. Story design levers (narrator persona, interactive prompts, real-time feedback) are hypothesised to strengthen presence, which then amplifies empathy, perceived impact and agency—each a direct antecedent of donation behavior.

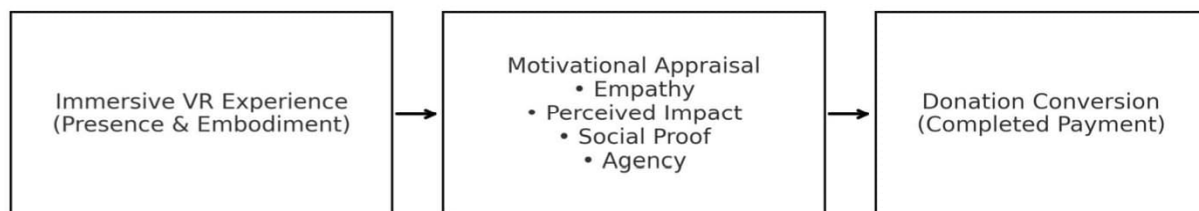


Figure P 46. Immersive Donation Funnel.

Methodology

A three-phase, qualitative-explorative design captures nuanced donor insights.

- **Phase 1 — Scoping Review.** Mapping 46 peer-reviewed VR-fundraising studies revealed four common outcomes: empathy, presence, donation intent and memory (Alliata et al., 2021).
- **Phase 2 — Expert Interviews.** Twelve agreed semi-structured interviews with Swiss NGO fundraisers (n = 7) and XR producers (n = 5) should identify best-practice CTA design and barriers (Granados, 2019).
- **Phase 3 — Focus Group & VR Session.** Five tech-savvy but low-frequency donors will experience a 7-min “Syria relief” 360° narrative (Meta Quest 3). Think-aloud plus post-experience discussion should trace emotional trajectory, usability and willingness to donate. Data will be analyzed via reflexive thematic analysis (Braun & Clarke, 2006, 2021) in MaxQDA.

Preliminary Insights & Expected Contributions

Pilot synthesis and interview pre-tests indicate four core motivational levers:

1. **Embodied Empathy.** First-person victim perspective elicits guilt and responsibility, lifting donation likelihood (Moriuchi & Murdy, 2022; Kandaurova & Lee, 2019).
2. **Perceived Impact.** Real-time feedback such as “CHF 10 builds one water filter” enhances tangible efficacy (Pimentel & Kalyanaraman, 2024).
3. **Social Proof 2.0.** Live donor tickers or avatar co-presence satisfy the need for communal effort (Wu & Dai, 2025).
4. **Seamless CTA Integration.** Eye-triggered or hand-tracking payment flows minimise friction between emotion and action (RaiseNow, 2024).

The study should deliver: (i) a **VR-fundraising motivation framework**, (ii) design heuristics for non-profit XR vendors, and (iii) a research agenda for mixed-methods efficacy trials.

Implications for Practice

- *NGOs*. Embed immersive storytelling early in donor acquisition funnels while pairing emotive perspective-taking with one-click payment flows.
- *XR Agencies*. Prioritise presence-enhancing cues—haptics, spatial audio—that reinforce agency without overwhelming novices.
- *Regulators*. Update fundraising guidance to address data privacy and psychological intensity in immersive solicitations

Limitations and Future Research

Qualitative sampling constrains generalizability. Future experiments should factorially manipulate key levers (e.g., empathy vs. impact cues) across larger donor pools to isolate causal pathways. Cross-cultural replications could test attitudinal variance in collectivist contexts.

References

- Alliata, G., Kenderdine, S., Hibberd, L., & Mason, I. 2021. *Towards immersive generosity: The need for a novel framework to explore large audiovisual archives through embodied experiences in immersive environments*. *PRESENCE: Virtual and Augmented Reality*, 30, 45–59. https://doi.org/10.1162/pres_a_00366
- Auer Frege, I., & Radtke, K. 2024. *WeltRisikoBericht2024*.
- Bekkers, R., & Wiepking, P. 2011. *A literature review of empirical studies of philanthropy*. *Nonprofit and Voluntary Sector Quarterly*, 40(5), 924–973. <https://doi.org/10.1177/0899764010380927>
- Bircher-Suits, B. 2025. Spendenrückgang: Wie Fundraiser gegensteuern können. *Fundraiso*.
- Braun, V., & Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Braun, V., & Clarke, V. 2021. *Thematic Analysis: A Practical Guide*. Sage.
- Cummings, J. J., Tsay-Vogel, M., Cahill, T. J., & Zhang, L. 2022. Effects of immersive storytelling on affective, cognitive, and associative empathy. *New Media & Society*, 24(9), 2003–2026. <https://doi.org/10.1177/1461444820986816>
- Dassler, B., Zangl, B., & van Meegdenburg, H. 2021. Mitleids- und Hilfsmüdigkeit bei humanitären Krisen. *Zeitschrift für Internationale Beziehungen*, 28(2), 246–264.
- Dethier, F., Delcourt, C., & Willems, J. 2021. Transparency of nonprofit organisations: An integrative framework and research agenda. *Journal of Philanthropy and Marketing*, 28(4), 1–14. <https://doi.org/10.1002/nvsm.1725>
- Granados, M. 2019. Virtual Reality als neue Kommunikationsmethode für NGOs. *Horizont*.
- Harken, S., & Avanzanto, D. 2024. *RaiseNow Fundraising-Studie 2024*.
- Kandaurova, M., & Lee, S. H. 2019. The effects of Virtual Reality on charitable giving. *Journal of Business Research*, 100, 571–580. <https://doi.org/10.1016/j.jbusres.2018.10.027>
- Leveau, P.-H., & Camus, S. (2023). Embodiment, immersion, and enjoyment in VR marketing experiences. *Psychology & Marketing*, 40(7), 1329–1343. <https://doi.org/10.1002/mar.21822>
- Moriuchi, E., & Murdy, S. 2022. Increasing donation intentions toward endangered species in VR. *Psychology & Marketing*, 39(7), 1302–1311. <https://doi.org/10.1002/mar.21650>
- Pimentel, D., & Kalyanaraman, S. 2024. How cognitive absorption influences responses to immersive narratives of environmental threats. *Cyberpsychology, Behavior, and Social Networking*, 27(1), 83–90. <https://doi.org/10.1089/cyber.2022.0401>

- RaiseNow. 2024. Die Psychologie des Spendens.
- Sooter, N. M., & Ugazio, G. 2023. Virtual reality for philanthropy: A promising tool to innovate fundraising. *Judgment and Decision Making*, 18, 1–12. <https://doi.org/10.1017/jdm.2023.15>
- Stiftung Zewo. 2021. Digitale Spenden—Aufbruch ins digitale Zeitalter.
- Spendenstatistik 2023.
- Swissfundraising & Stiftung Zewo. 2024. Spendenreport 2024.
- Veigl, S. 2023. Die Kunst des Fundraising. Impact ZHAW.
- Wu, X., Zhang, M., & Shi, S. 2022. Customers' interactive experience in immersive performing art. *Nankai Business Review International*, 15(2), 177–206. <https://doi.org/10.1108/NBRI-03-2022-0031>
- Wu, Z., & Dai, Y. 2025. Nonprofit transparency and donation willingness. *Public Performance & Management Review*, 48(3), 675–704. <https://doi.org/10.1080/15309576.2025.2465750>

Fitting the Future: How AI-Driven Virtual Try-Ons Re-shape Brand Positioning Across Fashion Segments

Natalija Paunovic, and Wolfgang Schäfer

ZHAW Zurich University of Applied Sciences, Switzerland

Corresponding author: wolfgang.schaefer@zhaw.ch

Keywords: Virtual try-on, Augmented Reality, Virtual Reality, Artificial Intelligence, Brand Positioning, Fashion Retail, Consumer Experience

Introduction

Virtual try-ons (VTOs) sit at the epicentre of artificial-intelligence (AI)–augmented-reality (AR) and virtual-reality (VR) convergence in fashion retail, letting shoppers “wear” digital garments on smartphones, smart mirrors or head-mounted displays. The economic upside—higher conversion, lower return rates and richer data—has been tracked for more than a decade (Kim & Forsythe, 2008; Chatterjee, Rana, & Chaudhuri, 2021). Yet VTOs also function as **symbolic artefacts** that can amplify or erode brand meaning, depending on how they materialize core segment values such as *speed* (fast fashion), *exclusivity* (luxury) or *accountability* (sustainable fashion). By zeroing in on VTOs, the paper aims to demonstrate how branding outcomes are as technically and academically pertinent as classic UX metrics.

Conception Background

Grounded in Service-Dominant Logic (SDL) (Vargo & Lusch, 2004), Kapferer’s Brand Identity Prism (Kapferer, 2004) and Porter’s value-chain model (Porter, 1985), we conceptualize VTOs as *operant resources* that orchestrate value-in-use. Prior AR literature focuses on hedonic experience (Rauschnabel, Bittner, & Heinrich, 2022) and operational efficiency; few studies map these mechanics to higher-order brand constructs. Segment sensitivity matters:

- **Fast fashion** leverages VTOs for algorithmic trend mirroring and social virality, aligning with “democratized novelty”.
- **Luxury fashion** demands high-fidelity rendering and human mediation, reinforcing “crafted exclusivity” (Baryannis et al., 2018).
- **Sustainable fashion** positions VTOs as eco-efficiency tools, overlaying CO₂ metrics and provenance data on garments (Williams, Martinez, & Smith, 2023).

Photorealistic rendering breakthroughs (Xu, Ma, & Wallace, 2023) and generative-AI styling suggest that symbolic stakes will grow.

Methodology

Design. A qualitative, exploratory design filtered the project's fourteen semi-structured interviews—nine senior brand executives (strategy, innovation, retail) and five multi-segment consumers—through a VTO lens.

Sampling. Purposive and snowball sampling ensured coverage across fast-fashion giants (e.g., Zara, H&M), luxury maisons (e.g., Gucci, Hermès), and certified sustainable labels (e.g., Patagonia, Armedangels).

Data collection. Interviews (45–65 min, online) combined functional prompts (“Describe your current VTO stack”) with symbolic probes (“What does a VTO say about your brand?”).

Analysis. Following Braun and Clarke's (2006) six-phase approach, 784 initial codes emerged; 143 referenced VTOs or adjacent AR features. These were re-coded into nine thematic clusters and distilled into the **VTO–Brand Positioning Model**. Credibility was bolstered by member checks and an audit trail of analytic memos.

Findings

Fast Fashion – “Instant Play & Social Share”

Participants framed VTOs as a gamified “scroll-to-dress” tool that mirrors TikTok micro-trends within hours (Kim & Forsythe, 2008). The visual language mimics social filters—low-latency but stylized. Consumers accept lower photorealism if novelty and price stay paramount; brand risk arises when identical filters proliferate across rival apps, flattening differentiation.

Luxury Fashion – “Augmented Bespoke”

Luxury managers likened state-of-the-art VR fitting rooms to the “silent maître” that enhances, never eclipses, human craft (Interview L3). Expected standards include sub-millimeter body scanning and studio-grade rendering (Xu et al., 2023). Over-automated chatbots or generic animations erode the aura of scarcity; successful maisons integrate *human-in-the-loop* concierge scripts where AI recommends ensembles and experts deliver them.

Sustainable Fashion – “Proof of Fit, Proof of Ethics”

Sustainable labels deploy VTOs as eco-efficiency levers and transparency enhancers. Overlaying life-cycle-assessment data (Williams et al., 2023) lets shoppers “see themselves and the footprint simultaneously”, boosting trust and willingness to pay. However, opaque AI decision-making triggers “data greenwashing” suspicions among eco-aware audiences.

Cross-Segment Synthesis

1. **Symbolic Alignment.** VTOs strengthen brand identity when design resonates with segment values.
2. **Authenticity Gradient.** Fast-fashion users tolerate stylized filters; luxury demands photorealism and tactile metaphors (e.g., fabric drape physics).
3. **Return–Relationship Trade-off.** VTO-led return cuts are offset by potential loss of human storytelling, especially in luxury contexts (Chatterjee et al., 2021).
4. **XR as Brand Stage.** Generative-AI styling plus spatial computing transforms VTOs into co-creative arenas for narrative co-production (Kouslis et al., 2024).

Discussion and Contribution

Theoretical. By mapping VTO design variables (visual fidelity, data transparency, human hand-off) onto Kapferer's six identity facets, we extend SDL: VTOs not only facilitate but *format* value co-creation by selecting which symbolic resources (heritage stories, ESG metrics, trend memes) surface in micro-interactions (Vargo & Lusch, 2004).

Managerial. The **VTO–Brand Positioning Canvas** (available on request) operationalizes the model: executives plot each planned VTO feature against identity promises, reveal misalignments and iterate before rollout. Fast-fashion brands can prioritize viral filter novelty; luxury maisons should invest in photorealism and human choreography; sustainable labels must make provenance overlays intelligible.

Technological. Advances in neural rendering (Xu et al., 2023) and cross-device spatial anchors lower fidelity barriers, but they also raise consumer expectations—turning VTO quality gaps into salient brand cues.

Limitations and Future Work

Qualitative and Europe-centric data limit generalizability. Next steps include: (1) a conjoint experiment manipulating VTO fidelity, advisor presence and transparency to quantify brand-equity shifts; (2) longitudinal field studies tracking returns, brand sentiment pre- and post-VTO launch; (3) cross-cultural replication in Asian and North-American markets where XR adoption curves differ. Integrating physiological immersion metrics (gaze, heart-rate variability) could further link VTO embodiment to brand emotion.

Conclusion

AI-powered VTOs are no longer peripheral utilities; they are brand theatres in which consumers rehearse and negotiate brand meaning. When choreographed to segment logic—agility for fast fashion, intimacy for luxury, accountability for sustainable fashion—VTOs elevate brand equity; when misaligned, they accelerate distrust. As spatial-computing hardware moves from “tech demo” to mass adoption, the strategic imperative for segment-sensitive VTO design grows, underscoring the contribution of this study to the EuroXR community.

References

- Baryannis, G., Dani, S., Marchese, C., & Tsolakis, N. 2018. Predictive analytics and artificial intelligence in supply chain management. *Computers & Industrial Engineering*, 137, 106024.
- Braun, V., & Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Chatterjee, S., Rana, N. P., & Chaudhuri, R. 2021. Augmented-reality try-on technology and its influence on consumer decision-making. *Journal of Retailing and Consumer Services*, 61, 102549.
- Deepa, S., & Abirami, A. 2024. *The impact of AI on customer experience*. IGI Global. <https://doi.org/10.4018/979-8-3693-2276-5.ch014>
- Kapferer, J.-N. 2004. *The new strategic brand management* (3rd ed.). Kogan Page.
- Kim, J., & Forsythe, S. 2008. Adoption of virtual mannequin technology for online apparel shopping. *International Journal of Retail & Distribution Management*, 36(6), 502–516. <https://doi.org/10.1002/dir.20113>
- Kouslis, P., Pal, R., & Lin, C. 2024. Generative AI in fashion: A value-chain perspective. *International Journal of Production Economics*, 259, 108700.

- Porter, M. E. (1985). *Competitive advantage: Creating and sustaining superior performance*. Free Press.
- Rauschnabel, P. A., Bittner, J. V., & Heinrich, D. 2022. Augmented reality marketing: A decade-long review. *Journal of Business Research*, 142, 569–590.
- Vargo, S. L., & Lusch, R. F. 2004. Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68(1), 1–17.
- Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. 2020. Influence of artificial intelligence on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893–1924. <https://doi.org/10.1108/BPMJ-10-2019-0411>
- Williams, K., Martinez, A., & Smith, P. 2023. Augmented-reality visualisation for sustainable fashion communication. *Journal of Cleaner Production*, 384, 135658.
- Xu, F., Ma, X., & Wallace, B. 2023. Photorealistic rendering for omnichannel AR commerce. *IEEE Transactions on Visualization and Computer Graphics*, 29(7), 1–14.

Auditory Presentation of Stimuli for Cognitive Load Dual Tasks in Redirected Walking

Michelle Meyer, Carolin Reichherzer, and Markus Zank

Lucerne University of Applied Sciences and Arts, Switzerland

Corresponding author: markus.zank@hslu.ch

Keywords: Virtual Reality, Redirected Walking, Cognition

Abstract

As more studies on redirected walking (RDW) in virtual reality (VR) arise, the importance of measuring its impact on cognitive load increases. An examination of existing publications reveals that the stimulus for the cognitive task was always displayed visually. Those existing implementations are prone to issues if used in richer virtual environments (VE) than studies until now used. Instead, an auditory presentation is proposed. We present an analysis of existing research on working memory tasks in RDW, indicating the feasibility of auditory presentation for verbal working memory paradigms.

Introduction

Using real walking as locomotion technique in virtual reality provides a realistic and well-tolerated experience as the proprioceptive and vestibular senses match the visually perceived optical flow caused by moving through a virtual environment (Razzaque et al., 2025). However, this approach limits the VE to the size of the physically available space. Redirected walking (RDW) methods allow to influence the users traversed path and thus create a seemingly bigger space (Razzaque et al., 2025). The factor by which the user's trajectory is altered is called gain. The commonly used gain is just below the perceivable threshold. Without the gain being consciously disruptive to the user the question arises, what the unconscious impact of RDW on the human cognition is.

It has been shown, that cognitive load influences gait in reality (Ebersbach et al., 2025; Nadkami et al., 2010) as well as in VR (Kannabe et al., 2014). Additional cognitive load can be created using the dual task paradigm (Woollacott & Shumway-Cook, 2002) where a concurrent secondary task is added to the walking task. This secondary task is often chosen to be taxing for working memory (WM).

Kourtesis et al. compared established tasks for WM assessment (Digit Span Task, Corsi Block Task, and Deary-Liewald Reaction Time Task) on desktop PC with VR. The results showed convergent validity and they concluded that the assessments in VR replicates the established cognitive measurements (Kourtesis et al., 2025).

In the following section an overview of studies investigating walking in VEs with working memory dual tasks is given. Then auditory stimulus presentation is proposed for studies employing RDW as a response to limitations of visual presentation.

Related work

Overall, few studies have been published on the cognitive impact of walking in VEs or the assessment of said impact with WM tasks so far.

In a CAVE VE Marsh et al. compared the impact of different locomotion techniques by administering a WM dual task. Participants had to remember and recall a sequence of either spatial (patterns) or verbal (random sequence of numbers). They found that locomotion techniques use spatial memory and thus compete with a concurrent spatial tasks (Marsh et al., 2012). To a similar conclusion came Janeh et al. when they studied the effect of cognitive load on velocity over an extended period of walking comparing walking with and without VR (Janeh et al., 2019). As dual task they utilized an auditory two-back (letter) task, thus impacting verbal WM. They recorded a significant decrease in velocity in reality as well as in VR while performing the dual task. Despite comparable results during the dual task, Janeh et al. noted a significant velocity decrease in VR conditions compared to reality, which they interpreted as cognitive demand of the task competing for resources with the increased cognitive load of the locomotion in the VE (Janeh et al., 2019).

Bruder et al. found that RDW significantly impacted verbal and spatial working memory thus requiring cognitive resources. Analysing gait during a dual task they found that the cognitive task had a significant effect on walking behaviour. They noted that participants made significantly more errors in the spatial task compared to the verbal task (Bruder et al., 2015). While Lai & McMahan compared three walking techniques for consumer VR applications they evaluated whether a verbal and spatial two-back task should be used as dual task (Lai & McMahan, 2020). They decided to use the verbal task, because they found the spatial stimuli too challenging in combination with their steering techniques (Lai & McMahan, 2020).

Nguyen et al. employed a dual task of serial seven subtraction. They determined that cognitive load had a significant impact on curvature redirection detection thresholds and participants could use higher gains without noticed during the task (Nguyen et al., 2020). Also studying redirection detection thresholds but using a different cognitive task Mostajeran et al. employed a dual-task spatial WM task while walking with different gains aiming to determine detection thresholds (Mostajeran et al., 2024).

Table P 13. Studies with working memory tasks in VEs.

Paper	Cognitive (WM) task	Stimulus presented
Marsh et al, 2012	Digit span (3-5), Spatial span (3-5)	Visually
Bruder et al, 2015	Verbal 2-back, Spatial 2-back	Visually
Janeh et al, 2019	Verbal 2-back	Auditory
Nguyen et al, 2020	Serial 7-subtraction	n/a
Lai & McMahan, 2020	Verbal 2-back	Visually
Mostajeran et al, 2024	Spatial 2-back	Visually

Findings

Visual stimulus usually takes the form of a UI element with the stimulus symbol or letter. For non-VR scenarios this is displayed on the screen (Kourtesis et al., 2025; Hilbert et al., 2015) while in VR it is often displayed ahead of the user roughly at head height (Figure P 47a) (Bruder et al., 2015; Lai & MaMahan, 2020; Mostajeran et al., 2024). While the procedure is not detailed in the descriptions, it can be assumed from the depictions that the stimulus UI is placed ahead at a fixed location in the VE.

In these scenarios participants followed a straightforward path in very open and visually sparse environments that gave little incentive to look around. Although this matches the typical scenario used in RDW threshold detection, following a single straight path is a rather artificial and limited use case. However, this approach can encounter issues if the VE becomes more narrow or visually busy which will happen for scenarios that are closer to potential applications of RWD. For example, if the stimulus is placed in a fixed location in the VE, curving paths, corners, or junctions where the user may choose between multiple routes can cause problems for the placement of the stimulus (Figure P 47). Placing the stimulus closer to the participant can solve the problem for certain environments but cannot solve the problem of the user looking around and potentially missing one of the stimuli (Figure P 47c).

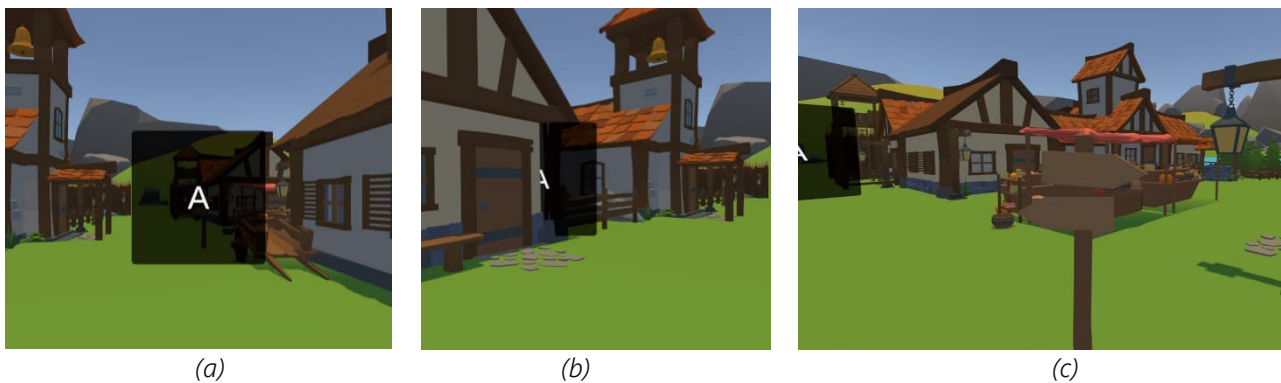


Figure P 47. Demonstration of issues with visually presented cognitive stimulus in VE as commonly used (a) stimulus obstructing scene (b) stimulus partially occluded by scene objects (c) stimulus partially out of view

Additionally, if placed naïvely the stimulus can intersect with the VE as depicted in Figure P 47b. While this could be solved with a shader, that renders the stimulus over the VE, this introduces contradicting depth cues since the stimulus is occluding closer geometry.

Therefore, we propose to use WM tasks with auditory presentation instead of visual stimuli. This could be particularly feasible for verbal n-back, serial 7-subtraction, and digit span (forward and backward) task as those need a small set of stimuli to implement.

The results of multiple unconnected studies give a first indication to the validity of the proposed solution. One study investigated auditory versus visual presented stimuli in WM assessment using Digit Span Backwards (DSB). Nakagawa et al. found indications that the user preferred cognitive strategy is more impactful than the presentation mode (Hilbert et al., 2015).

N-back is a commonly used cognitive task in VR studies as evident in Table P 13. In a study comparing the n-back performance of visual versus auditory performance, Amon & Bertenthal found that participants responded more accurately but slower for auditory presented stimulus than to the visual stimuli. They suggest that auditory stimulus creates more cognitive load than visual and conclude that stimulus modality influences encoding and cognitive processing on WM tasks. However, the performance advantage afforded by the auditory presentation was less pronounced with more difficult

n-back tasks and they noted the tentativeness of this result because the advantage was not significant when analysed with a d' signal detection measure (Amon & Bertenthal, 2018).

One drawback of the proposed solution is that for spatial WM task, which have also been used in RDW studies (see Table P 13), there is no simple conversion to auditory. However, a study comparing gait changes between a working memory dual task and a spatial attention dual task rendered no significant difference and could thus imply that changes in gait are impacted by the sharing of attentional resources rather than the cognitive mode of gait interference (Nadkarni et al., 2010). This suggests that verbal instead of spatial WM assessment tasks, which can be implemented auditory, could be used in RDW studies.

Discussion

While it might be suited to visually present stimulus to create cognitive load while testing for RDW detection thresholds or environments with mainly straight paths, it seems prone to issues if used in VE with variable paths. The proposed solution of presenting auditory stimuli would eliminate these issues for verbal WM task. The results of multiple studies on WM support the feasibility of the proposed auditory presentation with verbal working memory paradigms.

Additionally, we see the concurrent processing of auditory stimulus as relevant and realistic for practical use cases such as training where information should be retained.

Future work

The authors will continue by investigating the cognitive impact of redirected walking (and other locomotion techniques) in a maze where an auditory working memory task will be administered. Additional research is needed to further evaluate scenario-independent assessment implementations for redirected walking.

Acknowledgements

This work was supported by the Hasler Foundation under project number 23049 (haslerstiftung.ch/en).

References

- Amon, M. J., & Bertenthal, B. I. 2018. Auditory versus visual stimulus effects on cognitive performance during the N-back task. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 40). [Online]. Available: <https://escholarship.org/uc/item/95g7x626>
- Assessment, M. 2015. The digit span backwards task. *European Journal of Psychological Assessment*, 31(3), 174-180., <https://doi.org/10.1027/1015-5759/a000223>
- Bruder, G., Lubos, P., & Steinicke, F. 2015. Cognitive resource demands of redirected walking. *IEEE transactions on visualization and computer graphics*, 21(4), 539-544., <https://doi.org/10.1109/TVCG.2015.2391864>.
- Ebersbach, G., Dimitrijevic, M. R., & Poewe, W. 1995. Influence of concurrent tasks on gait: a dual-task approach. *Perceptual and motor skills*, 81(1), 107-113. Accessed: Jul. 01, 2025. [Online]. Available: <https://journals.sagepub.com/doi/epdf/10.2466/pms.1995.81.1.107>
- Janeh, O., Katzakis, N., Tong, J., & Steinicke, F. 2019. Infinity walk in vr: Effects of cognitive load on velocity during continuous long-distance walking. In *ACM Symposium on Applied Perception 2019* (pp. 1-9). <https://doi.org/10.1145/3343036.3343119>.

- Kannape, O. A., Barré, A., Aminian, K., & Blanke, O. 2014. Cognitive loading affects motor awareness and movement kinematics but not locomotor trajectories during goal-directed walking in a virtual reality environment. *PLoS one*, 9(1), e85560. <https://doi.org/10.1371/journal.pone.0085560>.
- Kohn, R.Z., & Whitton, M. C. 2001. Redirected Walking Accessed: Jul. 03, 2025. [Online]. Available: <https://doi.org/10.2312/egs.20011036>
- Kourtesis, P., Lizarraga, A., & MacPherson, S. E. 2025. Immersive Virtual Reality Assessments of Working Memory and Psychomotor Skills: A Comparison between Immersive and Non-Immersive Assessments. arXiv preprint arXiv:2503.06333., <https://doi.org/10.48550/arXiv.2503.06333>.
- Lai, C., & McMahan, R. P. 2020. The cognitive load and usability of three walking metaphors for consumer virtual reality. In 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (pp. 627-638). IEEE. <https://doi.org/10.1109/ISMAR50242.2020.00091>.
- Marsh, W. E., Putnam, M., Kelly, J. W., Dark, V. J., & Oliver, J. H. 2012. The cognitive implications of semi-natural virtual locomotion. In 2012 IEEE Virtual Reality Workshops (VRW) (pp. 47-50). IEEE. <https://doi.org/10.1109/VR.2012.6180878>.
- Mostajeran, F., Schneider, S., Bruder, G., Kühn, S., & Steinicke, F. 2024. Analyzing cognitive demands and detection thresholds for redirected walking in immersive forest and urban environments. In 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR) (pp. 61-71). IEEE. <https://doi.org/10.1109/VR58804.2024.00030>.
- Nadkarni, N. K., Zabjek, K., Lee, B., McIlroy, W. E., & Black, S. E. 2010. Effect of working memory and spatial attention tasks on gait in healthy young and older adults. *Motor control*, 14(2), 195-210., <https://doi.org/10.1123/mcj.14.2.195>.
- Nguyen, A., Rothacher, Y., Efthymiou, E., Lenggenhager, B., Brugger, P., Imbach, L., & Kunz, A. (2020, November). Effect of cognitive load on curvature redirected walking thresholds. In Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology (pp. 1-5). <https://doi.org/10.1145/3385956.3418950>.
- Woollacott, M., & Shumway-Cook, A. 2002. Attention and the control of posture and gait: a review of an emerging area of research. *Gait & posture*, 16(1), 1-14., [https://doi.org/10.1016/S0966-6362\(01\)00156-4](https://doi.org/10.1016/S0966-6362(01)00156-4).

CoVA: LLM Data Augmentation to Enhance User-Defined Intents Recognition for Voice-Based Assistance in Collaborative XR Environments

Alexis Lombard, Galo Castillo-López, Nasredine Semmar, and Gaël de Chalendar

Paris-Saclay University, CEA, List, Palaiseau, France

Corresponding author: gael.de-chalendar@cea.fr

Keywords: virtual assistant; retrieval-augmented generation; intent recognition

Introduction

In this work we present CoVA, a virtual assistant designed to participate in business meetings, remote training scenarios, and industrial collaboration sessions. To ensure versatility across use case scenarios, a virtual assistant should excel at supporting not only predefined conversational patterns for key information delivery and action triggering, but also dynamic navigation through shared documents and structured knowledge sources to remain aligned with user-provided materials related to the assisted work session (previous meeting reports, project-related content, domain-specific graphs). Offering both capabilities requires efficient and reliable intent classification to distinguish in-scope from out-of-scope queries and route them to the appropriate processing pipeline, despite potentially limited and heterogeneous intent examples. Thus, we present an intent-agnostic method to recognize user intent. Additionally, we describe how CoVA implements a Retrieval-Augmented Generation approach to provide information contained in shared documents.

System Architecture Overview

CoVA is implemented as a modular system of containerized microservices orchestrated via Kubernetes supporting real-time speech to text transcription, intent recognition, retrieval-augmented answer generation, and speech synthesis.

Audio input is streamed via WebRTC and undergoes preprocessing, including voice activity and wake-word detection. Wake-word activation relies on custom lightweight models trained with OpenWakeWord and is triggered by the phrase "Hi CoVA." Upon activation, user speech is transcribed in real time and forwarded for query understanding.

Utterances are first analyzed by the intent recognition module to determine whether they correspond to predefined commands. If a matching intent is detected, the system executes the associated response or action, which may include delivering a predefined reply, summarizing documents, executing XR-related commands, performing a web search when no relevant documentation is shared. If no matching intent is found, the system falls back to a retrieval-augmented generation pipeline retrieving semantically relevant information from shared documentation and constraining the generated response strictly to that context.

CoVA includes a web-based configuration interface that enables users to upload documents, define and edit intents, manage language and wake-word settings, and tailor assistant behavior to their specific use case. Each conversation bubble maintains its own configuration and dataset, ensuring data isolation and session-specific customization.

Use Case Scenarios

CoVA is deployed in three pilot settings of the CORTEX²¹ project, each associated with different usage patterns and interaction constraints.

In business meetings, CoVA operates within the Rainbow video conferencing platform and associated XR extensions. Each conversation is managed within a dedicated bubble, where users may upload relevant documents and define intent–response pairs through a web-based configuration interface. CoVA supports in-session question answering over shared documents, predefined response generation based on recognized intents, and high-level actions such as document summarization. Authorized users can update intent configurations dynamically during the meeting.

In remote training scenarios, trainers define the assistant's behaviour by specifying intents tied to instructional materials. Trainees can interact with CoVA via speech to obtain immediate clarification or navigation assistance. Document retrieval supplements predefined answers by providing flexible support based on technical content. To preserve session stability, administrative commands are restricted to trainers only.

In industrial collaboration, the assistant supports maintenance and intervention tasks by allowing experts to upload technical documentation and define procedural intents. In addition, intents can be auto-generated from metadata describing available 3D objects in the virtual environment, enabling voice-based manipulation of the XR workspace. CoVA facilitates both contextual information access and real-time control of immersive elements.

Across all pilots, CoVA maintains strict data isolation through tenant-specific storage, with one tenant per bubble. Intent recognition is prioritized over document retrieval to ensure interpretability and control over the assistant's output.

Intent Recognition in Resource-Constrained Settings

At the core of CoVA's interaction model lies the Intent Recognition Module, which enables deterministic responses to frequently expected queries. Each conversation bubble maintains a tenant-specific set of intents, defined by users through a configuration interface. These intents are composed of representative utterances paired with predefined responses or actions. During runtime, user queries are embedded using a multilingual BGE-m3 sentence transformer (Chen et al., 2024) and matched to precomputed intent vectors, generated using the same model, via cosine similarity. If the similarity score for the closest match exceeds a predefined confidence threshold, the intent is classified as in-scope and the associated action is triggered. The system supports a variety of action types, including uttering predefined replies, providing document summarization, performing retrieval-based answer generation, attributing sources, and manipulating the XR environment.

²¹ <https://cortex2.eu/pilots/>

While this hybrid architecture enables controlled responses for predefined intents and extends coverage through fallback to retrieval-constrained generative answers, the reliability of intent classification remains sensitive to the quality, representativeness, and lexical diversity of user-provided examples. In practice, intent categories are often defined with limited and heterogeneous utterance sets, resulting in suboptimal generalization, particularly in semantically nuanced or overlapping domains. To mitigate this limitation, we investigate data augmentation strategies based on synthetic utterance generation using Large Language Models (LLMs). Our intent-agnostic data augmentation method and its evaluation on existing intent recognition corpora are described in the next section.

Data Augmentation for Intent Recognition

We propose an LLM-based approach to generate synthetic utterances in resource-constrained settings, i.e., few-shot scenarios. Our method uses the name of the intent as well as In-Context Learning (ICL) in the generative task using Large Language Models. We conduct experiments using the instruct version of Mistral 7B on two intent recognition datasets: CLINC150 and BANKING77. The CLINC150 corpus corresponds to 23K utterances which cover 150 intent classes over 10 domains, including domains such as banking, home, travel, and others (Larson et al., 2019). Additionally, the CLINC150 dataset includes out-of-scope utterance examples. However, we do not use such examples in this work, as we are only interested in classifying in-scope utterances. The BANKING77 corpus includes 13K customer service utterances from one single domain, labelled with 77 intents related to banking such as transfer cancellation, pin change, and others (Casanueva et al., 2020).

We simulate the resource-constrained scenario by randomly sampling 1 to 5 utterance examples for every intent. We perform our sampling process 5 times to obtain multiple utterance subsets of sizes 1 to 5, in order to compute more robust evaluations. Such sampled utterances are used as ICL examples for our generative task. We use every sampled utterance subset to generate from 1 to 10 synthetic utterances. To evaluate the utility of our data augmentation method, we fine-tune the uncased version of BERT Base to classify intents, i.e. a multi-class classification task (Devlin et al., 2019). We argue that better sets of training examples provide better performance metrics. Thus, we investigate how generating synthetic utterances to augment the training sets impacts the predictive quality of our classification models. BERT is fine-tuned over 5 different seeds on each sample, and the selected performance metric is averaged over those runs to ensure evaluation robustness. Additionally, we select the best model out of 20 epochs after each fine-tuning according to the macro-F1 score on the validation sets.

We display the performance using different training data samples in Figure P 48 We observe that the macro-F1 scores increase when introducing synthetically generated utterances, regardless of the number of seed utterance examples. We also note that the intent recognition performances tend to stabilize as the number of generated utterances increases. Moreover, we observe that our few-shot generations with more than 1 seed example significantly outperform one-shot and zero shot results, where no seed examples are used. We believe that our intent-agnostic data augmentation method can be used in frameworks where the systems have no control over the precise definition of both intents and domains, as is the case with CoVA.

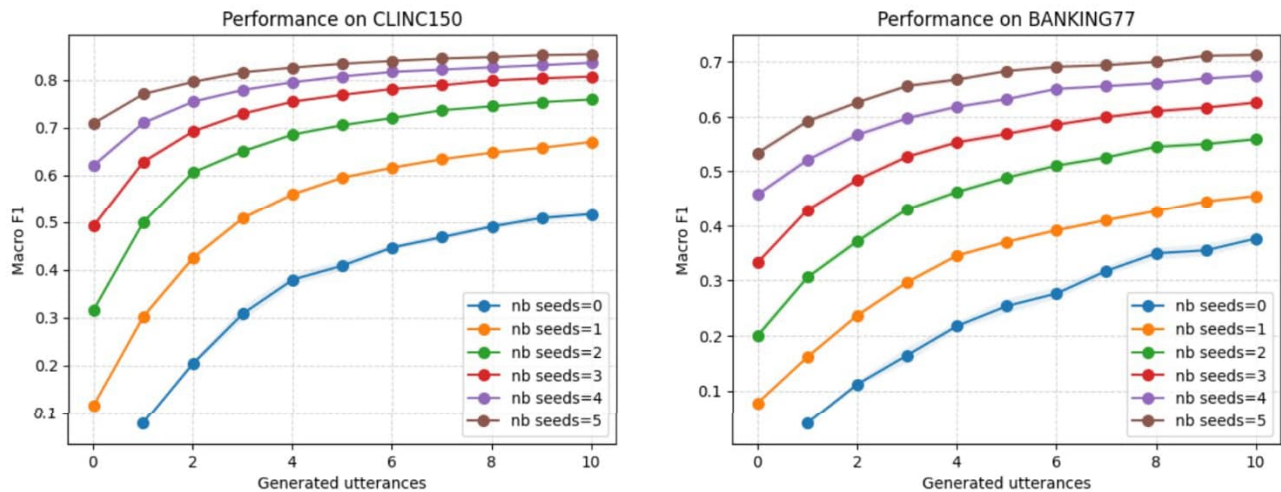


Figure P 48. Macro-F1 score of the intent recognition models fine-tuned on the generated utterances. The evaluation is performed on the test sets of the CLINC150 and BANKING77 datasets.

Retrieval-Augmented Generation

When a user query does not match any predefined intent with sufficient confidence, CoVA activates a retrieval-augmented generation (RAG) fallback mechanism. Each conversation bubble maintains its own namespace within a Weaviate vector store, where the associated documents are indexed. Prior to indexing, the document encoder service performs automated text extraction, segmentation into semantically coherent chunks, and embedding using the multilingual BGE-m3 model. At inference time, the user query is embedded using the same model and compared to the stored document chunks via semantic similarity search. The top-ranked passages are aggregated to construct a prompt, which is passed to a large language model (LLM) for response generation. To minimize latency in interactive settings, the response pipeline operates in a fully streamed manner. Tokens generated by the LLM are buffered into sentence-level units and passed incrementally to the Kokoro TTS service, enabling real-time voice synthesis as the response is being generated. Importantly, generative outputs are strictly conditioned on retrieved content, constraining the model's responses to the user-shared document context. This design mitigates hallucination and promotes factual consistency across all use cases, with heightened importance in contexts such as remote industrial collaboration and remote training, where procedural instructions must remain unaltered.

Integration of Graph-Based Retrieval

To extend CoVA's retrieval capabilities beyond flat document structures, we have integrated support for graph-based knowledge representations using G-Retriever, a retrieval-augmented generation framework for textual graphs. This integration allows users to upload structured graphs—represented as collections of natural language triples—that can be indexed and queried during conversation. Once uploaded, graph elements are embedded and stored using G-Retriever's dual-encoder architecture. When a query is classified as out-of-scope, CoVA activates G-Retriever's subgraph retrieval module. Relevance scores between the query and graph nodes are computed, and a minimal subgraph is extracted using a Prize-Collecting Steiner Tree (PCST) formulation, balancing coverage of relevant nodes with edge sparsity.

This extension supports use cases where procedural logic, taxonomies, or causal links are more naturally represented as graphs rather than unstructured text. The integration is modular and coexists

with existing vector-based document retrieval, allowing mixed-mode interactions within the same session. Graph uploads and activation are managed through the configuration interface. This development expands CoVA's applicability to domains where structured knowledge is central, while maintaining the system's core conversational and document-grounded architecture.

Conclusions and Future Work

In this study, we described the implementation of an intent-agnostic approach to recognize user intents and retrieve information from shared documents for virtual assistants in XR environments. Future works include the investigation of modelling the virtual agent turn-taking management without the use of wake-words, to make its interactions be perceived more realistic, i.e. human-like.

Acknowledgment

This work was developed as part of the CORTEX² project, which has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101070192.

References

- Casanueva, I., Temčinas, T., Gerz, D., Henderson, M., & Vulić, I. 2020. Efficient intent detection with dual sentence encoders. In Proceedings of the 2nd workshop on natural language processing for conversational AI (pp. 38-45). <https://doi.org/10.48550/arXiv.2003.04807>
- Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. 2024. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. arXiv preprint arXiv:2402.03216. <https://doi.org/10.48550/arXiv.2402.03216>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).
- Larson, S., Mahendran, A., Peper, J. J., Clarke, C., Lee, A., Hill, P., ... & Mars, J. 2019. An evaluation dataset for intent classification and out-of-scope prediction. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 1311-1316).

RHYTHM (Real-time HYbrid Test for Harmonic Music)

Jochem Rommens

Dutch Rose Media, The Netherlands

Corresponding author: jochem@dutchrosemedia.com

Introduction

Building on the success of the HYMNE project, RHYTHM explores how immersive technologies can enhance hybrid music performances. The project investigates the impact of visual fidelity and interactivity on fan engagement using real-time volumetric streaming, adaptive formats, and spatial audio within live concert settings.

Objective

To develop and evaluate a scalable, real-time hybrid concert model that:

- Enables artist-audience interaction through live volumetric streaming
- Compares fidelity and latency trade-offs in immersive experiences
- Extends insights from HYMNE for broader application across venues and formats

Methodology

- Redesign of WITT's 'Insomnia' performance using SPIRIT's point-cloud and mesh-based streaming tools.
- Engagement assessment using ITC-SOPI and SPIRIT QoE tools.
- Use of testbeds in Bristol and the volumetric capture studio in Eindhoven to simulate multi-user scalability and latency thresholds.
- Comparative streaming format trials: pre-recorded, mesh, point-cloud
- Live showcase with public audience feedback (planned)

Future goals

- Demonstrate sub-50 ms latency with SPIRIT point-cloud streaming
- Validate hybrid concert scalability with over 20 concurrent users
- Publish results on the role of interactivity in audience engagement
- Define guidelines for low-cost, accessible immersive performances
- Explore broader application in gaming, education, and healthcare

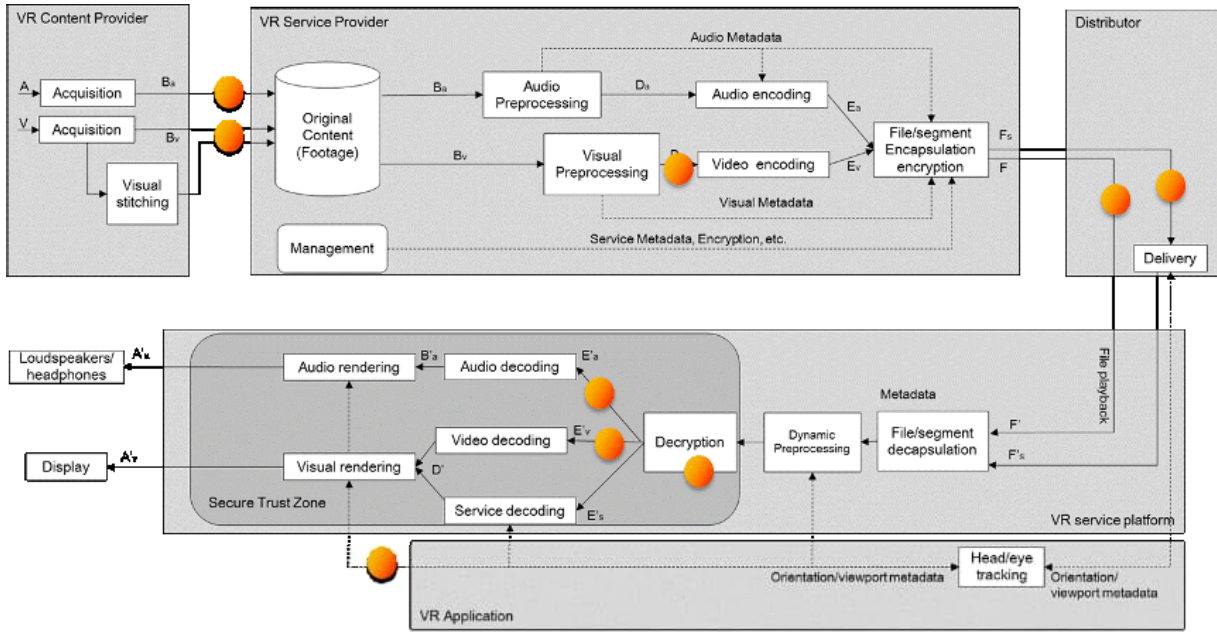


Figure P 49. Technical architecture live volumetric video.



Figure P 50. Hologys Live volumetric video (~10s delay).

Predicting Success Rate of Target Hitting in VR Ball Throwing for Physical and Psychological Support

Yusuke Goutsu, and Tetsunari Inamura

Brain Science Institute, Tamagawa University, Japan

Corresponding author: goutsu@lab.tamagawa.ac.jp

Keywords: Virtual Reality, User Model, Stochastic Approach, Gaussian Process

Introduction

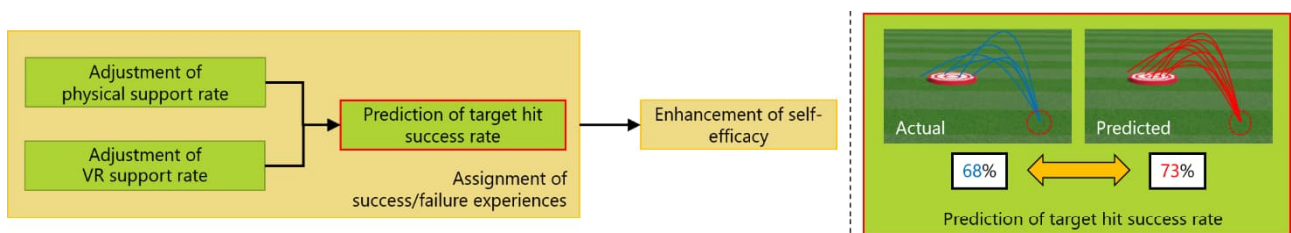


Figure P 51. Roadmap of our work (left) and on-going work (right).

In recent years, many assistive robots for caregiving have been studied, and conventional robot assistance has often focused on improving the user's motor skills and reducing physical stress. However, users have diverse personalities and individual characteristics, and it is not easy for even experienced caregivers or physical therapists to provide support that meets the needs of all users. Therefore, it is necessary to consider not only physical support to achieve specific performance, but also the nature of support from a psychological perspective. As previously related work, (Inamura et al., 2024) focused on the concept of self-efficacy and proposed a VR ball-throwing system that integrates support with an upper limb shoulder exoskeleton robot and VR illusion support to enhance self-efficacy. The roles of the upper limb shoulder exoskeleton robot and VR illusion in this system are to adjust the speed of raising the hand and the speed of the ball, respectively, for underhand throwing. However, these support rates are not adjusted based on the user's ball-throwing skills. When considering adjusting the support rate according to the user's personality and individual characteristics, it is important to be able to predict the degree of adjustment required to achieve the specified success rate of target hitting in order to provide successful and failure experiences and enhance self-efficacy, as shown in Figure P 51. The purpose of this paper is to investigate whether the predicted success rate of target hitting, based on the user's ball-throwing skills, matches the actual success rate, without considering the case where the support rate is adjusted in a simplified manner (refer to the right side of Figure P 51).

Implementation

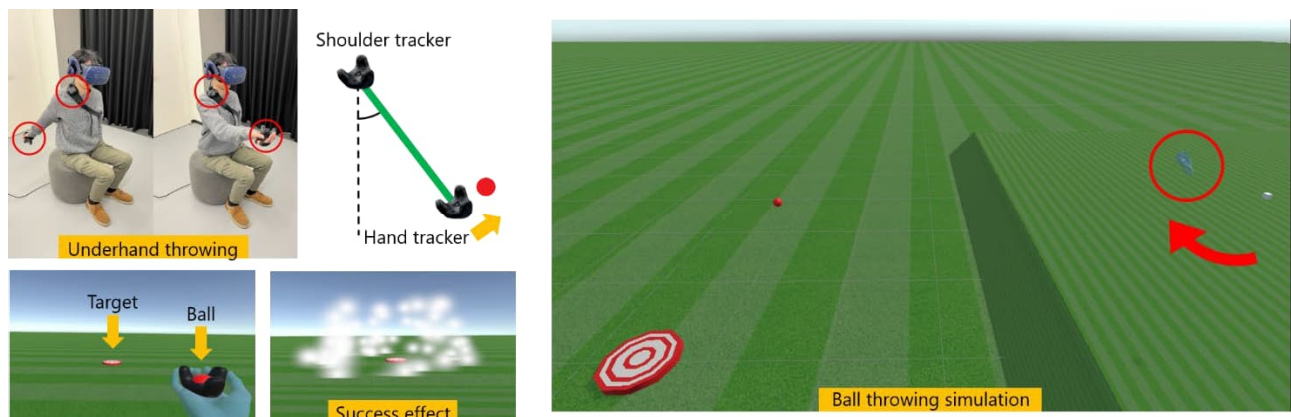


Figure P 52. VR ball-throwing system (left) and ball-throwing simulation using generated data (right)

VR Ball-throwing system

Figure P 52 (left) shows an overview of a system for throwing balls in VR space (Fisciale et al., 2024). The system uses an HTC VIVE Pro Eye head-mounted display (HMD) and VIVE trackers. Users throw balls at targets by swinging their hands downward (underhand throwing) while watching a first-person view of the VR space displayed on the HMD. The position and posture of the trackers are measured in real time, and two trackers are used to track ball throwing motions, as shown in the figure.

We use the Unity platform to construct a VR space. It includes a physics simulation engine, so objects in the VR space are subject to gravity acceleration. In the VR space, the right hand is displayed as shown in Figure 2 (left). When both the angle between the line extending from the shoulder to the hand and the vertical direction of the shoulder and the ball's velocity exceed specified values, the ball separates from the hand. Furthermore, the distance to the target and the size of the target can be freely set in the VR space. When the ball hits the target, an effect occurs as shown in Figure 2 (left).

Learning sequential patterns based on GPDM - Building a skill model -

We use Gaussian process dynamical model (GPDM) (Wang et al., 2005) to construct a skill model. GPDM can predict sequences by learning latent states in a low-dimensional space that captures the essential features of sequential patterns and the dynamics of the states. It is also a probabilistic method that considers prediction uncertainty, with latent states being generated probabilistically as the step progresses. We consider this variety in predictions to be a measure of user skill. In other words, in GPDM-based sequential prediction, given an initial state in the latent space, the model repeatedly conducts probabilistic mappings and transitions between states following a Gaussian distribution to generate sequential data in the observation space.

Probabilistic generation of sequential pattern and success rate prediction using skill model

As training data for GPDM, we utilize sequential data related to hand-raising movements and ball-throwing parabolic trajectories obtained from the VR ball-throwing system. Using the trained GPDM, multiple probabilistic predictions are performed generating a set of sequential patterns related to the hand's movement and the ball's trajectory. Based on the Monte Carlo method, multiple sequential patterns are probabilistically generated (sampled) and the target hit success rates are predicted by determining whether they are successful. Specifically, the success decision is based on the ball's landing point, the target's center point, and the target's size. For the ball's parabolic trajectories, the success rate is calculated directly from the sequential data of the ball's position. For the hand-raising movements,

we perform a ball-throwing simulation in VR space using the generated data of the hand and shoulder positions (see Figure 2 on the right), and predict the success rate based on the results.

Experimentations and Preliminary Results

We use a GPDM-based skill model that has learned sequential data related to ball-throwing movements to confirm (1) whether differences between the unskilled state and the skilled state are reflected in the skill model and (2) which sequential training data (hand-raising movement or ball trajectory) is better for predicting success rates of target hitting.

Procedure

The experiment was conducted with six healthy participants (five men and one woman) who had no physical problems wearing VR equipment. Participants were given instruction time prior to the actual experiment to ensure they could perform the ball-throwing task correctly. The experiment consisted of three sessions, with 20 ball throws performed in each session. The VR ball-throwing system settings were adjusted to increase difficulty from the instruction phase, with the target distance and target size set to 6[m] and 70[cm], respectively.

The number of successful hits was recorded for each session. In addition, for the first and last sessions, sequential data related to the hand-raising movement and the ball's parabolic trajectory were used for GPDM training to construct skill models for each participant. Next, multiple probabilistic sequential predictions were performed using the trained skill models to generate sequential data for the hand and shoulder positions and the ball's position. Here, the number of sequential data points generated was set to 40. Furthermore, the conditions for the angle and velocity when the ball leaves the hand were set to approximately 45[deg] and 1.5[m/s], respectively.

Result

Table 1 (see numbers in parentheses) shows the results of the target success rate for session 1 and 3 during the learning data collection. When comparing session 1 with session 3, the success rate tended to increase as the number of ball throws increased, indicating that the target throwing skill was acquired.

Table 1 also shows the results of predicting the success rate of hitting the target using both a skill model trained on sequential data related to hand raising movements and ball-throwing simulations (left side of the table), as well as a skill model trained only on sequential data related to ball-throwing parabolic trajectory (right side of the table). Since the VR environment settings at the time of prediction (such as the distance to the target and the size of the target) were the same as those during data collection, the same success rate as in the table is expected. In the table, the average error between the actual success rate and the predicted success rate was calculated for each skill state when session 1 and session 3 were designated as before and after skill improvement, respectively. The results in the table show that the method predicting based on the ball's parabolic trajectory yielded prediction values relatively close to the actual success rate, while the method predicting based on the hand-raising movement resulted in errors that cannot be ignored.

Table P 14. Predicted success rate of target hitting and average error between predicted and actual success rates

ID	Success rate predicted from hand-raising movement [%]		Average error	Success rate predicted from ball parabolic trajectory [%]		Average error
	Session 1	Session 3		Session 1	Session 3	
1	15.0 (40.0)	25.0 (55.0)	27.5	47.5 (40.0)	50.0 (55.0)	6.3
2	12.5 (40.0)	25.0 (55.0)	28.8	27.5 (40.0)	55.0 (55.0)	6.3
3	0.0 (10.0)	2.5 (15.0)	11.3	5.0 (10.0)	20.0 (15.0)	5.0
4	20.0 (45.0)	10.0 (55.0)	35.0	45.0 (45.0)	55.0 (55.0)	0.0
5	12.5 (30.0)	15.0 (60.0)	31.3	45.0 (30.0)	65.0 (60.0)	10.0
6	15.0 (35.0)	35.0 (70.0)	27.5	22.5 (35.0)	75.0 (70.0)	8.8
All			26.9			6.0

Conclusion and Future Work

In our VR ball-throwing system that integrates both the upper limb shoulder exoskeleton robot and VR illusion, it is important to adjust the support rate based on the user's ball-throwing skills. To this end, this paper investigates whether a skill model trained on sequential data of hand raising movements and ball parabolic trajectories can predict success rates of target hitting for users with different skill levels, assuming that the support rate is not adjusted in a simplified scenario. The experiments revealed that (1) the GPDM-based skill model can reflect user skills and that (2) the method of predicting the success rates based on ball parabolic trajectory results in smaller errors with actual performance.

Future challenges include (A) whether the success rates can be accurately predicted even when support rates are adjusted, (B) how to provide success and failure experiences through the adjustments of support rate to enhance self-efficacy, and (C) how to improve the method of predicting the success rates based on hand-raising movement.

Acknowledgement

This work was supported by JST Moonshot R&D Grant Number JPMJMS2034.

References

- Fiscale, V., Inamura, T. & Soccini, A. M. 2024. Adaptive training in virtual reality through dynamic alien motion support. In Proc. of 2024 IEEE Gaming, Entertainment, and Media Conference (GEM), pp.1-5.
- Inamura, T., Murakami, H., Wen, W., Okabe, T., Suzuki, M., Yamamoto, J., Teramae, T., Hirata, Y., Kubota, N. & Noda, T. 2024. Enhancing self-efficacy through integrated physical robotics and virtual success illusions: Implementation and discussion from an ethical perspective. In Proc. of the International Conference on Robot Ethics and Standards.
- Wang, J., Hertzmann, A. & Fleet, D. J. 2005. Gaussian process dynamical models. In Proc. of the 18th International Conference on Neural Information Processing Systems (NIPS), pp.1441–1448.

Demonstrating Immersive Teleoperation of Robots via Adaptive Point Cloud Streaming

Nunzio Barone, Walter Brescia, Gabriele Santangelo, Antonio Pio Maggio,
Ivan Cisternino, Luca De Cicco, and Saverio Mascolo

Politecnico di Bari, Italy

Corresponding author: nunzio.barone@poliba.it, walter.brescia@poliba.it, luca.decicco@poliba.it

Keywords: Point cloud real-time streaming; Immersive teleoperation; mobile robots;

Introduction

Autonomous Mobile Robots (AMRs) are increasingly deployed across various domains for their ability to navigate and perform tasks without direct human control, relying on sensors like LiDARs and stereo cameras. However, in complex environments involving occlusions, narrow passages, or subtle obstacles, autonomy may fail, requiring human teleoperation.

To improve remote operation, immersive technologies such as Head Mounted Displays (HMDs) and volumetric video (e.g., point clouds) provide operators with real-time, spatially rich feedback. However, streaming volumetric content poses challenges due to its high bandwidth and low-latency demands, especially on resource-limited robotic platforms.

This work proposes a WebRTC-based teleoperation framework to stream point cloud data from mobile robots to HMDs in real time. The system integrates a full streaming pipeline for transmitting video, telemetry, and control data; a bandwidth-adaptive sampling strategy for reducing latency; and a real-time rendering module for HMDs.

Achieving the necessary end-to-end latency below 150 ms (5GAA, Musicant et al., 2023) is particularly difficult due to the size of point cloud data. Existing compression methods such as G-PCC (ISO/IEC 23090-9:2023), V-PCC (ISO/IEC 23090-5:2025), and Google Draco²² offer high compression rates but are impractical for real-time decoding on embedded systems. Other approaches, like the one in (De Fré et al., 2024), use WebRTC for low-latency volumetric streaming in videoconferencing. However, assuming static background is unsuitable for teleoperation, where retaining full environmental context is essential for safe robot control.

In contrast, our system leverages WebRTC's media track and data channel to jointly stream sensory and control data with low latency, without compromising scene detail. The system was validated through

²² <https://github.com/google/draco>

experiments on a real AMR and a remote HMD—including remote tests across countries— proving its scalability and effectiveness for immersive robot teleoperation under constrained network and hardware conditions.

The proposed framework

The proposed framework, presented in Figure P 53, enables real-time teleoperation of a mobile robot using volumetric video streamed to a commercial head-mounted display (HMD). The system consists of two main components: a mobile robot equipped with an Nvidia Jetson Xavier NX and a Meta Quest 3 HMD. The robot captures, compresses, and streams point cloud data, while the HMD receives and renders it to provide immersive spatial awareness to the operator.

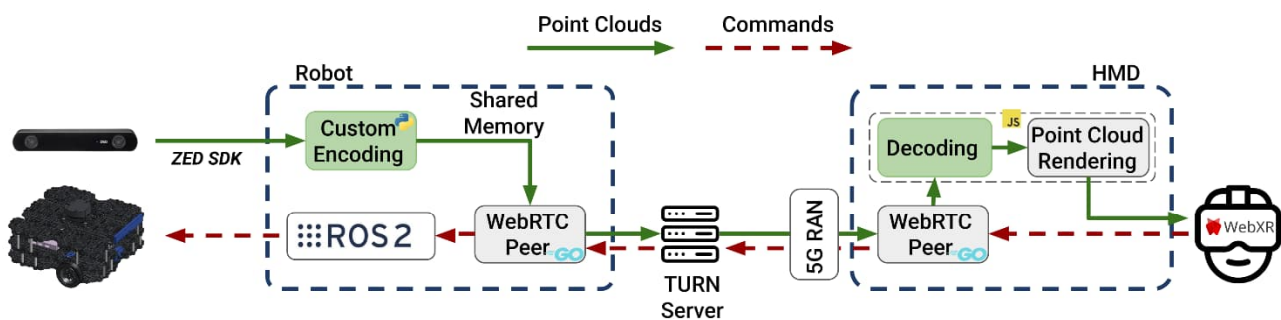


Figure P 53 The overall immersive teleoperation system employed in the demonstration.

Robot's Software Stack

On the robot side, the software stack is composed of a point cloud producer, ROS2 nodes, and a WebRTC peer implemented in GoLang. The point cloud producer uses the ZED SDK to generate 3D point clouds at 18 FPS from stereo images captured at 1080p. To reduce bandwidth, the system applies quantization and filtering strategies: each point's position and color is quantized to 16 bits, significantly reducing the data size. Additionally, a filtering mechanism prioritizes points closer to the robot based on a depth-weighted (z-coordinate) importance function, ensuring the most relevant data for navigation is preserved while distant points are dropped.

Let the validity v represent the importance, then we define it as:

$$v = \frac{1}{\left(\frac{z}{\max(z)}\right)^\lambda + \epsilon}, v \in \mathbb{R}^{n \times 1}$$

where, ϵ is a small positive constant to avoid division by zero and λ is a tunable parameter that controls the steepness of the distance-based prioritization. The validity values of all points below the guaranteed distance D are then brought to the maximum value of v for all points to associate them with the highest priority.

The GoLang WebRTC component establishes the peer connection using Pion, which does not natively support point clouds. To work around this, a Custom Media Track is implemented to send 3D data over RTP using a 90 kHz video payload type. Each point cloud frame is split into 1200-byte RTP packets, where each packet contains 150 points encoded with 8 bytes per point (6 for coordinates and 2 for color). Custom headers attached to each packet carry metadata including frame number and byte

offsets. Pion's interceptors—GCC, TWCC, NACK, and PLI—enable congestion control and bandwidth estimation, which is sent back to the Python encoder through shared memory to dynamically adjust the bitrate. The GoLang peer also opens three WebRTC DataChannels: one for receiving JSON-encoded movement commands, one for receiving bandwidth feedback from the HMD, and one for controlling the start and stop of the media stream.

The signaling and connection setup uses a WebSocket-based protocol. The robot connects to the signaling server on port 3001, sending its name and SDP offer. The client connects on port 3002, retrieves the list of available robots, and completes the SDP negotiation and ICE candidate exchange via the server. The robot can also update its offer to support reconnection, and deregisters either explicitly or when its WebSocket disconnects.

HMD's software stack

On the HMD side, the WebRTC peer acts as a local bridge between the streaming data and the Web user interface. Upon initialization, it establishes three WebSocket connections: the control socket, telemetry socket, and streaming socket. The control socket manages robot selection, triggers WebRTC initialization, and updates the UI with the list of available robots. It also forwards only the selected robot's RTP stream while suspending others, reducing bandwidth usage. The telemetry socket transmits teleoperation commands from the HMD to the robot via the bridge. The streaming socket receives the incoming point cloud packets, reconstructs the coordinates and color data using preallocated buffers, and sends the final frame—structured as 75% positions and 25% colors—to the rendering pipeline.

Point cloud decoding begins with the HMD calculating the number of points from the buffer size and casting the raw data into typed arrays. This approach minimizes memory access by referencing the same memory block directly. The data is then structured into 3D geometries, updating the position and color attributes in real-time. Rendering is handled by WebGL, with a custom ShaderMaterial composed of a vertex and fragment shader. The vertex shader adaptively adjusts point size based on scene scale and depth, enhancing visual clarity and depth perception.

To maximize performance, point cloud decoding is handled inside a WebWorker, while rendering is executed on the main browser thread. The two components communicate using asynchronous messages and zero-copy buffer transfers, ensuring smooth frame updates and reducing the risk of motion sickness caused by rendering delays.

Demo execution



Figure P 54. Example of navigation scenario.

The proposed demonstration replicates a typical teleoperation scenario, such as the one depicted in Figure P 54, involving two robots: the Turtlebot and, for logistic reasons, a stereo camera connected to a laptop acting as a second robot. The user will be able to seamlessly switch between the two robots to experience different points of view.

In particular, the following operational flow will be carried out:

- Turtlebot autonomous navigation: While the robot is performing autonomous navigation the user can monitor it without any intervention;
- Robot switching: It will be possible to switch between the PoV of the Turtlebot and the PoV of a camera connected to a laptop, representing a different robot;
- User Teleoperation: When desired, the user can take control of the robot using the controllers of the Headset to send linear and angular velocity setpoints to the robot.
- Parameter tuning: the user can fine-tune the λ and D values to assess the impact of those parameters on the user experience.

Conclusions

We propose and demonstrate a WebRTC-based teleoperation framework that allows real-time transmission of point clouds acquired from a stereocamera mounted on a mobile robot to an operator wearing an HMD device used for rendering the received point cloud. The framework comprises a distance-based filter to adapt the bitrate of the acquired point cloud stream to the available network bandwidth. Points are dropped with a probability that is function of the distance considering that points farther from the sensors are less likely to have a short-term impact on obstacle avoidance. The demonstration allows users to change the distance-based filter parameters to experiment their effect on user experience while teleoperating the robot.

Acknowledgements

This work has been funded by the European Union through the DIANE project winner of the Open Call n.1 of the SPIRIT project (101070672).

References

- 5GAA. 2020. 5GAA: Tele-operated driving (tod): System requirements, analysis and architecture, <https://5gaa.org/tele-operated-driving-tod-system-requirements-analysis-and-architecture/>.
- Barone, N., Brescia, W., Santangelo, G., Maggio, A. P., Cisternino, I., De Cicco, L., & Mascolo, S. 2025. Real-time Point Cloud Transmission for Immersive Teleoperation of Autonomous Mobile Robots. In Proceedings of ACM MMSys 2025.
- De Fré, M., van der Hooft, J., Wauters, T., & De Turck, F. 2024. Scalable mdc-based volumetric video delivery for real-time one-to-many webrtc conferencing. In: Proceedings of the 15th ACM MMSys.
- ISO/IEC 23090-5:2025. Group, M.P.E.: Information technology - coded representation of immersive media - part 5: Visual volumetric video-based coding (v3c) and video-based point cloud compression (v-pcc), <https://www.iso.org/standard/89030.html>, March 2025
- ISO/IEC 23090-9: 2023. Group, M.P.E. Information technology - coded representation of immersive media - part 9: Geometry-based point cloud compression (g-pcc), <https://www.iso.org/standard/78990.html>, March 2023
- Musicant, O., Botzer, A., Shoval, S. 2023. Effects of simulated time delay on teleoperators' performance in inter-urban conditions. Transportation Research Part F: Traffic Psychology and Behaviour 92, 220–237. <https://doi.org/10.1016/j.trf.2022.11.007>

VR CHARLIE: Chat Assistant for Digital OSCE Examinations

Armin Grasnick, Ali Mehmood Khan, Daniela Becker, Julia Drejer, and Julia Nuss

IU International University of Applied Sciences, Germany

Corresponding author: armin.grasnick@iu.org

Keywords: Human-Machine Interaction, Wearable Computing, Virtual Reality (VR), Speech-to-Text (STT), Text-to-Speech (TTS), Glasgow Coma Scale (GCS)

Abstract

The VR system CHARLIE enables an examiner-free, automated Objective Structured Clinical Examination (OSCE) for the Glasgow Coma Scale (GCS) in a virtual space without the need for dummies or actors. Developed in Unity with an external speech recognition model for speech-to-text and an internal text-to-speech feature, CHARLIE enables realistic, voice-driven interaction with a virtual patient avatar that simulates various GCS scenarios. Students determine the GCS score by assessing the avatar's verbal and non-verbal responses mirroring real-life clinical examinations; the assessments are automatically compared to the preset ground truth. Pilot studies with students and clinicians showed transcription accuracy of up to 98% and demonstrated high user engagement. Currently, VR CHARLIE is being used to assess the GCS domains of orientation and speech, with development being extended to motor responses and various scenarios. The system works on desktop and in VR, is scalable for further OSCE cases and is to be integrated into standard medical assessment processes.

Background and System

Skills education and examination as part of Bloom's taxonomy for competence development (Bloom, 1956; Anderson et al., 2001; Krathwohl, 2002) have become fundamental components of medical curricula, ensuring students can demonstrate essential clinical competencies (Winterton et al., 2006). Among various approaches, the Objective Structured Clinical Examination (OSCE) is recognized as the global standard for high-stakes, hands-on assessment in medical faculties (Harden et al., 1975; Khan et al., 2013), with specific scenarios such as the Glasgow Coma Scale (GCS) increasingly being assessed in this format (Jain & Ivarson, 2025). Traditionally, OSCEs require real-time, in-person interaction: an examiner observes a student performing clinical tasks on a dummy or actor-patient, rating performance instantly with a standardized matrix. This process is resource-intensive and logistically challenging, especially for large student cohorts.

Recent breakthroughs in deep neural networks for speech recognition and synthesis (Hinton et al., 2012; Xiong et al., 2017; Chan et al., 2016; Van den Oord et al., 2016; Shen et al., 2018; Zen et al., 2013) now make it possible to shift OSCEs into fully digital and remote environments enabling upscaling.

While recent systematic reviews highlight the growing impact of virtual patient (VP) simulations in medical curricula, these are almost exclusively web- or menu-based; immersive, VR-based, speech-driven assessment scenarios (such as for the GCS) remain absent from the literature (Lee et al., 2020; Faferek et al., 2024). Our VR CHARLIE system addresses this gap by enabling examiner-free, fully automated assessment in immersive VR for OSCE scenarios.

Implemented in Unity, VR CHARLIE utilizes OpenAI Whisper models for reliable, real-time speech-to-text transcription and Unity's built-in text-to-speech capabilities, enabling smooth, speech-based interaction in GCS scenarios with a responsive virtual patient avatar. In the current prototype, CHARLIE mimics different GCS states, with the actual score randomly preset by the system. During the exam, students interact with CHARLIE through speech and motion and must determine the patient's GCS score based solely on the avatar's verbal and nonverbal reactions. After submitting their assessment, the student's evaluation is automatically compared to the preset (ground truth) score, which can later be reviewed by instructors or examiners for verification or grading purposes.

Pilot Evaluation and Future Development

In early pilot deployments—drawn from physicians in neurological practice and our B.Sc. Physician Assistants students—participants were able to perform realistic GCS evaluations with CHARLIE by checking patient orientation for time/place/person/situation and verbal response (including free-text dialogue), with transcription accuracy rates of up to 98% when using cloud-based STT services. Feedback from initial user sessions suggests that students considered the interaction with the speech-based avatar to be engaging and helpful for clinical reasoning. They reported on a straightforward exam workflow as well as timely and comprehensible manner system responses.

To date, VR CHARLIE can assess two of the three GCS domains (orientation and speech). Ongoing development aims to implement full motor assessment, further enhance avatar realism, and to introduce randomized and adaptive scenarios that reflect authentic clinical variations. The platform is designed for both desktop and immersive VR use and can be scaled for other OSCE cases and learning environments. The next steps planned include user studies with larger cohorts to quantify learning gains and acceptance relative to traditional OSCEs, as well as integration into the regular assessment pipeline.

VR CHARLIE is unique in enabling end-to-end, examiner-free digital OSCE assessment, thus reducing the need for physical infrastructure and staff resources, expanding access to high-quality skills assessment, and offering a scalable platform for digital skills assessment (see Table 1 for development roadmap).

CHARLIE in Action

To illustrate the current functionality and setup of VR CHARLIE, Figure P 55 shows the virtual OSCE assessment environment. The left panel depicts the immersive 3D scene as experienced by students, while the right panel offers a close-up of the virtual patient avatar during interaction.

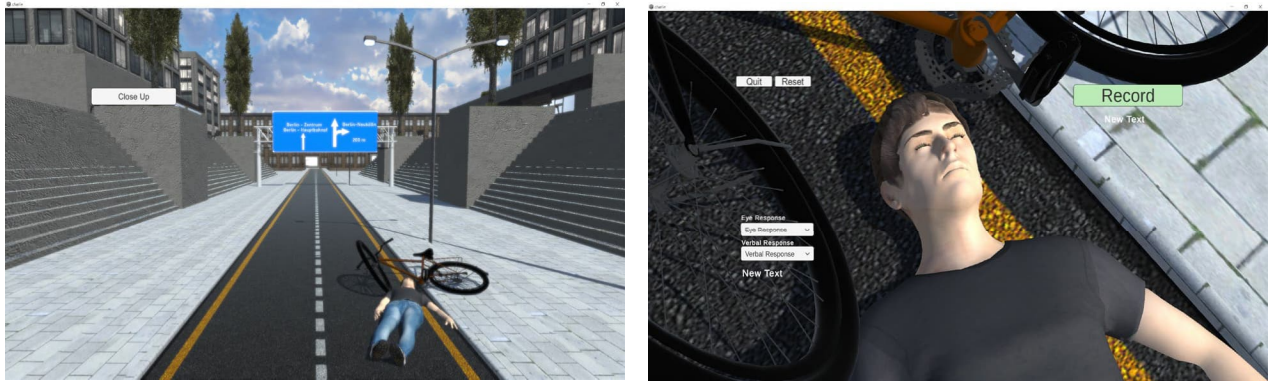


Figure P 55 3D-Scene (left) and Close up (right)

CHARLIE's Core Features

The core implemented features include full support for speech-based assessment of GCS orientation and verbal response using realistic scenarios. Speech recognition in the current online version (cloud-based STT) achieves up to 98% transcription accuracy, enabling reliable natural-language interaction with the patient avatar. Work is ongoing to develop the assessment of motor responses, improve avatar realism, and expand to more diverse and adaptive clinical situations. At present, direct feedback to learners is limited to visual cues and verbal reactions from the avatar; future enhancements will explore adaptive, AI-driven feedback mechanisms and extend support to additional languages and offline (edge) deployment environments.

Table P 15. Core Features.

Feature	Status	Next Steps
GCS Orientation/Verbal	Implemented	Enhance scenario diversity and realism
GCS Motor Response	In development	Add motor assessment and facial cues
Speech Technology	98% (cloud STT)	Edge models, multilingual extension
Feedback	Visual + Verbal	AI-augmented, adaptive feedback

References

- Bloom, B.S. 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. Longmans, Green, and Co., London.
- Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Richard, M., Pintrich, P., Raths, J., Wittrock, M. 2001. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Addison Wesley Longman, New York.
- Chan, W., Jaitly, N., Le, Q., Vinyals, O. 2016. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4960–4964. <https://doi.org/10.1109/ICASSP.2016.7472621>
- Fąferek, J., Cariou, P. L., Hege, I., Mayer, A., Morin, L., Rodriguez-Molina, D., ... & Kononowicz, A. A. (2024). Integrating virtual patients into undergraduate health professions curricula: a framework synthesis of stakeholders' opinions based on a systematic literature review. BMC Medical Education, 24(1), 727. <https://doi.org/10.1186/s12909-024-05719-1>

- Harden, R.M., Stevenson, M., Downie, W.W., Wilson, G.M. 1975. Assessment of clinical competence using objective structured examination. *British Medical Journal*, 1(5955), 447–451. <https://doi.org/10.1136/bmj.1.5955.447>
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6), 82-97. <https://doi.org/10.1109/MSP.2012.2205597>
- Jain, S., & Iverson, L.M. 2025. Glasgow Coma Scale. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing. <https://pubmed.ncbi.nlm.nih.gov/30020670/>
- Khan, K.Z., Ramachandran, S., Gaunt, K., & Pushkar, P. 2013. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. *Med Teach*. 35(9), e1437-46. <https://doi.org/10.3109/0142159X.2013.818634>
- Krathwohl, D.R. 2002. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice* 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
- Lee, J., Kim, H., Kim, K.H., Jung, D., Jowsey, T., Webster, C.S. 2020. Effective virtual patient simulators for medical communication training: A systematic review. *Med Educ* 54(9), 786–795. <https://doi.org/10.1111/medu.14152>
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Wu, Y. (2018, April). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 4779-4783). IEEE. <https://doi.org/10.1109/ICASSP.2018.8461368>
- Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 12, 1. <https://arxiv.org/abs/1609.03499>
- Winterton, J., Delamare-Le Deist, F., & Stringfellow, E. 2006. *Typology of Knowledge, Skills and Competences: Clarification of the Concept and Prototype*. Cedefop Reference Series; 64. Office for Official Publications of the European Communities, Luxembourg. https://www.cedefop.europa.eu/files/3048_en.pdf
- Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., & Zweig, G. 2017. The Microsoft 2016 conversational speech recognition system. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5255–5259. <https://doi.org/10.1109/ICASSP.2017.7953159>
- Zen, H., Senior, A., & Schuster, M. 2013. Statistical parametric speech synthesis using deep neural networks. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7962–7966. <https://doi.org/10.1109/ICASSP.2013.6639215>

Individual Influences in Highly Immersive VR: Two Exploratory Studies on Personality, Perception, and Presence

Xinyu Bi, Hannah Hayes, and Carlo Campagnoli

School of Psychology, University of Leeds, United Kingdom

Corresponding author: c.campagnoli@leeds.ac.uk

Keywords: Virtual Reality, Presence, Spatial Presence, Individual Differences, Cybersickness

Introduction

As part of the wider XR4HUMAN initiative to promote equitable, inclusive, and human-centered Extended Reality (XR), understanding the variability in user experience is paramount. The efficacy of VR applications is largely dependent on eliciting a sense of presence, or "being there" (Slater, 2009), yet the influence of individual user traits remains a key challenge in developing universally effective experiences. While it is often argued that the influence of these traits is diminished by the power of modern, highly immersive systems, this work-in-progress, developed within the XR4HUMAN framework²³, presents two complementary studies that investigate whether and how psychological factors remain significant predictors of presence. The first study explores broad personality traits such as Curiosity (Berlyne, 1960), Empathy (Davis, 1983), and Neuroticism (Costa & McCrae, 1980), while the second study takes a more focused approach, examining the role of Trait Absorption (Tellegen & Atkinson, 1974) in the context of a more specific construct, *spatial presence* (Wirth et al., 2007). Together, these studies aim to clarify the role of individual differences in immersive VR, laying the groundwork for larger-scale investigations needed to build more robust, human-centered models of user experience.

Methods & Demo

Two exploratory correlational studies were conducted with a combined sample of 85 university undergraduates.

Materials & Procedure

Participants first completed validated self-report measures for a range of personality traits. Study 1 (N=45) focused on Curiosity (Seligman, Park, & Peterson, 2004), Empathy (Goldberg et al., 2006), and Neuroticism (DeYoung, Quilty, & Peterson, 2007). Study 2 (N=40) measured Trait Absorption (Tellegen

²³ <https://xr4human.eu/>

& Atkinson, 1974) and depressive symptomatology (Beck, Steer, & Brown, 1996; Topper et al., 2014). All participants were then immersed in the same "Leonardo da Vinci" VR workshop scenario²⁴ using either a Meta Quest 2 (Study 1) or Meta Quest 3 (Study 2) headset.

Immediately following the VR exposure, participants completed questionnaires measuring presence, using either the general Presence Questionnaire (Witmer & Singer, 1998; Study 1) or the more specific MEC-Spatial Presence Questionnaire (Vorderer et al., 2004; Study 2). Finally, to align with the objectives of the XR4HUMAN project, all participants completed its dedicated Post-XR Survey to gather critical data on the overall user experience, including prior VR exposure and any physical discomfort (cybersickness).



Figure P 56. The "Leonardo da Vinci" VR Studio Environment.

The Conference Demonstration

For the EuroXR 2025 demo, attendees will be able to:

1. Engage with the key questionnaires used across both studies, including the Post-XR survey.
2. Experience a brief version of the "Leonardo da Vinci" VR painting task on a Meta Quest headset.
3. Explore the anonymized datasets and the statistical models from both studies.

Data Analysis

The data were analyzed using methodologies appropriate to each study's hypotheses. Study 1 employed hierarchical multiple regression to test the direct and interactive effects of personality traits on presence. Study 2 used a moderated mediation model (PROCESS Model 14) to test the relationships

²⁴ Stega, <https://stega.io/davincivr/>

between absorption, attention, depression, and spatial presence, with post-hoc analyses conducted to explore alternative models.

Results

As is often the case in exploratory studies with limited sample sizes, the primary hypothesized models did not all achieve statistical significance, but they revealed promising and distinct insights. In Study 1, the personality traits of Curiosity, Empathy, and Neuroticism did not emerge as significant predictors of general presence ($p > .05$), though the data revealed trends in the hypothesized directions (e.g., for curiosity, $\beta = .241$, $p = .124$). In Study 2, while the primary model investigating trait absorption and attention was not significant, post-hoc analysis uncovered a highly promising interaction effect. Trait absorption was found to significantly moderate the relationship between users' perception of the system's immersion and their resulting spatial presence. For individuals with medium-to-high levels of absorption, a stronger perception of the system's quality led to significantly greater spatial presence ($p < .001$), an effect not present for those with low absorption.

Discussion

Taken together, these two studies challenge the notion that individual differences become negligible in highly immersive VEs. Instead, they suggest a more nuanced reality where user traits interact with technology in complex ways. The findings from Study 1 highlight that the direct effects of broad personality traits may be subtle and easily overshadowed by powerful situational factors like cybersickness or the specific emotional context of the VR experience. The emotionally neutral "Leonardo da Vinci" experience may not have contained the necessary stimuli to robustly activate trait-level responses for constructs like Empathy or Neuroticism.

The results from Study 2 provide a compelling example of a more indirect mechanism. Here, a trait like absorption does not influence presence directly, but rather shapes how a user perceives the technology itself, which in turn facilitates the feeling of being there. It appears individuals high in absorption are more attuned to the sensory and technical qualities of the VE, leading them to perceive it as more realistic. It is this enhanced perception of the system, shaped by individual traits, that facilitates a stronger sense of spatial presence. This work provides a strong rationale for continued, well-powered investigation into both direct and indirect pathways. Understanding these interactions is critical for the XR4HUMAN project's goal of developing guidelines for human-centered design, informing the creation of more effective, inclusive, and personalized applications for all users.

References

- Beck, A. T., Steer, R. A., & Brown, G. 1996. Manual for the Beck Depression Inventory-II. Psychological Corporation.
- Berlyne, D. E. 1960. Conflict, arousal, and curiosity. McGraw-Hill.
- Costa, P. T., & McCrae, R. R. 1980. Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 38(4), 668–678.
- Davis, M. H. 1983. Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. 2007. Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880–896.

- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. 2006. The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Seligman, M. E. P., Park, N., & Peterson, C. 2004. The Values In Action (VIA) classification of character strengths. *Ricerche di Psicologia*, 27(1), 63-78.
- Slater, M. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 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.
- Topper, M., Emmelkamp, P. M. G., Watkins, E., & Ehring, T. 2014. Development and assessment of brief versions of the Penn State Worry Questionnaire and the Ruminative Response Scale. *British Journal of Clinical Psychology*, 53(4), 402-421. <https://doi.org/10.1111/bjc.12052>
- 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: Measurement, Effects, Conditions (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>
- Witmer, B. G., & Singer, M. J. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225-240.

CORE-MHC: CO-Facilitated and REmote Mixed-Reality Mental Health Care Interventions

Samuel Navas-Medrano¹, José Luis Soler-Dominguez¹, Josep Carreres-Fluixà¹,
Marta García-Ballesteros¹, Lorena Adam-Ferri², Laura Moreno-Renart²,
Susana Ortega-Bataller², Jesús Rienda-Jimena², and Patricia Pons¹

¹*Instituto Tecnológico de Informática, Spain*

²*Fundación SASM, Spain*

Corresponding author: ppons@iti.es

Keywords: remote therapy, mixed reality, extended reality, emotional regulation, biofeedback.

Introduction

Despite meaningful progress in the field of mental healthcare, access to effective treatment remains deeply unequal. Several structural barriers continue to limit access, such as a persistent shortage of qualified mental health professionals (Harp 2023), limited-service availability in rural and underserved areas (Palomin et al. 2023), and the enduring stigma associated with seeking psychological support (Hazel et al. 2022; Mezuk et al. 2021). Extended Reality (XR) technologies offer a promising avenue to address these limitations. In particular, remote XR-based interventions have the potential to reduce stigma by presenting therapy as an engaging and immersive experience (Rodríguez-Rivas et al. 2024; Rodríguez-Rivas et al. 2022). These interventions can also extend the reach of mental health services to communities that are often excluded from traditional care (Altaf Dar et al. 2023; Hendl and Shukla 2024). Additionally, by enabling at-home therapeutic activities, XR can ease the burden on healthcare infrastructure, decrease waiting times, and improve the overall efficiency of mental health service delivery (Bratt and Kalmendal 2023).

Within XR, Mixed Reality (MR) introduces a distinct approach to therapeutic design, especially for remote settings. By blending digital elements with the user's real environment, MR allows individuals to engage in structured mental health activities without leaving the comfort and familiarity of their personal space. This continuity with their surroundings can help foster a greater sense of safety, agency, and emotional regulation, factors that are particularly important in mental health contexts. While MR can also support co-present interactions by allowing users to maintain visual contact with one another, the present work focuses solely on remote use cases, where such features are secondary to spatial and emotional continuity.

In this work, we introduce CORE-MHC, a novel MR platform developed to support remote, gamified, multi-user therapeutic activities. Designed within the framework of the CORTEX2 initiative, CORE-MHC

reimagines remote therapy as a collaborative, engaging, and context-sensitive experience, using MR to create a meaningful bridge between therapeutic goals and the user's everyday environment.

System Design

CORE-MHC was co-designed in collaboration with Fundación SASM, a mental healthcare provider serving individuals with chronic psychiatric disorders through a range of services, including day centers, assisted living, and residential care. Their multidisciplinary team, comprising psychologists, occupational therapists, among others, was actively involved in the design process to ensure the platform addressed real therapeutic needs and operational realities. Together, we defined two remote therapy modalities to accommodate both therapist preferences and clinic resource allocation: one in which therapists use a standard PC while patients connect via head-mounted displays (PC-to-HMD), and another where both therapist and patient engage using MR headsets (HMD-to-HMD). The therapeutic scenario chosen for this initial implementation was animal-assisted therapy, as it allows professionals to work across multiple dimensions of mental health—such as emotional regulation, cognitive stimulation, and psychomotricity—within a single, flexible framework. The system integrates real-time biofeedback to provide therapists with objective measurements of the patient's physiological state (ECG and respiration rate), supporting more personalized and responsive interventions.



Figure P 57. PC-to-HMD modality (left), HMD-to-HMD modality (right)

User Study

We performed a user study in order to systematically assess the system in terms of effectiveness, usability and overall user satisfaction, from both patient and mental health professional (MHP) perspectives. The methodology employed a mixed-methods approach, combining quantitative data collection through standardized questionnaires with qualitative insights derived from open-ended feedback and observations. This comprehensive evaluation was conducted using different system configurations, specifically the PC-to-headset and headset-to-headset modalities, to determine their respective strengths and areas for improvement in facilitating remote mental health care interventions.

The evaluation involved a total of 10 patients and 4 mental health professionals, who participated in 24 sessions in total, with average durations of 20 to 25 minutes. For patients, the assessment included the Self-Assessment Manikin (SAM) (Lang, 1980) administered both before and after the sessions to measure emotional states related to valence, arousal, and dominance. Post-session, patients also completed the UMUX-LITE (Usability Metrics for User Experience - Lite) questionnaire (Lewis et al., 2020) to evaluate perceived overall usability, and an adapted version of the Temple Presence Inventory (TPI) (Lombard et al., 2000) to assess their sense of presence within the virtual environment.

For mental health professionals, usability was primarily measured using the System Usability Scale (SUS) (Brooke, 1996). User experience aspects were captured through the Short User Experience Questionnaire (UEQ-S) (Schrepp et al., 2017), and their sense of presence was also assessed using an adapted subset of questions from the Temple Presence Inventory (TPI) questionnaire (Lombard et al., 2000). In addition to quantitative data, qualitative feedback was gathered from professionals through open-ended questions after the sessions, providing valuable context to the numerical results.

Results

In the evaluation by MHPs, SUS results showed significantly higher usability for the PC-to-HMD modality (mean: 87.5) compared to the HMD-to-HMD one (mean: 60.83). As shown in Figure 1 (left), the PC-to-HMD scored higher across nearly all questions, with more consistent responses (SD: 7.83 vs. 22.85). This suggests greater reliability and fewer usability issues for PC-to-HMD, potentially reflecting individual preferences or technical difficulties in the headset-to-headset condition. Open-ended responses supported these findings: professionals appreciated the immersive nature of headset-to-headset interactions but preferred the PC version for its familiarity, ability to take notes, and reduced technical demands.

UEQ-S results further confirmed these differences in perceived user experience. The 8-item questionnaire (4 pragmatic, 4 hedonic) showed that the PC-to-HMD modality matched the headset-to-headset in efficiency, novelty, and clarity, while scoring slightly higher in other areas, particularly ease of use and user interest. These results align with the SUS findings and reinforce that while both modalities are usable and innovative, the PC-to-HMD setup provides a more familiar and consistently positive experience for professionals.

Quantitatively, the PC-to-HMD scored 2.5 in pragmatic qualities and 2.75 in hedonic ones, compared to 1.92 and 2.0 for the headset-to-headset setup. Both scored “excellent” on the UEQ benchmark, confirming CORE-MHC’s ability to support both usability and enjoyment.

Regarding presence, results from the adapted TPI questionnaire showed similar outcomes across both modalities, with variation more dependent on the individual therapist. All professionals reported that their partner felt present and that virtual objects felt tangible. Most also felt connected to their avatar, immersed in the session, and had a clear sense of “being there.” The only exception came from a professional who experienced technical issues, likely affecting their sense of immersion.

Regarding patients, The SAM questionnaire, administered before and after each session, showed significant changes in emotional state across all three measured dimensions—valence, arousal, and dominance, regardless of the intervention modality. Valence scores decreased significantly post-session in both modalities (PC-to-HMD: 3.09 → 1.73; HMD-to-HMD: 2.91 → 1.64) what reflects a more positive auto-perception. Similarly, arousal scores increased (PC-to-HMD: 2.36 → 3.91; HMD-to-HMD: 3.27 → 3.64), showing an increase on calmness. Dominance scores, indicating perceived control, also rose post-session (PC-to-HMD: 3.27 → 4.09; HMD-to-HMD: 3.27 → 4.00).

The UMUX-LITE showed high perceived usefulness and ease of use for both setups. The PC-to-HMD configuration scored slightly higher in both categories (Usefulness: 4.73 vs. 4.27; Ease of Use: 4.55 vs. 4.18), suggesting a marginal usability advantage.

Results from the TPI-derived presence questions revealed that users experienced a strong sense of immersion across both setups, with the PC-to-HMD configuration yielding slightly higher scores in key dimensions such as plausibility, co-presence with the therapist, and engagement.

Conclusion

Post-sessions feedback further highlighted the system's therapeutic potential. Professionals emphasized the system ability to enhance engagement, emotional connection, and motivation. While there was an initial learning curve, most therapists adapted quickly and felt increasingly comfortable using the system. The HMD-to-HMD setup offered a more immersive and empathetic experience, fostering a stronger sense of shared space. However, some therapists reported physical discomfort and technical interruptions. In contrast, the PC-to-HMD mode was praised for its practicality, ease of documentation, and alignment with typical therapy workflows. Ultimately, the choice of modality for MHPs may depend on personal preference, as both are suitable for achieving therapeutic goals, though the PC-to-HMD modality may be more favorable when consistency and ease of use are critical.

Patients' evaluations of the CORE-MHC system revealed consistent positive trends across both setups, with notable differences in emotional impact, perceived usability, and overall preference. Better results on the PC-to-HMD modality may reflect greater familiarity with traditional screen-based interactions or reduced technical complexity in that configuration. Despite this, both modalities were rated highly overall, confirming the CORE-MHC app's general accessibility and effectiveness from the user perspective.

These findings suggest that while the PC-to-HMD modality may provide a slightly more stable and usable experience, the HMD-to-HMD configuration offers a richer and more emotionally resonant interaction for patients. This highlights a trade-off between usability and experiential quality, suggesting that future iterations of CORE-MHC may benefit from optimizing the headset-to-headset setup to preserve its immersive strengths while addressing any remaining technical or usability issues.

Acknowledgments

Funding for this project is provided through the European Union's Horizon Europe research and innovation programme, under the CORTEX2 Open Call 1 (Grant Agreement no. 01070192).

References

- Altaf Dar, M., Maqbool, M., Ara, I., & Zehravi, M. 2023. The intersection of technology and mental health: enhancing access and care. *International journal of adolescent medicine and health*, 35(5), 423-428.
- Bratt, A., & Kalmendal, A. 2023. eHealth interventions can potentially provide cost-effective and accessible support for stress-related problems in healthcare professionals. *Evidence-Based Nursing*.
- Brooke, J. 1996. SUS—A quick and dirty usability scale. In P. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Harp, J. J. 2023. The shortage of healthcare workers in the United States: A call to action. In *Assessing the Need for a Comprehensive National Health System in the United States* (pp. 123-138). IGI Global Scientific Publishing.
- Hazell, C. M., Berry, C., Bogen-Johnston, L., & Banerjee, M. 2022. Creating a hierarchy of mental health stigma: testing the effect of psychiatric diagnosis on stigma. *BJPsych Open*, 8(5), e174. <https://doi.org/10.1192/bjo.2022.578>
- Hendl, T., & Shukla, A. 2024. Can digital health democratize health care? *Bioethics*, 38(6), 491-502. <https://doi.org/10.1111/bioe.13266>
- Lang, P. J. 1980. Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in Mental Health Care Delivery Systems* (pp. 119-137). Ablex Publishing.

- Lewis, J. R., Utesch, B. S., & Maher, D. E. 2020. UMUX-Lite: An accurate, validated, and reliable short-form usability questionnaire. *International Journal of Human-Computer Interaction*, 36(16), 1541-1550.
- Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., Rossman, J., & Park, S. 2000. Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. Presented at the Third International Workshop on Presence, Delft, The Netherlands.
- Mezuk, B., Needham, B., Joiner, K., Watkins, D., Stoddard, S., Burgard, S., & Link, B. 2021. What elephant? Pedagogical approaches to addressing stigma toward mental disorders in undergraduate public health education. *Pedagogy in Health Promotion*, 7(3), 183-190. <https://doi.org/10.1177/23733799209228>
- Palomin, A., Takishima-Lacasa, J., Selby-Nelson, E., & Mercado, A. 2023. Challenges and ethical implications in rural community mental health: the role of mental health providers. *Community mental health journal*, 59(8), 1442-1451.
- Rodríguez-Rivas, M. E., Cangas, A. J., Martín, A., Romo, J., Pérez, J. C., Valdebenito, S., ... & Corrigan, P. 2024. Reducing stigma toward people with serious mental illness through a virtual reality intervention: a randomized controlled trial. *Games for Health Journal*, 13(1), 57-64. <https://doi.org/10.1089/g4h.2023.0118>
- Rodríguez-Rivas, M. E., Cangas, A. J., Cariola, L. A., Varela, J. J., & Valdebenito, S. 2022. Innovative technology-based interventions to reduce stigma toward people with mental illness: systematic review and meta-analysis. *JMIR Serious Games*, 10(2), e35099. <https://doi.org/10.2196/35099>
- Schrepp, M., Hinderks, A., & Thomaschewski, J. 2017. Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). *International Journal of Human-Computer Interaction*, 33(10), 803-812. <https://doi.org/10.9781/ijimai.2017.09.001>

3D Selection through Hand Tracking in XR for Point Clouds Using Signed Distance Fields

Luca Fluri, and Arzu Çöltekin

*Institute of Interactive Technologies, School of Computer Science,
FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland*

Corresponding author: arzu.coltekin@fhnw.ch

Keywords: XR, Hand-tracking, Data Selection, Point Clouds, Signed Distance Fields

Introduction & Related Work

Visualizing scientific phenomena in XR can provide unique spatial insights (Fonnet & Prié, 2021). However, data selection in 3D space remains challenging. Traditional 3D selection methods are based on 2D interfaces and lack immersive feedback, requiring users to mentally translate between 2D projections and 3D relationships (Çöltekin et al., 2016). Various 3D selection methods do exist, such as the Slice-n-Swipe, i.e., a knife metaphor for slicing point clouds (Bacim et al., 2014), and Tangible Brush, a 6DOF input with touch gestures (Serenio et al., 2016). However, current 3D selection methods are typically limited to selecting simple, pre-defined geometric primitives such as spheres and boxes, which limits their use (Baloup et al., 2019). Furthermore, many of the current 3D selection methods require specialized hardware, they are not optimized for large datasets, leading to computational bottlenecks, thus do not sufficiently support real-time XR applications. Similarly, automated approaches to 3D selection based on machine learning needs extensive preprocessing to train the necessary models, making it high-barrier for real-time exploratory tasks (Chen et al., 2020).

Existing 3D selection methods in XR can be classified into two categories: *ray-casting* and *volumetric*. Early tools were restricted to single-point selection, while recent hand-tracking methods are much more flexible, even though there are still limitations due to tracking quality issues (Lubos et al., 2014). However, hand-tracking based selection is attractive as it aligns with natural human behaviors, potentially providing spatially intuitive data interaction (Napier, 2021). Despite the advantages of hand-tracking, advances in XR interaction and GPU processing, the combination of hand-tracking, volumetric selection, and real-time point cloud processing remains mostly unexplored. A specifically promising method that lends itself well in this junction is the *SDFs* (*Signed Distance Fields*), which, compared to traditional methods, allow for compact geometry encoding that supports smooth interpolation and anti-aliased rendering across collision detection, mesh generation, and 3D rendering applications (Chermain et al., 2021). Our work combines SDF based algorithmic accuracy for volumetric 3D data selection with arguably intuitive hand-based interaction. We introduce a new GPU-based SDF approach that connects best of traditional 2D selection methods with XR paradigms, while still offering real-time performance for large-scale point cloud datasets.

Our Approach

To address the limitations of existing 3D selection methods, we created a Unity-based XR application that combines hand tracking with 3D selection algorithms that are GPU accelerated and SDF based. We use SDFs to represent arbitrary 3D selection volumes as compact 3D textures, which are processed entirely on the GPU with compute shaders for parallel point-in-volume testing. The system keeps all point cloud data (positions, colors, selection states) in GPU memory to reduce CPU-GPU transfer bottlenecks, while utilizing Unity's indirect instanced rendering pipeline to visualize millions of points in real-time (Neuman-Donihue et al., 2023). Figure P 58 shows the overall architecture of our solution.

Hand-tracking acts as the main input method, enabling natural gesture interaction with volumetric selection tools (Buckingham, 2021). This GPU based setup supports efficient processing of large point clouds and offers flexibility to implement various selection techniques, from geometric primitives to complex brushing-based interactions, all operating at real-time frame rates suitable for immersive XR environments.

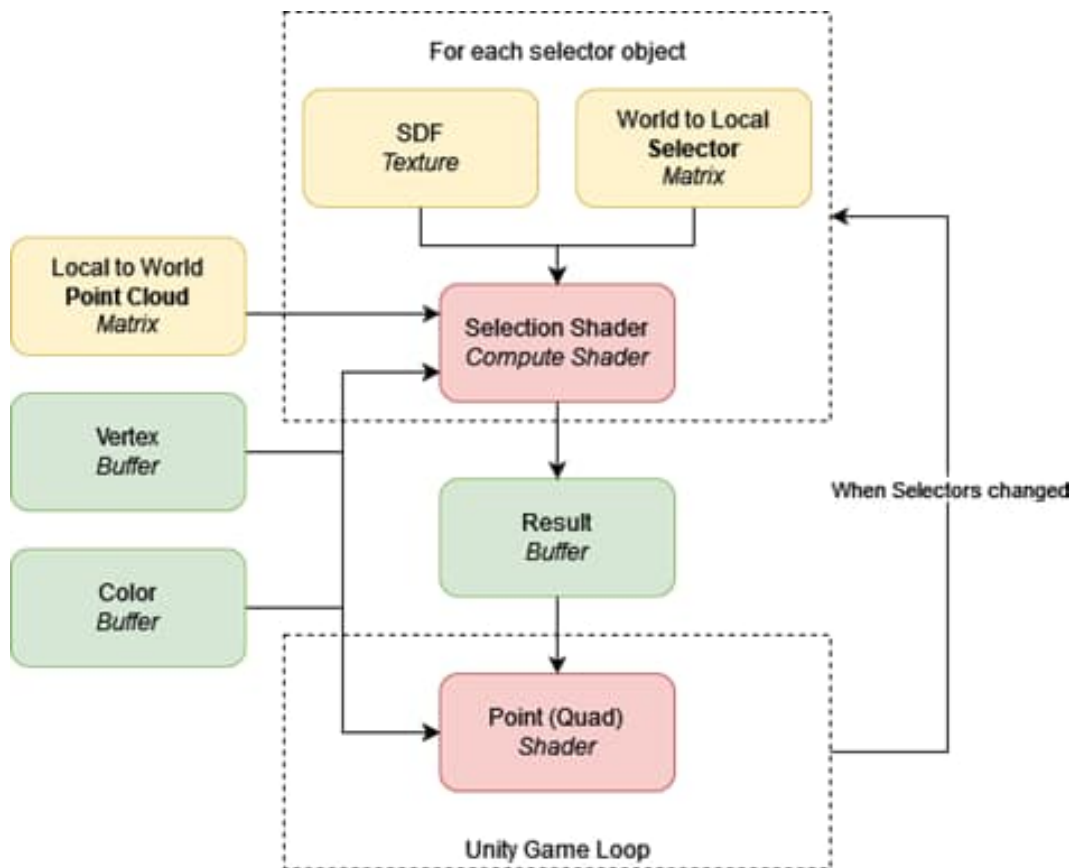


Figure P 58. Overview of our selection handling concept. Yellow represents transformation data, red represents shaders, and green represents GPU data buffers.

Selection Techniques

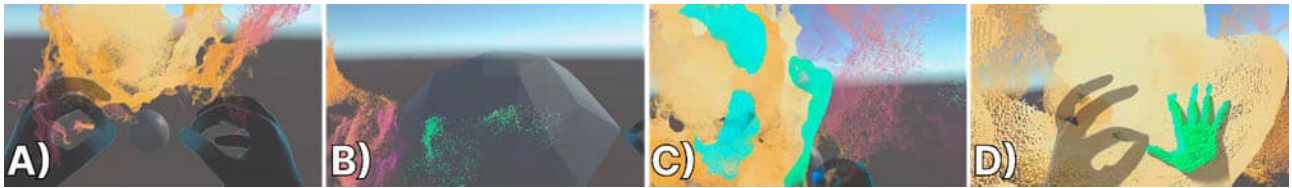


Figure P 59. Our four selection techniques. A) Pinch-Drag to spawn shapes, B) pinch-drag to draw convex hull, C) spherical brushing (brush sphere) and D) using hands model directly as brush (brush hands).

We implemented four distinct selection techniques to showcase the versatility of our SDF-based framework, categorized into two interaction paradigms: direct geometry-based *Shapes* and *Convex hull* (Figure P 59A and B) and brushing-based approaches *brush sphere* and *brush hands* (Figure P 59C and D). We elaborate each technique below:

- **Shapes:** Users create predefined geometric primitives (sphere, box, cylinder) using a "pinch and pull" gesture, where the distance between hands determines the scale during pinching, and the position is automatically centered between the hands (Benko & Feiner, 2007). Once created, shapes can be moved, rotated, and scaled through ongoing hand manipulation, enabling precise volumetric selection with familiar geometric forms.
- **Convex Hull.** A new mid-air drawing method allows users to sketch convex selection volumes with one-handed pinching gestures. Collected pinch points are smoothed using Laplacian filtering before creating a 3D convex hull mesh, which is then converted into an SDF volume. This approach bridges the gap between rigid geometric primitives and freeform selections, though it is limited to convex shapes.
- **Brush Sphere.** Implements a dynamic, sphere-based selection system where users can spawn and control a spherical brush volume. Users can sweep the sphere through the point cloud with real-time scaling and positioning, enabling intuitive "painting" of selections similar to traditional 2D brush tools but in 3D space.
- **Brush Hands.** A novel technique utilizing live hand mesh geometry as selection volumes. Real-time SDF creation from tracked hand shapes allows natural hand movements to directly select objects. To avoid accidental selections (the "Midas touch" problem) (Conte, 2020), the method requires activation by pinching with the other hand, giving users intentional control over when their hands act as selectors.

All implemented techniques are supported by a hand-attached palm menu for mode switching and parameter adjustments, accessible through poking interactions. A bimanual transformation system uses a "double-fist" gesture implementing the handlebar metaphor (Song et al., 2012), where the midpoint between fists controls translation, hand distance controls scaling, and hand orientation determines rotation of the entire point cloud and selection volumes.

Discussion & Conclusion

Based on the literature and initial testing of our prototypes, we believe SDF-based selection approach successfully addresses key limitations of traditional 3D selection techniques while maintaining real-time performance for practical XR applications. Importantly, GPU-accelerated design achieves linear scalability, enabling interactive selection of point clouds up to 5 million points at acceptable frame rates (>60 FPS), with rendering rather than selection computation being the main bottleneck. We

believe that the integration of real-time SDF generation with hand tracking marks a significant technical achievement, enabling the arbitrary selection of volumes that go beyond simple geometric primitives. Our work shows that efficient, real-time selection of nearly arbitrary point cloud subsets is possible through careful GPU optimization and SDF-based volume representation. At this stage, our prototype implementation only supports convex shapes. Furthermore, we observed performance drops beyond 10 million points and some variability in hand-tracking accuracy. Future improvements should focus on solutions such as hierarchical data structures (Schütz et al., 2020) to better handle datasets larger than 10 million points, support concave selections, and implementing AI-assisted methods to make hand-tracking more robust. Despite these limitations, our work demonstrates that intuitive, accurate 3D selection is possible in XR environments. Specifically, combining SDF processing with hand-tracking enables efficient spatial data exploration, and we believe it offers significant potential for wider adoption of XR-based analysis tools in scientific and industrial fields.

References

- Bacim, F., Nabyouni, M. & Bowman, D.A. 2014. Slice-n-Swipe: A freehand gesture user interface for 3D point cloud annotation. In *2014 IEEE Symposium on 3D User Interfaces 3DUI 2014*, 185–186.
- Baloup, M., Pietrzak, T. & Casiez, G. 2019. RayCursor: A 3D Pointing Facilitation Technique based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.
- Benko, H. & Feiner, S. 2007. Balloon Selection: A Multi-Finger Technique for Accurate Low-Fatigue 3D Selection. In *2007 IEEE Symposium on 3D User Interfaces 3DUI 2007*, 79–86.
- Buckingham, G. 2021. Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. *Frontiers in Virtual Reality* 2. <https://doi.org/10.3389/frvir.2021.728461>
- Chen, Z., Zeng, W., Yang, Z., Yu, L., Fu, C.-W. & Qu, H. 2020. LassoNet: Deep Lasso-Selection of 3D Point Clouds. *IEEE Transactions on Visualization and Computer Graphics* 26(1), 195–204.
- Chermain, X., Lucas, S., Sauvage, B., Dischler, J.-M. & Dachsbacher, C. 2021. Real-Time Geometric Glint Anti-Aliasing with Normal Map Filtering. *Proc. ACM Comput. Graph. Interact. Tech.* 4(1), 1:1–1:16.
- Çöltekin, A., Lokka, I. & Zahner, M. 2016. On the usability and usefulness of 3d (geo) visualizations—a focus on virtual reality environments. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41, 387–392. <https://doi.org/10.5167/uzh-133353>
- Conte, T. 2020. The Midas Touch effect: the most unknown phenomenon in UX design. Online blog, <https://uxdesign.cc/the-midas-touch-effect-the-most-unknown-phenomenon-in-ux-design-36827204edd>. Last visited 20th of July 2025.
- Fonnet, A. & Prié, Y. 2021. Survey of Immersive Analytics. *IEEE Transactions on Visualization and Computer Graphics* 27(3), 2101–2122.
- Lubos, P., Beimler, R., Lammers, M. & Steinicke, F. 2014. Touching the Cloud: Bimanual annotation of immersive point clouds. In *2014 IEEE Symposium on 3D User Interfaces 3DUI 2014*, 191–192.
- Napier, J. 2021. *Hands*. Princeton: Princeton University Press.
- Neuman-Donihue, E., Jarvis, M. & Zhu, Y. 2023. FastPoints: A State-of-the-Art Point Cloud Renderer for Unity. arXiv preprint arXiv:2302.05002. <https://doi.org/10.48550/arXiv.2302.05002>
- Schütz, M., Ohrhallinger, S. & Wimmer, M. 2020. Fast Out-of-Core Octree Generation for Massive Point Clouds. *Computer Graphics Forum* 39(7), 1–13. <https://doi.org/10.1111/cgf.14134>
- Sereno, M., Ammi, M., Isenberg, T. & Besançon, L. 2016. Tangible Brush: Performing 3D Selection with Portable and Position-aware Devices. In *IEEE VIS2016*.
- Song, P., Goh, W.B., Hutama, W., Fu, C.-W. & Liu, X. 2012. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 1297–1306.

Toward Efficient Storage and Real-time Rendering of Dense 3D Gaussian Splatting Reconstruction in Virtual Reality

Sergio Montoya, Albert Ruiz, Marc Martos, Ruggero Ragonesi, Leonel Toledo,
Ivan Huerta, Diego González, and Sergi Fernández

Fundació i2CAT, Spain

Corresponding author: diego.gonzalez@i2cat.net

Keywords: 3D Gaussian Splatting, Neural Rendering, Volumetric Compression, Unity, WebGPU, Virtual Reality

Introduction

Recent advances in neural rendering have led to powerful new representations for photo-realistic scene reconstruction, with 3D Gaussian Splatting (3DGS) emerging as a state-of-the-art technique due to its ability to render high-quality results in real time (Kerbl et al., 2023). Unlike traditional mesh-based approaches, 3DGS models scenes as collections of anisotropic Gaussians with learned properties such as position, scale, color, and view-dependent appearance, allowing for continuous and differentiable volumetric rendering (Kerbl et al., 2023; Mildenhall et al., 2020). Despite its strengths, 3DGS poses challenges in terms of storage, transmission, and rendering performance, especially when targeting interactive or resource-constrained environments like virtual reality (VR). To address these limitations, recent efforts have focused on representation compression (Bagdasarian et al., 2025), SH distillation (Hedman et al. 2021) and platform-specific rendering optimizations (Aras, 2023), which this work builds upon and extends. Finally, we present an optimized 3D Gaussian Splatting rendering for XR devices which relies on the data reduction pipeline that we describe in this paper.

3D Gaussian Splatting Pruning, Distillation, and Quantization

Recent advances in point-based neural rendering have significantly improved the quality of 3D scene reconstruction, with 3D Gaussian Splatting (3DGS) emerging as a particularly effective approach. However, the scalability of these techniques is limited by the exponential increase in storage requirements, often reaching several gigabytes per scene. This storage burden, alongside the real-time computational demands of rendering large numbers of Gaussians, limits their deployment in interactive and mobile virtual reality (VR) applications. Therefore, it becomes essential to develop efficient compression strategies that maintain high visual fidelity while reducing both storage footprint and rendering complexity.

To address these challenges, we propose a multi-stage pipeline for storage-efficient and real-time-friendly 3D Gaussian Splatting representations, built upon the LightGaussian (Zhiwen, et al., 2023) framework: (i) Pruning: we apply visibility-, opacity-, and redundancy-based pruning techniques to

reduce the number of Gaussians while preserving scene geometry and appearance. (ii) Spherical Harmonics (SH) Distillation: we simplify the modeling of view-dependent effects by reducing SH degree, trained through a teacher-student distillation pipeline. (iii) Vector Quantization (VQ): we introduce a K-means-based quantization method targeting only the least significant SH coefficients, enabling a compact encoding without affecting critical visual features. This pipeline is designed to strike a balance between compression and fidelity, and to make 3DGS feasible for deployment in bandwidth-constrained or latency-sensitive VR systems. Our main contribution is not only the adaptation of LightGaussian pipeline to our particular use case, but also quantitative evaluating the impact each of these steps have on the quality, storage requirements and rendering performance. To test our implementation, we use two public datasets: (i) MipNeRF360 (Jonathan, 2022) & (ii) Tanks and Temples (T&T) (Knapitsch et al, 2017). Besides, we created our own dataset by capturing indoor scenes from a complex museum environment using an Insta360 X4 camera.

We conduct a comprehensive evaluation by applying our pipeline on all three datasets. The pruning step alone yields substantial data size reduction while preserving perceptual quality. Further application of SH distillation—by lowering the SH degree from 2 to 0—enables up to a 10× compression ratio, although with increasingly noticeable degradation in view-dependent lighting fidelity. In the final stage, vector quantization is applied selectively to the 60% least relevant SH coefficients. This step introduces an additional 2× compression with minimal loss in rendering quality. Across all datasets, we observe that our complete pipeline—pruning + SH distillation + VQ—achieves compression rates of up to 25× compared to the original uncompressed 3DGS representation, while showing a maximum PSNR loss of around 1.5 dB in the most compressed scenario. It is key to note that the quality loss is only perceivable if areas with high reflections due to Spherical Harmonics distillation. A summary of the obtained results are shown in Table 1.

Table P 16. Summary of Pruning, Spherical Harmonics distillation and Vector Quantization quantitative results.

Dataset	Reduction Steps	CR	Avg. PSNR (dB)	PSNR Loss (%)
MipNeRF360	Vanilla GS	N/A	31.17	N/A
	Pruning	~3×	31.19	-0.06%
	Pruning + SH (deg 2)	~6×	30.74	1.38%
	Pruning + SH (deg 1)	~8×	30.18	3.18%
	Pruning + SH (deg 0)	~10×	29.86	4.20%
	Pruning + SH (deg 0) + VQ	~20–25×	29.83	4.30%
T&T	Vanilla GS	N/A	20.63	0.00%
	Pruning	~3×	20.82	-0.92%
	Pruning + SH (deg 0)	~10×	20.14	2.38%
	Pruning + SH (deg 0) + VQ	~20–25×	20.12	2.47%
VNG_Museum	Vanilla GS	N/A	26.66	N/A
	Pruning	~3×	26.74	-0.30%
	Pruning + SH (deg 0)	~10×	25.72	3.53%
	Pruning + SH (deg 0) + VQ	~20–25×	25.70	3.60%

3D Gaussian Splatting Encoding and Compression

To extend the 3D Gaussian Splatting size reduction described in the section above, the present work proposes a novel encoding and compression pipeline that transforms raw 3D Gaussian parameters into

a structured RGB image representation which can later be compressed using traditional video or image compression tools. The method is built upon grouping Gaussian attributes into four distinct sets: position (x, y, z), color and opacity, scale, and rotation (expressed as quaternions). Each group is converted into RGB images where individual channels are assigned to specific components (e.g., x, y, z for position; r, g, b for color; etc.). These encoded textures are then compressed using LZ4 lossless compression standard.

Although position data is crucial for preserving scene geometry, our preliminary experiments revealed that this attribute is highly sensitive to quantization and cannot currently be encoded effectively through 2D texture-based representations without significant degradation, forcing position data to remain in raw format. On the other hand, the other parameters can successfully be encoded in 2D textures while showing a final quality loss after decoding below 0.4 dB. Even without transforming the positions data, our encoding and compression pipeline still achieves slight reductions in the final storage footprint of 3D Gaussian Splatting scenes (an average of 25% smaller data sizes), particularly due to the efficient compression of the remaining non-spatial attributes. Regarding performance, we measured total encoding and decoding times below 40 ms when running in a Nvidia RTX 4090 GPU. The overall results demonstrate that significant compression is possible without compromising rendering quality, with the added benefit of a modular design that enables straightforward hardware integration. While the current pipeline does not yet compress spatial information effectively, ongoing work aims to explore alternative strategies—such as adaptive encoding schemes or hybrid data representations—that may enable efficient and accurate compression of geometric data in future iterations. The proposed method is a promising step toward scalable, real-time deployment of high-fidelity 3D reconstructions in constrained environments.

VR Rendering: Real-Time Gaussian Splatting for Web and Native XR Platforms

This work presents a real-time Gaussian Splatting renderer specifically designed for Unity, with the goal of supporting both web-based and native virtual reality (VR) applications. The renderer was developed as an extension of the open-source implementation provided by aras-p. We significantly adapted this library to enable efficient volumetric rendering across a range of devices and platforms, including Android-based headsets and in-browser XR environments. The main goal is to unify and streamline cross-platform rendering workflows while maintaining performance and visual fidelity in real-time scenarios.

For browser-based rendering, a WebGPU-compatible backend was developed and integrated into Unity's rendering pipeline. WebGPU introduces a more restrictive compute model requiring architectural changes to the original renderer. To address these constraints, a FidelityFX-compatible sorting algorithm was integrated to replace less portable compute routines. Radix sorting was retained for native desktop and mobile platforms, preserving optimal performance in those contexts. Additional changes include duplicating edit buffers to circumvent WebGPU access restrictions and adding runtime detection of execution context, enabling dynamic switching between WebGPU and native logic. These adaptations allowed medium-complexity scenes to be rendered in real-time (>60 FPS) in-browser, demonstrating the feasibility of XR volumetric rendering via Unity and WebGPU.

In parallel, native standalone VR support was added for both Windows and Android platforms. The renderer was modified to support Unity's XR architecture and interface correctly with third-party XR packages and runtime environments. Key features include support for both single-pass instanced and multiview rendering modes, correct handling of per-eye projection matrices in shaders, and use of 2D texture arrays to manage stereo rendering within a unified compute pipeline. Compatibility with Unity's

XR framework ensures accurate headset tracking and stereo output, while preserving the flexibility to operate in multipass and non-VR modes. Crucially, the integration of pruning and distillation techniques—applied during preprocessing—was essential to reduce the number of splats and their associated computational cost without sacrificing rendering quality. These optimization steps significantly decreased the overall memory bandwidth and fragment overdraw, directly enabling higher framerates in resource-constrained environments.

Prototype tests conducted on an untethered Oculus Quest 3 using a custom Unity APK demonstrated stable real-time rendering above 100 FPS when operating in PCVR mode. However, when running fully standalone on the device, performance remains limited to approximately 20 FPS due to increased per-eye computation and hardware constraints. Further optimization, including frustum and occlusion culling and additional splat-level simplification, is planned to close this performance gap and enable fluid standalone VR rendering.

Conclusion

This work presents a comprehensive pipeline for enabling real-time, cross-platform deployment of 3D Gaussian Splatting in virtual and augmented reality applications. By combining pruning, SH distillation, and vector quantization, we achieved up to 25× compression with minimal quality degradation (around 1.5 dB of maximum PSNR loss), making high-fidelity volumetric rendering feasible on memory- and bandwidth-constrained devices. Our additional contributions include a modular compression and encoding scheme that transforms Gaussian attributes into RGB representations, enabling further storage reduction using standard compression tools, and a real-time Unity-based renderer that supports both WebGPU for in-browser XR and native VR rendering on Android and Windows.

Prototype tests demonstrate real-time performance in PCVR and browser-based settings, and while standalone VR still faces hardware limitations, the integration of aggressive preprocessing techniques has proven crucial for maintaining visual fidelity under tight resource constraints. Our data reduction pipeline and lightweight rendering logic establishes a strong foundation for scalable deployment of neural representations across a wide range of devices. Future work will focus on improving spatial attribute compression, introducing advanced culling strategies, and refining real-time optimizations to further close the gap for standalone VR applications.

Acknowledgement

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement N° 101070192 (CORTEX2)

References

- Aras-P, 2023. Unity Gaussian Splatting Implementation, GitHub
- Bagdasarian, M. T., Knoll, P., Li, Y., Barthel, F., Hilsmann, A., Eisert, P., & Morgenstern, W. 2024. 3dgs. zip: A survey on 3d gaussian splatting compression methods. In Computer Graphics Forum (p. e70078). <https://doi.org/10.1111/cgf.70078>
- Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., & Hedman, P. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5470-5479).
- Fan, Z., Wang, K., Wen, K., Zhu, Z., Xu, D., & Wang, Z. 2024. Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. Advances in neural information processing systems, 37, 140138-140158.
- Hedman, P., Srinivasan, P. P., Mildenhall, B., Barron, J. T., & Debevec, P. 2021. Baking neural radiance fields for real-time view synthesis. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 5875-5884).
- Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph., 42(4), 139-1.
- Knapitsch, A., Park, J., Zhou, Q. Y., & Koltun, V. 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. ACM Transactions on Graphics (ToG), 36(4), 1-13. <https://doi.org/10.1145/3072959.3073599>
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106. <https://doi.org/10.1145/3503250>

AI-Powered Scene Generation in VR

Zobaer Ahmed, Tatiana Gerth, and Helmut Grabner

School of Engineering, Zurich University of Applied Sciences, Switzerland

Corresponding author: ahmd@zhaw.ch

Keywords: Virtual Reality, Generative AI, 3D Scene Creation, Rapid Prototyping, Interactive Games

Introduction and Motivation

The creation of 3D scenes and interactive content has traditionally been limited to individuals with expertise in modelling software and programming. This limitation reduces the speed of prototyping and restricts participation in creative processes to a narrow group of skilled users. To address this issue, a virtual reality (VR) tool has been developed to facilitate intuitive scene and game creation through natural interaction. The tool enables users to generate, edit, and animate 3D environments without requiring programming knowledge, aiming to broaden accessibility and streamline content development across domains such as education, visualization, and entertainment.

Tool Panels Overview

The system integrates several AI technologies into a modular VR interface. Inference for all AI components (speech recognition, large language model (LLM) processing, and image generation) is performed via Hugging Face's hosted APIs. Whisper (OpenAI 2023) handles speech-to-text, LLaMA 3.1 Instruct (Meta AI 2024) is used to generate descriptive prompts for object and texture generation, Meshy API (Meshy AI 2024) provides 3D model generation, and Stable Diffusion 3.5 (Stability AI 2024) is used for 2D image and texture creation. The main interface consists of three floating panels arranged around the user (Figure P 60):

1. Main Tool Panel

This panel includes three essential tools:

- **3D Object Generation Panel:** Users can enter text prompts, and these are sent to the Meshy API to retrieve realistic, textured 3D models generated from scratch based on the description.
- **Image Generation Panel:** Users can input descriptive prompts to generate images and use them as textures using Stable Diffusion (Stability AI 2024).
- **Log Console:** This displays generation status updates, warnings, and internal debug information, helping users understand system behaviour in real time.

2. Object Editing Panel

This contextual panel appears when the user points at an object and presses the right-hand secondary button. It allows detailed editing of object properties, including:

- **Physics Settings:** Adjust scale, mass, toggle gravity, and apply predefined physics materials (e.g., slippery, bouncy).
- **Movement Composition:** Users can combine simple motion primitives (linear, circular, elliptical and oscillatory) to create more complex dynamic behaviours such as vibrations or orbital paths.
- **Texture Management:** The current texture can be previewed, and a new one can be generated and applied using Stable Diffusion directly from this panel.

3. Inventory Panel

This panel provides persistent storage for generated content:

- **3D Object Storage:** Generated objects can be dragged into inventory slots for later use and dragged back into the scene to reinstantiate them.
- **Texture Storage and Application:** Generated textures are automatically stored. Users can grab a texture to attach it to their virtual hand. When the hand-held texture comes into contact with an object, it is automatically applied as the surface material.

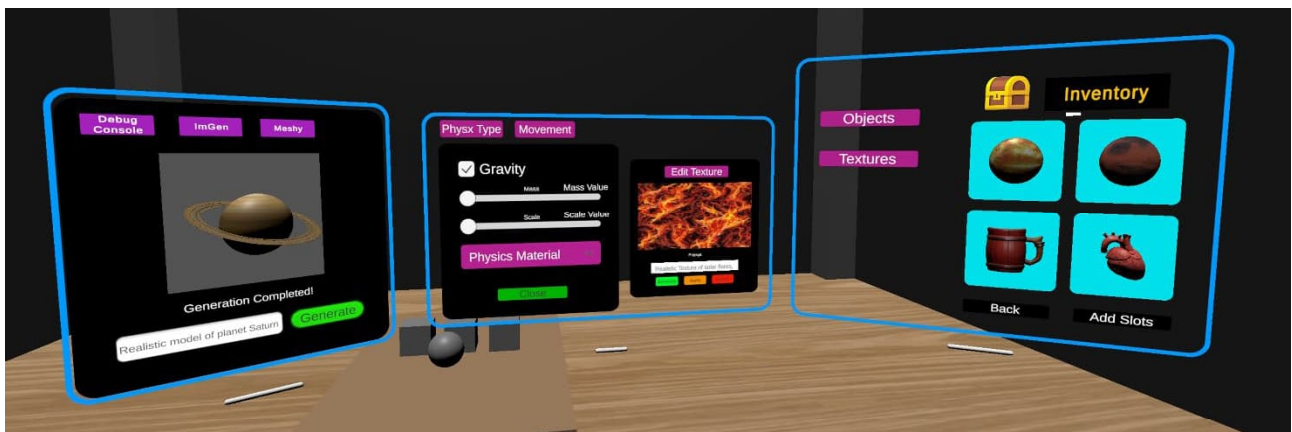


Figure P 60. User Interface: Main tool Panel on the left, Object Editing Panel in the middle and Inventory Panel on the right.

Work in Progress: Scene Generation from Description

A core feature currently under development is the scene generation panel, where users can either type or speak a complete scene description. The spoken input is transcribed using Whisper (OpenAI 2023), and passed to a LLaMA Instruct LLM (Meta AI 2024) via Hugging Face's inference API. The model then generates a structured JSON output that contains a list of objects and corresponding prompts for 3D model generation and texturing. These prompts are sent to the Meshy API and used to instantiate the described scene in VR without requiring manual asset selection. This creates a truly hands-free, voice-driven 3D prototyping environment.

Showcase Applications

Three demonstration scenes have been developed using the system to illustrate its capabilities:

1. **Ball Toss Game:** A simple game where a user tosses a ball using a virtual tennis racket. Each time the ball hits the ground, the score resets to zero.
2. **Educational Object Labelling:** A generated 3D object (e.g., a heart, engine, or plant) is labelled interactively. The user can point to parts of the object and attach descriptive text or images, allowing for effective visualization and learning.
3. **Planetary Orbit Simulation:** A scene where multiple planets are generated and animated using circular motion patterns to simulate orbits. This demonstrates the system's ability to compose dynamic, physics-driven scenes with minimal input. Note that the orbital motion is not physically accurate and does not follow real orbital mechanics; it serves purely as a demonstration of the movement composition functionality available in the tool.

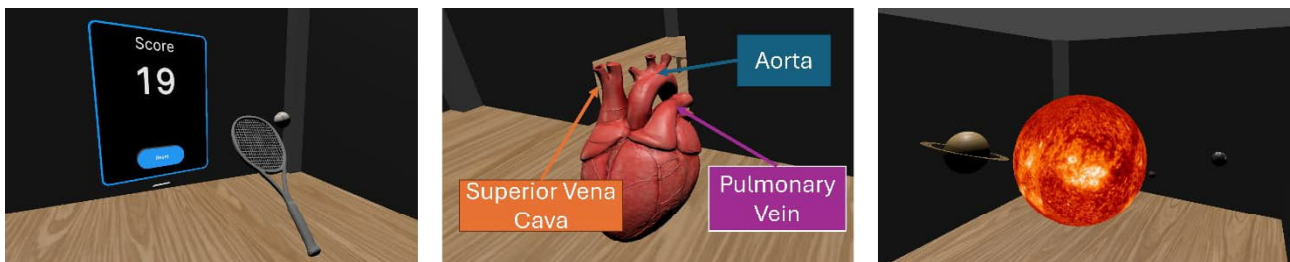


Figure P 61. Showcase Applications: Ball Toss on the left, Educational Object Labelling in the middle and Planetary Orbit Simulation on the right.

Future Work and Conclusions

Future development will focus on incorporating visual scripting tools for defining game logic—such as win/loss conditions, score tracking, and object behaviours—directly within VR environment. This will allow users to build complete mini-games without writing code. By combining speech input, generative AI, and immersive VR interaction, this project offers a highly accessible and creative tool for building interactive 3D content. It lowers the entry barrier for designers, educators, and hobbyists alike.

References

- OpenAI. 2023. *Whisper large-v3*. Available at: <https://huggingface.co/openai/whisper-large-v3>
- Meta AI. 2024. *Llama-3.1-8B-Instruct*. Available at: <https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct>
- Meshy AI. 2024. Meshy: AI-Driven 3D Model Generation API. [Online]. Available: <https://meshy.ai/api>
- Stability AI. 2024. *Stable Diffusion 3.5 Large*. Available at: <https://huggingface.co/stabilityai/stable-diffusion-3.5-large>

Genba SaVR: A Secure and Scalable Virtual Reality Platform for Industrial Training

Gabriel Danciu, Rebecca Acheampong, Alexandre Rekeraho, Ionut Alexandru Oprea, Vladut-Gabriel Anghel, Bogdan-Valentin Floricescu, and Ioana Corina Bogdan

Dept. of Electronics and Computers, Transilvania University of Brasov, Romania

Corresponding author: gabriel.danciu@unitbv.ro, ionut-alexandru.oprea@student.unitbv.ro

Keywords: Virtual Reality, Industrial Training, Microservices Architecture, Adaptive Learning, MageAI, Real-time Collaboration, XR Security

Abstract

Genba SaVR is a scalable, secure microservices platform designed to deliver immersive industrial training in virtual reality environments. Unlike traditional monolithic XR platforms, Genba SaVR integrates containerized microservices orchestrated by Kubernetes to enable dynamic scaling, high availability, and modular deployment. MageAI pipelines reduce task complexity by providing contextual hints and dynamically adjusting training materials to guide users through each scenario. Multi-layered security mechanisms, combining authentication, session management, and continuous DevSecOps practices, provide robust protection of user data and platform infrastructure. Performance evaluations demonstrate significant latency reductions, with WebRTC protocols achieving an average communication latency of approximately 17 ms. Usability assessments confirm high levels of engagement and responsiveness, supporting effective collaborative training scenarios across diverse operational contexts.

Introduction

Immersive technologies are transforming industrial training by enabling realistic, interactive experiences that improve safety, engagement, and knowledge retention (Naranjo et al., 2020). However, delivering effective virtual reality (VR) training at scale remains challenging due to latency constraints, security vulnerabilities, and the complexity of maintaining high-performance infrastructure (Van Damme et al., 2024; Acheampong, R., 2024). Traditional monolithic platforms often struggle to balance responsiveness, scalability, and robust security, resulting in limited adoption for critical applications such as safety training and operational audits.

To address these limitations, Genba SaVR has been developed as a modular and secure XR platform tailored to industrial training environments. The platform combines containerized microservices architecture with adaptive learning pipelines and real-time communication protocols to deliver low-latency, high-reliability experiences. By integrating advanced security mechanisms and DevSecOps

practices, Genba SaVR ensures that sensitive training data and user interactions are protected without compromising usability or performance. The system is designed to support dynamic scaling across diverse deployment scenarios, enabling organizations to train and collaborate securely from any location.

Methodology

Architectural Design

Genba SaVR is a distributed microservices platform that separates core functionalities into Docker containers managed by Kubernetes for scalable, fault-tolerant deployment. As shown in Figure P 62 Nakama handles authentication, user management, and session persistence using credential verification and token management. Adaptive learning pipelines powered by MageAI collect and process user input to deliver personalized training content in real time. WebSocket and WebRTC protocols enable low-latency communication for interactive collaboration. Security monitoring is integrated through Kubescape and ARMO, supporting continuous vulnerability assessment and compliance with DevSecOps practices.

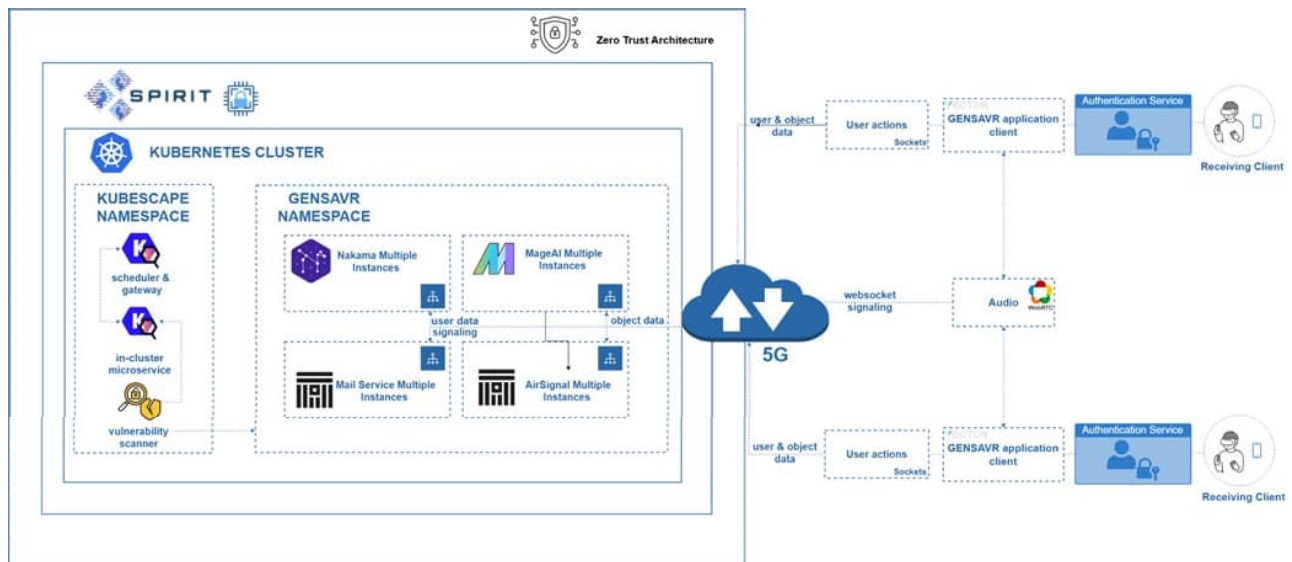


Figure P 62 Genba SaVR system architecture showing containerized microservices, Kubernetes orchestration, Nakama backend services, adaptive learning pipelines, and integrated security monitoring.

Adaptive Learning Integration

To enhance training effectiveness, Genba SaVR uses MageAI pipelines to collect user input and dynamically adapt learning content based on performance and engagement. As shown in Figure P 63, real-time data from VR interactions is processed through configurable pipelines that generate personalized feedback integrated directly into the training scenario.

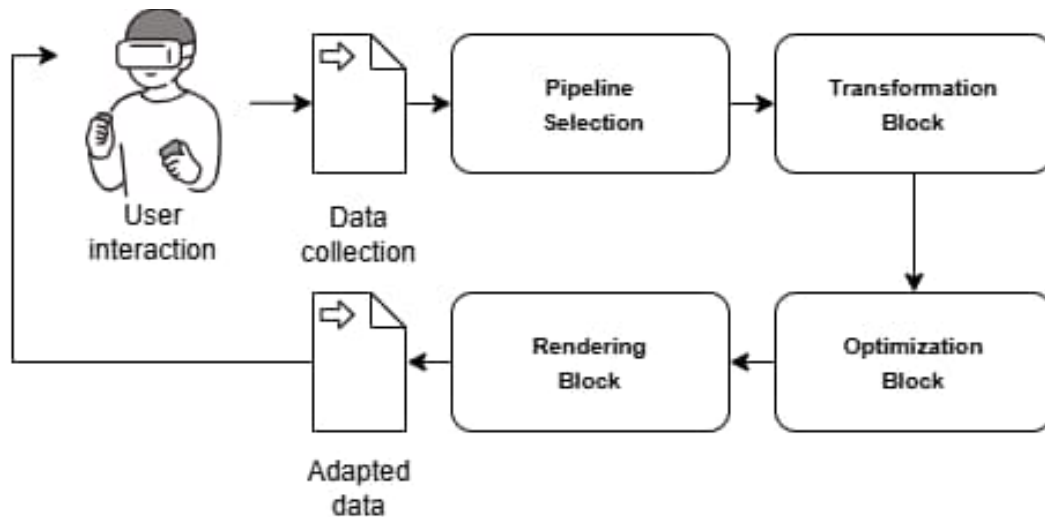


Figure P 63. Adaptive learning in Genba SaVR using MageAI for real-time content personalization.

Security Implementation

Security is integrated across all layers of Genba SaVR through multi-layered authentication combining credential verification, device validation, and short-lived tokens. Nakama manages secure password storage and session continuity. DevSecOps practices with Kubescape (Kamieniarz & Mazurczyk, 2024) and ARMO enable continuous vulnerability scanning and compliance checks, maintaining a strong security posture without compromising performance.

Evaluation and Results

The evaluation of Genba SaVR focuses on assessing performance, scalability, and security in a controlled deployment environment. Testing is conducted on a Kubernetes cluster configured to host the containerized microservices, including Nakama, MageAI pipelines, and real-time communication components.

Latency measurements indicate that the combination of WebSocket and WebRTC protocols achieves an average end-to-end communication latency of approximately 17 ms in optimal network conditions, comparable to established commercial collaboration platforms. Compared to initial configurations that relied exclusively on legacy signaling channels, this approach reduces latency by more than 60%, supporting smooth real-time interactions during training sessions. Figure P 64 illustrates the observed latency reduction achieved when deploying Genba SaVR on the SPIRIT platform infrastructure. Additional performance assessments demonstrate that Kubernetes orchestration maintains consistent resource allocation as concurrent user sessions increase, with automated scaling and load balancing sustaining stable response times and throughput.

Security validation is performed using Kubescape and the ARMO platform to identify potential vulnerabilities and verify compliance with recommended security practices. The assessment highlights a high level of adherence to Kubernetes security guidelines and compliance across the system. Identified risks primarily relate to medium-priority configuration issues, which are addressed through iterative updates and remediation procedures integrated into the platform's DevSecOps workflow.

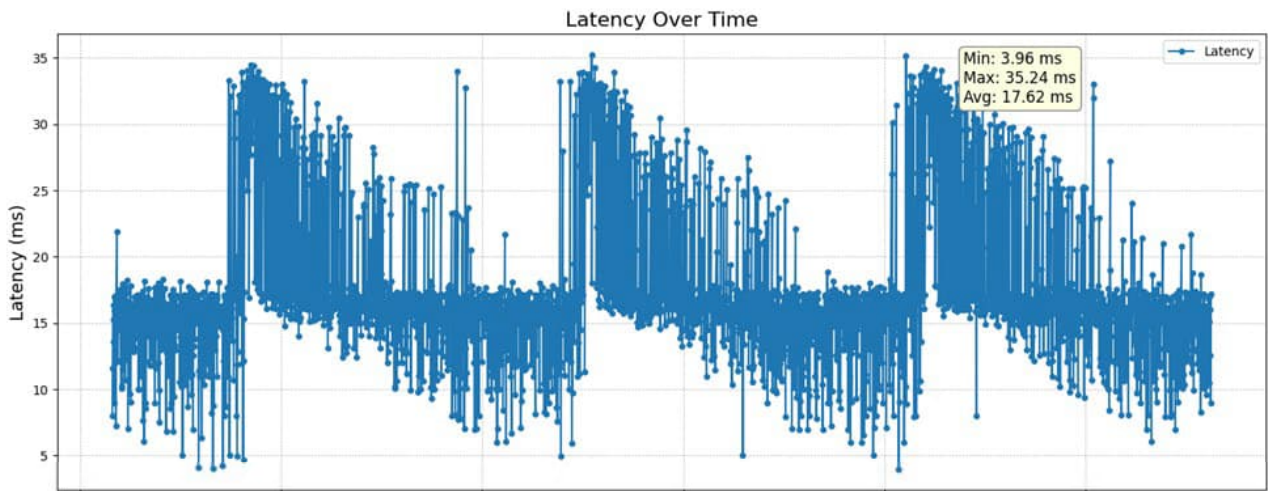


Figure P 64. WebRTC latency measurements with Genba SaVR deployed on the SPIRIT platform infrastructure.

Preliminary usability testing is conducted to evaluate the user experience within the VR training environment. Participants report a high degree of immersion and responsiveness, confirming that adaptive learning features and low-latency communication contribute to effective training engagement. Figure P 65 illustrates a representative training scenario where multiple participants collaborate in the virtual environment to perform shared evaluation of station-specific components during a line inspection exercise. Feedback indicates that the modular design simplifies user onboarding and supports consistent performance across different hardware configurations

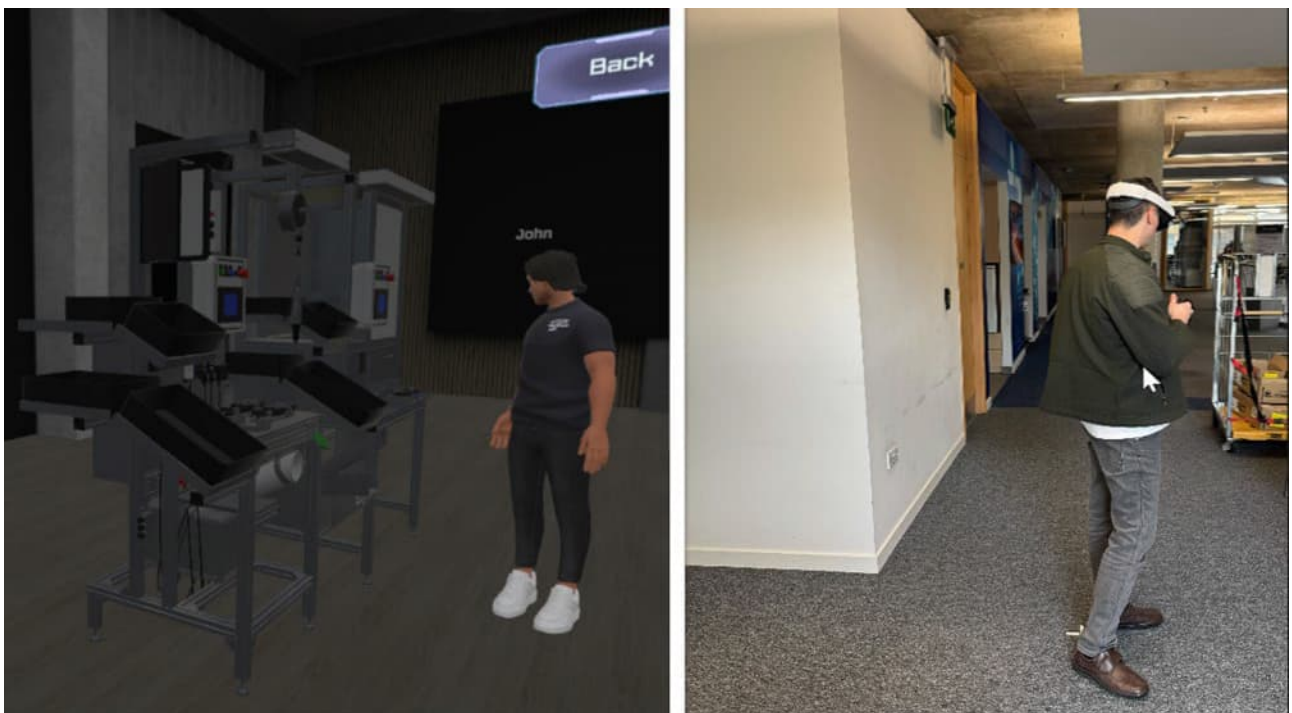


Figure P 65. Multi-user collaboration during line inspection, enabling shared evaluation of station-specific components in the virtual environment.

Conclusion

This work presents Genba SaVR, a modular and secure platform designed to support immersive industrial training with low-latency real-time collaboration. The system integrates containerized microservices, adaptive learning pipelines, and robust security practices into a unified infrastructure capable of dynamic scaling and high availability. Performance evaluations demonstrate that the combination of WebRTC and Kubernetes orchestration significantly reduces communication latency and maintains consistent responsiveness under variable workloads. Security assessments confirm a high level of compliance with Kubernetes best practices, supported by continuous vulnerability monitoring and DevSecOps workflows. Future development focuses on extending adaptive learning capabilities to incorporate additional behavioral analytics and personalized training pathways. Planned enhancements also include expanded testing across diverse industrial scenarios and integration of extended reality (XR) modules to validate effectiveness in production environments. These efforts aim to further improve training outcomes and ensure that the platform remains adaptable to evolving operational requirements and security challenges.

References

- Naranjo, J. E., Sanchez, D. G., Robalino-Lopez, A., Robalino-Lopez, P., Alarcon-Ortiz, A., & Garcia, M. V. 2020. A scoping review on virtual reality-based industrial training. *Applied Sciences*, 10(22), 8224. <https://doi.org/10.3390/app10228224>
- Van Damme, S., Sameri, J., Schwarzmann, S., Wei, Q., Trivisonno, R., De Turck, F., & Torres Vega, M. 2024. Impact of latency on QoE, performance, and collaboration in interactive Multi-User virtual reality. *Applied Sciences*, 14(6), 2290. <https://doi.org/10.3390/app14062290>
- Acheampong, R., Balan, T. C., Popovici, D. M., Tuyishime, E., Rekeraho, A., & Voinea, G. D. 2025. Balancing usability, user experience, security and privacy in XR systems: A multidimensional approach. *International Journal of Information Security*, 24(3), 1-18.
- Kamieniarz, K., & Mazurczyk, W. (2024, May). A comparative study on the security of kubernetes deployments. In 2024 International Wireless Communications and Mobile Computing (IWCMC) (pp. 0718-0723). IEEE.

Advancing Collaboration Paradigms in Extended Reality

Cédric Merz, Andreas Leu, Elif Gürçınar, and Arzu Çöltekin

*Institute of Interactive Technologies, School of Computer Science,
FHNW University of Applied Science and Arts Northwestern Switzerland, Switzerland*

Corresponding author: arzu.coltekin@fhnw.ch

Keywords: XR, Collaborative Systems, Awareness Cues, Interaction Paradigms

Introduction

Extended Reality (XR)—i.e., Virtual, Mixed and Augmented Reality (VR/MR/AR)— applications have been (re)emerging as promising environments for (co-located or remote) spatial collaboration in recent years (Çöltekin et al., 2020). If well-done, various forms of XR can be of great benefit for tasks that require visuospatial interaction, which can be challenging with traditional digital methods e.g., audio or video calls (Brägger et al., 2022; van Gent et al., 2024; Macchi & De Pisapia., 2024). Because such spatial interactions require extensive verbal exchange to describe spatial relationships, people might not be very effective or efficient with video calls in comparison to XR enabled environments (Baumgartner et al., 2022). For example, in scenarios where joint design decisions need to be made such as in architectural or urban planning contexts (Proulx & Çöltekin 2025); where a shared understanding needs to be reached such as in remote teaching and learning (Crogman et al., 2025); or an expert needs to guide someone in the field about operating a complex procedure in medical or industrial contexts (e.g., Baumgartner et al., 2022); the need for spatial interactions become very pronounced, and potentially critical (e.g., Elmasllari et al., 2025; Fedosov et al., 2025). However, despite its great potential in facilitating spatial interactions, collaboration in XR still faces several challenges. These challenges include both technological and human-computer interaction (HCI) issues, as well as psychological and social factors such as differences in visuospatial skills among individuals and groups, which can be influenced by age, expertise, anxiety levels, and spatial abilities (Çöltekin et al., 2020; Crogman et al., 2025). Tackling the above-mentioned challenges requires refining existing collaboration paradigms, and proposing, implementing, and testing new ones.

Our contributions

We present several prototypes that are based on a well cited concept known as "awareness cues", i.e., signals/symbols/signs in the user interface that is triggered by a remote partner which allows users to infer the remote partner's actions (Oulasvirta, 2009). Based on literature search, various interviews with stakeholders, and internal brainstorming sessions, we identified various visual and spatial ways one can implement awareness cues besides enabling voice or text to verbally communicate. Below we summarize the various experimental prototypes we developed and implemented based on the identified directions:

- **Gaze visualization:** Collaborators can see the gaze direction of each other on-demand, by which a stronger sense of shared attention can be created (Brägger et al., 2022, Figure P 66, left).
- **Pointing paradigms:** When a user wishes to bring the attention of their partner(s) to certain objects or areas, they can either simply look at them and activate gaze-pointing or activate hand-pointing by "finger extension" (Figure P 66, middle) laser or stick pointers (Figure P 66, right).
- **Shared views:** A synchronized visual space where all participants can view the same digital content either through a "picture-in-picture" screen e.g., from a moderator's or a selected partner's perspective (Figure P 66, red circles), or by rotating the scene, i.e., "switching seats".



Figure P 66. Three pointer styles (green ball indicates the position of the "pointer"). Left: Gaze pointer visualizes where the gaze hits the scene. Middle: Finger pointer extends the finger with a semi-transparent line, enabling a natural pointing behavior with an extended finger. Right: A representation of a physical pointer, similar the one presented by Brägger et al., 2022. One of the shared view concepts, i.e., picture-in-picture view is highlighted with a red circle in all three panels.

- **Avatar based interactions and eye contact:** In social XR and metaverse studies, it has been shown that use of avatars can lead to a deeper connection through non-verbal communication (Etienne et al., 2023), thus we experimented with simpler avatar presentations to more sophisticated ones including facial expressions eye contact while in XR. This is possible by combining avatar representations with eye tracking, and/or enabling pass-through in the right moments for co-located collaborations (Figure P 67).
- **Mixing smartphone AR and VR:** In some scenarios, remote participation by many people as "audience" might be desired, such as in a virtual classroom setting, or in a virtual townhall meeting. Given that many citizens still do not have VR headsets at home, utilizing a many-to-one AR-VR system, we facilitated collaboration on spatial content while minimizing hardware requirements. In this setup, a single VR user is represented by an avatar and can use their hands or a pointing stick to convey information. Meanwhile, AR users can observe the scene on their smartphone through an AR app and use a ping functionality to communicate spatial information to the VR user (Figure P 67, middle).



Figure P 67. Example of the avatar representations included in the prototyping phase. Left: Abstract avatar hovering over a city model. Middle: Smartphone AR users enter the VR, and their phones are displayed as avatars with their names on it in the shape of a smartphone. Right: We integrated examples of customizable semi-realistic looking avatars Meta's avatar library that can mimic facial expressions of the users well. We also experimented with eye contact with the latter.

Technologies In the initial stages of our prototype development, we leveraged HoloLens 2 and Mixed Reality Development Toolkit (MRTK), and subsequently, consolidated all our collaborative XR prototypes in Quest 2 Pro (utilizing its passthrough mode for AR/MR needs) and developed the scenes with the Unity game engine.

Conclusions and a brief discussion

Human-computer interaction in XR is a research field with well-understood challenges given the familiar keyboard and mouse are taken away from the users. When human-human interactions mediated by XR become necessary, the complications are amplified, as two or more people's needs to be met and synchronized. At this junction, while there are interesting and promising developments, more research and new concepts are needed. Our comprehensive prototyping efforts contribute towards filling this gap. Based on initial interviews with stakeholders and user studies we conducted with each prototype, the interaction paradigms we explored above appear promising, and warrant further investigation towards user-validated XR collaboration frameworks.

Acknowledgements

This effort was (partially) funded by various grants over time including 50321.1 IP-SBM Digital Collaboration Hubs for Spatial Planning and Design Teams, FHNW's Lehrfonds program, 61697.1 INNO-EE SolarVR and Swiss Government Excellence Scholarships program. We are grateful for all intellectual contributions from all project members at various stages of the projects mentioned.

References

Baumgartner, L., Bragger, L., Koebel, K., Scheidegger, J., & Çöltekin, A. 2022. Visually annotated responsive digital twins for remote collaboration in mixed reality environments. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. V-4-2022, 329–336, <https://doi.org/10.5194/isprs-annals-V-4-2022-329-2022>

- Bragger, L., Baumgartner, L., Koebel, K., Scheidegger, J., & Çöltekin, A. 2022. Interaction and visualization design considerations for gaze-guided communication in collaborative extended reality. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-4-2022, 205–212, <https://doi.org/10.5194/isprs-annals-V-4-2022-205-2022>
- Çöltekin, A., Lochhead, I., Madden, M., Christophe, S., Devaux, A., Pettit, C., ..., & Kubiček, P. 2020. Extended Reality in Spatial Sciences: A Review of Research Challenges and Future Directions. *ISPRS International Journal of Geo-Information*, 9(7), 439. <https://doi.org/10.3390/ijgi9070439>
- Crogman, H. T., Cano, V. D., Pacheco, E., Sonawane, R. B., & Boroon, R. 2025. Virtual Reality, Augmented Reality, and Mixed Reality in Experiential Learning: Transforming Educational Paradigms. *Education Sciences*, 15(3), 303. <https://doi.org/10.3390/educsci15030303>
- Elmasllari, E., Kim, K.G., & Çöltekin, A. 2025. Does extended reality work for skills training? RIMMA2025 - International Conference on Forecasting, Preparedness, Warning and Response: Visualization, Communication and Information Management 28 - 30 January 2025, University of Bern, Switzerland. Abstracts of the ICA.
- 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.
- Fedosov, A., Gürçınar, E., Fluri, L., Bernasocchi, M., Kuhn, M., & Çöltekin, A. 2025. Designing AR Viewer for QField: Towards Supporting Handling Geospatial Data In Situ For Emergency Response Situations. RIMMA2025 - International Conference on Forecasting, Preparedness, Warning and Response: Visualization, Communication and Information Management 28 - 30 January 2025, University of Bern, Switzerland. Abstracts of the ICA.
- Lochhead, I., Hedley, N., Çöltekin, A., & Fisher, B. 2022. The Immersive Mental Rotations Test: Evaluating Spatial Ability in Virtual Reality. *Frontiers in Virtual Reality*, 3. <https://doi.org/10.3389/frvir.2022.820237>
- Macchi, G., & De Pisapia, N. 2024. Virtual reality, face-to-face, and 2D video conferencing differently impact fatigue, creativity, flow, and decision-making in workplace dynamics. *Scientific Reports*, 14(1), 10260.
- Proulx, M. J., & Çöltekin, A. 2025. Use of immersive technologies in design. *Frontiers in Computer Science*, 7, 1620569. <https://doi.org/10.3389/fcomp.2025.1620569>
- Oulasvirta, A. 2009. Social Inference Through Technology. In *Awareness Systems: Advances in Theory, Methodology and Design*, Panos Markopoulos, Boris De Ruyter, and Wendy Mackay (Eds.). Springer London, London, 125-147.
- Van Gent, S., Langefeld, A., Held, N., Cremers, A., Loots, G., & Gunkel, S. N. 2024. An experimental comparison of participants' experience in face-to-face, video, and virtual reality meetings. *Frontiers in Virtual Reality*, 5, 1463189.

Demos

Z-3PO: Seamless Switching Between 3 Representation Perspectives for Real-Time Immersive Teleconferencing

Simon N.B. Gunkel, Tessa Klunder, Galit Rahim, and Gianluca Cernigliaro

TNO (Netherlands Organization for Applied Scientific Research)

Corresponding author: Simon.Gunkel@tno.nl

Keywords: Holographic Communication, Extended Reality (XR), Social XR, RGBD Capture, Photorealistic 3D Representation, Remote Training, Immersive Teleconferencing, Education

Abstract

In this (demo) paper we present Z-3PO a capture solution to support seamless switching between 3 real-time representations for humans and objects in immersive 3D applications. We integrated this capture application into a unity demonstrator connecting end devices via a commercial WebRTC pipeline (openrainbow.com). Our prototype allows a teacher and student to collaboratively engage in a Lego block building task while seamlessly explore the different hybrid representation perspectives (3D stereo, 2D billboard and 3D Mesh) in Mixed Reality.

Introduction

The accelerating pace of technological advancement across industrial sectors has exposed a critical global challenge: a growing shortage of skilled experts (Gunkel et al., 2025a). As industries evolve, the demand for specialized knowledge and technical proficiency increasingly outpaces the capacity of traditional training systems (Islam et al., 2025). This imbalance is further exacerbated by limited access to training facilities, geographic constraints, and the complexity of modern industrial tasks. Consequently, organizations face mounting difficulties in maintaining operational efficiency, ensuring safety, and fostering innovation, all of which are contingent on a well-trained workforce.

To address this skills gap, immersive technologies such as Extended Reality (XR)—encompassing virtual, augmented, and mixed reality—offer a compelling solution. XR enables realistic, interactive training environments that simulate complex industrial scenarios, allowing for hands-on learning without the constraints of physical presence (Gunkel et al., 2025b). Photorealistic, real-time representations of humans and objects enhance the fidelity and effectiveness of remote instruction, quality control, and collaboration. These capabilities not only mitigate the impact of expert scarcity but also pave the way for scalable, accessible, and high-quality industrial training solutions.

The main goal of Z-3PO is to enable the next paradigm for collaborative business meetings and remote training experiences in XR, with a specific focus on natural communications between humans based on the following three limitations.

Limitation 1, **Human presence**: bringing real people in XR is key for realistic interactions. Remote XR experiences are normally based on virtual avatars not able to convey the true feeling of “being there together”. Z-3PO enables real humans remotely located to join shared XR experiences, specifically considering the use cases: i) business meetings, where users communicate with each other in a many-to-many scenario and ii) remote training where a trainer and a trainee interact in a one-to-one scenario.

Limitation 2, **Real-life experience**: XR applications are usually based on computer generated assets that, in case of collaborative experiences, can be controlled by multiple users. However, the integration of virtual assets in real-world scenarios often lacks the tactile authenticity of object manipulation. Thanks to the multiple holographic representation formats available, Z-3PO allows augmentation of reality with different virtual representations of users and objects.

Limitation 3, **Real-time and multi-modal communication in XR**: Holographic communications encounter challenges due to high volumes of data to be processed and solutions vulnerable to real-time constraints. Furthermore, even though multiple representation modalities were introduced and tested in the past (Gunkel, 2024). It is not always clear how suitable these formats are under different conditions and use cases and direct comparisons are cumbersome. With Z-3PO we introduce a demonstrator that provides 3 photorealistic real-time representation formats and allows seamless switching between them without the need of complex reconfiguration or restarting of components or end-devices.

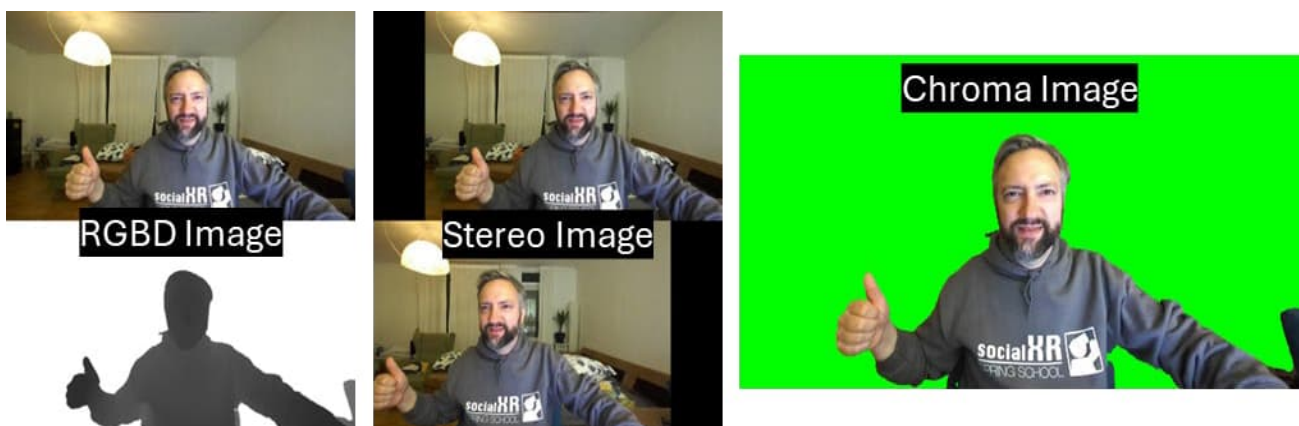


Figure D 1. Three capture and render perspectives for industrial trainings (3D mesh, 3D stereo, 2D Chroma).

Solution

The technical basis for Z-3PO is the RGBD 3D user representation format, and modular capture component presented in Gunkel et al. (2023), and the stereo format presented in the “remote training” prototype (Gunkel et al., 2025b). Ultimately the Z-3PO prototype integrates the following 3 representations (see Figure 1) into a flexible end-to-end (capture, networking, and rendering) pipeline:

- **2D billboard (chroma)**: detailed representation of humans or objects with greenscreen cutout effect and thus blended into the virtual environment
- **3D Stereo**: using stereo sensors for a highly detailed 3D view, this for example allows a view of objects or a workbench in complex (technical) training tasks
- **3D RGBD**: photorealistic 3D representation of humans as point cloud or mesh texture

All 3 formats are widely compatible with existing video compression and transmission technologies and allow real-time photorealistic capture and rendering. Figure D 2 shows the full end-to-end architecture of Z-3PO. A capture module (served by a stereo and depth camera, Zed 2i) is responsible for capturing the 3 formats, creating all necessary metadata and allowing switching between the formats. The capture module is connected to a unity client via virtual webcam interface (for video data) and a local webserver (for metadata). The unity client (running on a Laptop with dedicated graphics card) sends and receives video, audio and metadata via WebRTC (Rainbow SDK)²⁵ and renders the combined local and remote view (users and objects) into a mixed reality head mounted display (HMD). The rendering is changed based on the formats and metadata, to reproduce a photorealistic natural representation of the users and objects. The main benefits of the Z-3PO are:

- A. 3 representation formats integrated into a commercial (operational) WebRTC pipeline
- B. Easy way to compare different formats (real-time switch between formats)

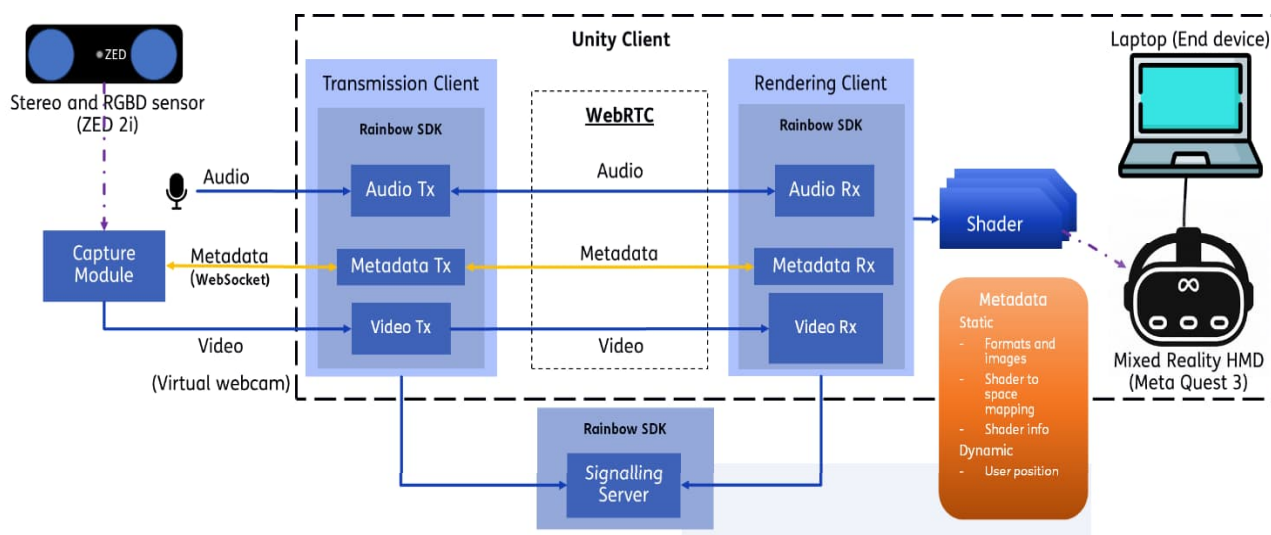


Figure D 2. System Architecture.

Demonstrator Outline

Participants can engage in a remote Lego building task with a "student" building an Lego object locally assisted by a remote teacher (similar as presented in Gunkel et.al., 2025b)²⁶. Users can see their local environment, including (Lego) objects and interaction, in a Meta Quest 3 Mixed Reality glasses (with pass-through activated) and the possibility to toggle between the 3 representation perspectives (stereo, chroma and 3D). Thus, the demo is similar to the "Ready Expert One" experience presented in Gunkel et al²⁶, (2025b) with the key difference of a different rendering HMD (the Meta Quest 3 offering higher detail in reading with a higher resolution and field of view) and the ability to directly compare different representation formats in a fluent experience (no restart needed during switching).

²⁵ openrainbow.com

²⁶ <https://youtu.be/5XFPV2Zzosg>

Initial Measurements

To evaluate the technical feasibility of our demonstrator we performed some initial measurements deployed in local setup connected via the Rainbow infrastructure. The test was executed on a laptop (Acer Predator Helios 300, PH315-55s-917f) with internet over Wi-Fi. We measured the CPU / GPU usage of the Z-3PO capture module and latency for the 3 representation formats. Latency was measured with the DelAyrUco tool²⁷ as capture to render (local) and full end-to-end latency (remote). For the latency measurements we utilized the Z-3PO capture module and the browser version of Rainbow²⁸. The results of our measurements can be seen in Table D 1. The performance (CPU / GPU) of the capture module is stable and similar across a representation modality. The remote delays measures might indicate some inconsistencies and need further investigation but overall show technical readiness of our approach.

Table D 1. Initial technical measurements of Z-3PO.

	3D-RGBD (MESH)	3D-Stereo	2D-Chroma
Capture CPU in %	4.22 (std. 0.21)	4.13 (std. 0.19)	4.96 (std. 0.26)
Capture GPU in %	35.76 (std. 2.13)	35.34 (std. 2.5)	37.54 (std. 1.0)
Capture Latency in ms	170 (std. 27)	148 (std. 21)	143 (std. 18)
Full Latency in ms	514 (std. 35)	498 (std. 27)	641 (std. 17)

Conclusion and Future work

We presented Z-3PO, a real-time XR teleconferencing solution that enables seamless switching between three photorealistic representation formats—2D chroma, 3D stereo, and 3D RGBD. Integrated into a commercial WebRTC pipeline, Z-3PO supports immersive, natural communication for remote training and collaboration. Initial technical tests confirm stable performance and acceptable latency across formats. Future work includes expert user evaluations to assess usability and representation preferences. We also plan to explore adaptive streaming (via Artificial Intelligence) to enhance performance and user experience.

Acknowledgments

This work was partially funded by the European Union: CORTEX2²⁹ OC#2 (101070192), R3in3D (RGBD Real-Time Representations of Humans and Objects in 3D).

References

- Gunkel, S.N.B., Dijkstra-Soudarissanane, S., Stokking, H.M., & Niamut, O.A. 2023. From 2D to 3D video conferencing: Modular RGB-D capture and reconstruction for interactive natural user representations in immersive extended reality (XR) communication. *Frontiers in Signal Processing*, 3, 1139897. <https://doi.org/10.3389/frsip.2023.1139897>

²⁷ <https://github.com/TNO/DelAyrUco>

²⁸ <https://web.openrainbow.com/>

²⁹ <https://cortex2.eu/>

- Gunkel, S.N.B. 2024. From 2D Video Conferencing to Photorealistic Immersive 3D Communication.
- Gunkel, S.N.B., Klunder, T., & Cernigliaro, G. 2025a. R3-D3: (Photo-) Realistic Real-time Representations of Dynamic 3D-Driven Industrial Trainings. In ACM International Conference on Interactive Media Experiences Workshops (IMXw) (pp. 140–144). SBC.
- Gunkel, S.N.B., Klunder, T., & Cernigliaro, G. 2025b. Ready Expert One: Universal 3D Workbench for Remote Industrial Training. In Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (pp. 359–361).
- Islam, M.T., Sepanloo, K., Woo, S., Woo, S.H., & Son, Y.-J. 2025. A review of the industry 4.0 to 5.0 transition: Exploring the intersection, challenges, and opportunities of technology and human–machine collaboration. *Machines*, 13(4), 267. <https://doi.org/10.3390/machines13040267>

XR Meets AI and Biofeedback: Technology-Enhanced Stress Management Training for Healthcare Professionals

Mirella Moser, and Katja Fraedrich

ZHAW School of Management and Law, Switzerland

Corresponding author: mosl@zhaw.ch

Keywords: Virtual Reality (VR); Real-Time Biofeedback; Conversational AI; Speech Interfaces; GPT-based Dialogue; Physiological Monitoring; Immersive Learning; Stress Management; Nursing Education; Healthcare Training

Context

In the fast-paced world of the 21st century, work-related stress has become a ubiquitous phenomenon. Mental disorders related to stress and anxiety demonstrate alarmingly high prevalence rates (Salari et al., 2020; World Health Organization, 2016). The consequences of stress-related disorders such as burnout syndrome are manifold, impairing not only individual health and professional performance (Salvagioni et al., 2017) but also creating significant societal burdens, including substantial economic costs (Hassard et al., 2018) and potential threats to the stability of essential public services. This is especially evident within the healthcare sector, where nursing is consistently classified as one of the most stressful professions (Johnson et al., 2005).

Stressors faced by nursing professionals are varied, with problematic interactions involving challenging patients playing a particularly significant role (Treviranus et al., 2021). Dealing professionally and goal-oriented with patients and their relatives often requires nurses to suppress their spontaneous emotional reactions, placing high demands on their emotional self-control (Schmucker, 2020). As stress-related strain negatively affects nurses' health, the quality of healthcare, and retention of essential skilled professionals (BAG, 2022), effective, practical, and scalable interventions for building stress management skills among current and future healthcare professionals are urgently needed.

Addressing this critical need, our demonstration presents a proof-of-concept developed within an advanced Innosuisse-funded project together with PH Zurich, Careum Verlag and Bandara VR. We introduce an innovative training solution that leverages Extended Reality (XR), Artificial Intelligence (AI), and biofeedback mechanisms to enhance stress management competencies among nursing professionals. By integrating immersive simulations, real-time physiological monitoring, personalized AI-driven feedback, and AI-generated immersive dialogues with avatars, this solution aims to empower healthcare workers to proactively manage workplace stress, thereby improving individual health outcomes and sustaining quality patient care.

Description of the VR Learning Application

The VR application presented in this demonstration is structured as a comprehensive learning journey designed explicitly to enhance stress management competencies among nursing professionals. Throughout the VR experience, learners' heart rates are continuously measured using a Bluetooth-connected heart rate monitor. Real-time biofeedback is visualized directly within the VR environment through color-coded indicators and numerical heart rate displays, thereby allowing immediate awareness and reflection on physiological stress responses.

The VR learning journey unfolds systematically across five sequential and distinct phases:

- **Lobby Entrance:** Participants initially enter a virtual lobby, providing an introductory orientation to the VR technology alongside an overview of the training objectives.
- **First Stress Scenario:** Learners encounter a simulated interaction with avatars representing challenging patient relatives. This scenario incorporates deliberately introduced stressors, including disruptive stimuli, to realistically mimic emotionally demanding situations common in nursing.
- **Relaxation Scenario:** Participants engage in guided breathing exercises specifically aimed at promoting relaxation and fostering self-regulation skills. This phase also enables learners to become familiar with their individual physiological stress patterns.
- **Second Stress Scenario:** Building on previous experience, learners face another interaction with challenging avatars, this time under altered conditions to test newly acquired stress management strategies.
- **Reflection Phase:** The VR experience concludes with a structured reflection and discussion session, in which participants analyze their experiences and insights gained from observing their heart rate data and their communication behavior throughout the scenarios.



Figure D 3. Stress Scenario

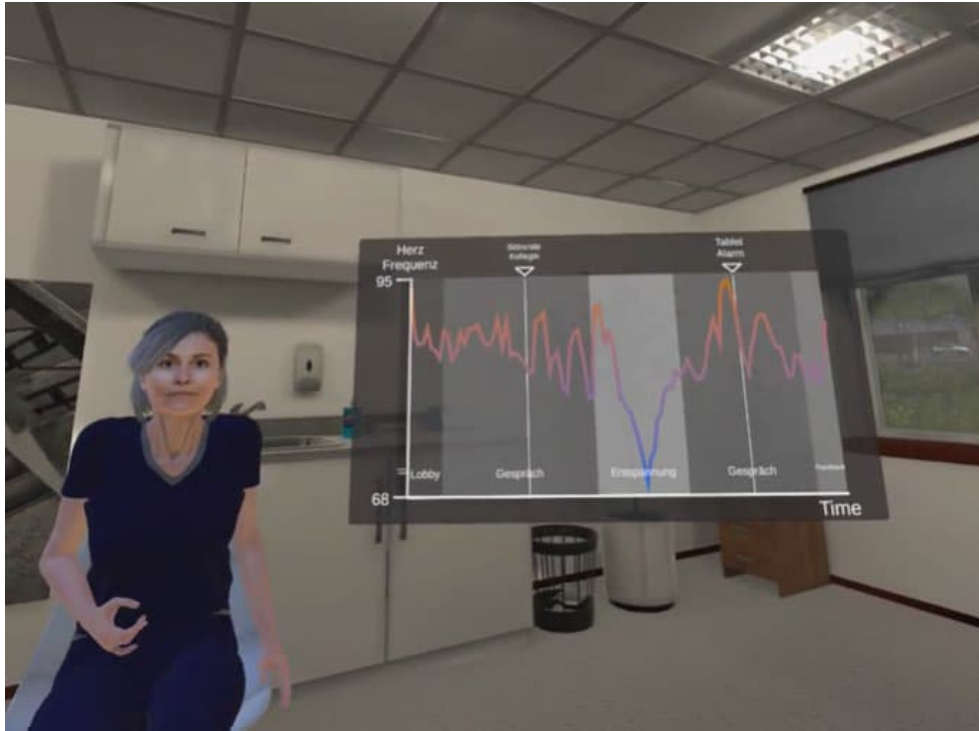


Figure D 4. Reflection Phase

System Architecture and Interaction Design

At the heart of the training system is a standard-alone VR application built in Unity (Unity Technologies, 2024) and optimized for the Meta Quest 3 (Meta, 2023). Users interact naturally through hand tracking, removing the need for controllers and allowing for intuitive engagement with virtual scenarios. To monitor physiological responses during stress-inducing interactions, the system connects to a Bluetooth Low Energy (BLE) heart rate sensor, like the Polar OH1+ (Polar Electro Oy, 2024), which streams live pulse data into the VR environment. This data is visualized continuously, giving users a tangible sense of their own stress reactivity as they navigate emotionally charged situations.

A central feature of the experience is the ability to engage in free-flowing conversation with virtual characters. Using Azure's speech-to-text service (Microsoft Azure, 2024), spoken input is transcribed in real time and fed into a custom GPT-based dialogue system (OpenAI, 2024). This enables participants to speak naturally with avatars, who respond contextually based on the ongoing exchange. Visual realism is heightened through expressive facial animation and real-time lip syncing, while voices are generated using ElevenLabs' lifelike speech synthesis (ElevenLabs, 2024). These design choices aim to create emotionally authentic encounters that feel immediate and personal.

Beyond the moment-to-moment interaction, the AI system plays an additional role in supporting reflection. After completing the second stress scenario, participants enter a guided debrief phase where the AI reviews the recorded conversation and prompts users to reflect on their performance. Questions such as "What did you do well?" are followed by tailored follow-ups based on earlier dialogue, helping learners identify effective strategies and areas for improvement.

Together, these components form an integrated system that combines immersive interaction, real-time biofeedback, and adaptive AI to support emotional resilience training in a deeply engaging and personalized way.

References

- BAG. 2022. Faktenblatt Berufsverweildauer in der Pflege. Bern: Bundesamt für Gesundheit BAG.
- ElevenLabs. 2024. ElevenLabs: AI Speech Synthesis. <https://elevenlabs.io> (accessed: 01.07.25).
- Hassard, J., Teoh, K. R. H., Visockaite, G., Dewe, P. & Cox, T. 2018. The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology* 23, 1–17. <https://doi.org/10.1037/ocp0000069>
- Johnson, S., Cooper, C., Cartwright, S., Donald, I., Taylor, P. & Millet, C. 2005. The experience of work - related stress across occupations. *Journal of Managerial Psychology* 20(2), 178–187. <https://doi.org/10.1108/0268394051057980>
- Meta. 2023. Meta Quest 3: Technical Specifications. Menlo Park, CA: Meta Platforms Inc. <https://www.meta.com/quest/quest-3/> (accessed: 01.07.25).
- Microsoft Azure. 2024. Azure Speech-to-Text Service. Redmond, WA: Microsoft Corporation. <https://azure.microsoft.com/services/cognitive-services/speech-to-text/> (accessed: 01.07.25).
- OpenAI. 2024. GPT (Generative Pre-trained Transformer). San Francisco, CA: OpenAI. <https://openai.com/blog/chatgpt> (accessed: 01.07.25).
- Polar Electro Oy. 2024. Polar OH1+: Optical Heart Rate Sensor. Kempele, Finland: Polar Electro Oy. <https://www.polar.com/ch-de> (accessed: 01.07.25).
- Salari, N., Hosseini-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulpoor, S., Mohammadi, M., Rasoulpoor, S. & Khaledi-Paveh, B. 2020. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health* 16(1), 57.
- Salvagioni, D. A. J., Melanda, F. N., Mesas, A. E., González, A. D., Gabani, F. L. & Andrade, S. M. de. 2017. Physical, psychological and occupational consequences of job burnout: A systematic review of prospective studies. *PLOS ONE* 12(10), e0185781. <https://doi.org/10.1371/journal.pone.0185781>
- Schmucker, R. 2020. Arbeitsbedingungen in Pflegeberufen: Ergebnisse einer Sonderauswertung der Beschäftigtenbefragung zum DGB-Index Gute Arbeit. *Pflege-Report 2019: Mehr Personal in der Langzeitpflege – aber woher?*, 49–60.
- Treviranus, F., Mojtahedzadeh, N., Harth, V. & Mache, S. 2021. Psychological stress factors and resources in outpatient nursing. *Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie* 71, 32–37.
- Unity Technologies. 2024. Unity Real-Time Development Platform. San Francisco, CA: Unity Software Inc. <https://unity.com> (accessed: 01.07.25).
- World Health Organization. 2016. Monitoring and evaluating digital health interventions: A practical guide to conducting research and assessment. Geneva: World Health Organization. <https://apps.who.int/iris/handle/10665/252183> (accessed: 08.09.23)

Control Your XR: A Dashboard for XR Applications

Urs Riedlinger, Yücel Uzun, Leif Oppermann

Fraunhofer FIT, Germany

Corresponding author: urs.riedlinger@fit.fraunhofer.de

Keywords: XR, dashboard, application, control, demonstrator

Abstract

This demo showcases our ongoing work on a prototype system we call the “MARS dashboard,” designed to support live control and monitoring of XR applications. “MARS” stands for Mixed and Augmented Reality Solutions. In scenarios such as public demonstrations, research environments, or even smaller internal events, XR users are often immersed in headsets with no clear line of communication to or from facilitators or researchers. Our dashboard bridges this gap, enabling an external operator to monitor user progress, adjust system parameters, and provide support, without requiring the user to remove the headset or interrupt their experience. Implemented with Node.js³⁰ server on the backend and Angular³¹ on the frontend, the dashboard seamlessly connects to XR applications developed in Unity. It is designed to be engine-agnostic, supporting future connections with Unreal and other platforms. This submission demonstrates the functionality of the dashboard and the implications for usability, accessibility, and experimenter control in real-time XR scenarios.

Introduction and Motivation

Extended Reality (XR) systems are becoming increasingly prevalent in both public and research-oriented contexts; yet, one persistent challenge remains: how can those outside the headset meaningfully observe, influence, or assist the immersive experience in real-time? Whether it’s a user trying an XR prototype at a trade fair, a participant in a user study, or a student in an educational training, external guidance and monitoring are often necessary to ensure a smooth and practical session. In our experience and observation, it is common practice for exhibitors to request that users remove or return the headset whenever a problem or ambiguity arises, resulting in a disruption of the immersive experience.

To address this challenge, we have developed a web-based dashboard that functions as a central control and observation panel for our XR applications. Rather than relying on headset mirroring or

³⁰ <https://nodejs.org> 2025

³¹ <https://angular.dev/> 2025

hidden admin interfaces in XR, our dashboard provides a simple, flexible and intuitive browser-based UI for controlling parameters that's been defined and required by the application, such as difficulty, resetting the application, monitoring logs, or evaluating performance, all without interfering with the user's XR immersion. The solution enables facilitators to stay in control and remain informed, without requiring prior XR experience or extensive technical knowledge.

Technical Implementation

The dashboard is built as a lightweight, flexible and modular application. The backend consists of a Node.js server that uses WebSocket to enable real-time, bidirectional communication between the frontend, database, and XR applications. All communication is handled using consistently formatted and structured JSON data to ensure flexibility for connected systems. Its frontend is implemented in Angular, offering a modern and responsive interface that runs directly in any standard browser. This choice ensures high compatibility with various devices, from desktop monitors at trade show booths to tablets in mobile field studies. The XR side is currently developed in Unity Technologies³² and uses a custom WebSocket client to interpret commands from the dashboard and return telemetry and state updates.

Our architecture is designed to be flexible and engine-agnostic. As long as an application can connect to our WebSocket server and interpret the structured JSON data, it can be integrated with the dashboard, regardless of the technology stack or programming language used. The JSON data types are stored in a separate repository and converted semi-automatically from TypeScript to, for example, C#, ensuring that the client and server share the same types. This makes our solution suitable not only for our internal demos but also as a framework that could be adopted by other teams working with different XR platforms.

One important feature is the ability to operate the dashboard also in a local network. This is particularly important for larger events, where a stable internet connection cannot be relied upon. Our client hence searches for a dashboard connection locally and remotely, and it is possible to prioritise one over the other.

The current functionalities of the dashboard include application selection, configuration sliders, buttons, real-time user logging with filters and tagging, and access to usage statistics and leaderboard rankings (see Figure D 5, Figure D 6, and Figure D 7). These features have been integrated to meet the demands of varied real-world deployments, ranging from quick live demos to extended experimental sessions.

Use Cases and Applications

The dashboard was initially developed for Hannover Messe 2025 as part of an interactive XR installation showcasing the future of work, with and without AI support. Within a single application, visitors could explore two distinct demo scenarios: one focused on a recycling plant and the other on an automotive wing inspection process. Both experiences highlighted how AI can augment human performance in industrial tasks. The dashboard served as the central control tool, enabling facilitators to monitor progress, adjust settings per visitor, and compare AI-assisted versus manual performance in real-time. Designed for both public and professional audiences, it enabled smooth operation even for

³² <https://unity.com/> 2025

non-technical staff. While initially built for this event, the dashboard is intended for broader use and currently being integrated into additional projects and XR applications.

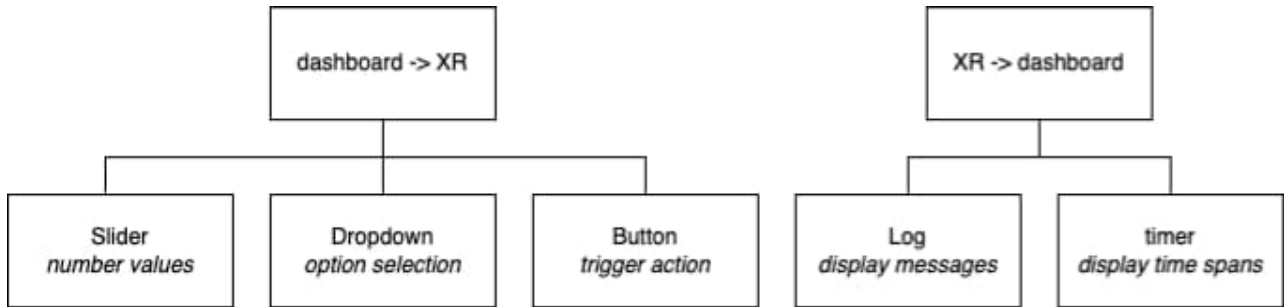


Figure D 5. Overview of the functionalities of the dashboard

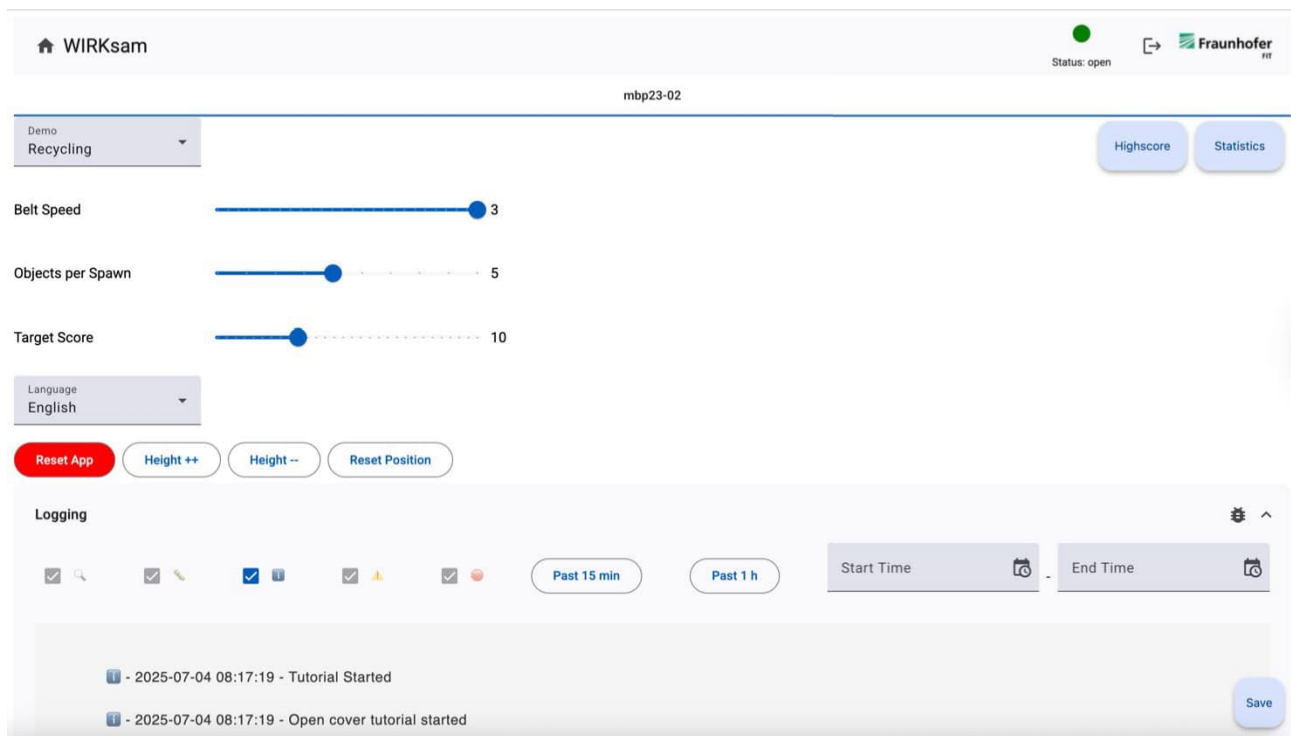


Figure D 6. Screenshot of the dashboard control interface

🏆 Tag Daily	📅 Woche Weekly	Statistics
Driving-Pink-Muskox 16.08.42 🤖 AI: 533 🧑 No AI: 325	Agreed-Cranky-Cat 2025-03-20 🤖 AI: 12000 🧑 No AI: 12000	Total Users: 435
Static-Coffee-Eel 16.23.03 🤖 AI: 416 🧑 No AI: 250	Agreed-Drunk-Puma 2025-03-20 🤖 AI: 10000 🧑 No AI: 10000	Wing Demo Total Users: 354 Users that were better with AI: 134 Users not finished: 71
	Agreed-Coffee-Puma 2025-03-20 🤖 AI: 1000 🧑 No AI: 1000	Total Wings: 2586 Total Errors Found: 3905 Total Errors Missed: 1219 Total Warnings: 801
	Relaxed-Bronze-Platypus 2025-03-20 🤖 AI: 888 🧑 No AI: 791	Recycling Demo Total Users: 81 Users that were better with AI: 20 Users not finished: 7
	Visual-Brown-Capybara 2025-03-21 🤖 AI: 975 🧑 No AI: 624	Total Correct Sortings: 311 Total Wrong Sortings: 138
	Uptight-Rose-Frog 2025-03-20 🤖 AI: 996 🧑 No AI: 499	
	Perfect-Aqua-Cuckoo 2025-03-20 🤖 AI: 778 🧑 No AI: 697	

Figure D 7. Screenshot of the highscore (left) and statistics (right) collected by the dashboard

Data Logging, Monitoring, and Visualisation

A core feature of the dashboard is its robust logging system. Anything an application developer wants to log, such as user interaction, system event, or AI intervention, is timestamped and categorised using a set of icons for clarity. Events can be filtered by type (e.g., information, warning, error), and the timeframe can be adjusted to view only recent activity (e.g., the past 15 minutes) or to specify a specific start and end timestamp. These logs support real-time decision-making during demos, as well as post-hoc analysis in usability testing or experimental evaluations.

The dashboard also provides access to live statistics and high scores. These metrics can be defined by the application and are presented in a clean, scrollable format, comparing daily and weekly performance, both with and without AI assistance. For instance, in the Wing Demo, out of 354 total users, 134 performed better when using AI support. In the Recycling Demo, 20 out of 81 users improved their scores with the help of AI. Such quantitative insights, updated live, offer both entertainment value at public events and meaningful feedback in research studies.

Discussion and Future Work

This dashboard is more than a control tool – it is an enabler of effective XR deployment in diverse real-world settings. By reducing the technical barrier to interaction, it empowers facilitators, researchers, and even casual users to manage XR sessions with minimal training. The real-time nature of the tool encourages adaptive support, improving user experience and reducing the likelihood of disengagement or confusion. Nonetheless, we see several paths forward for extending its capabilities. We are exploring the integration of camera previews from XR headsets to further enhance the observer context without requiring direct mirroring. Furthermore, we aim to integrate more device statistics, such as the current battery or audio level, so that users can prevent device failures. Additionally, we are considering introducing analytics dashboards that automatically generate summary reports after each session. Finally, gesture and speech recognition for facilitators could offer hands-free control in constrained environments. Our long-term vision is to establish the dashboard as a general-purpose interaction layer for XR systems, especially those designed for public, experimental, or cooperative use. By providing a consistent, responsive, and engine-agnostic control panel, we hope to support a broader range of XR experiences that are both immersive for users and manageable for facilitators.

Acknowledgments

This work is part of ongoing research and prototyping efforts at Fraunhofer FIT. We thank all our colleagues and participants who contributed feedback and insights throughout the development. Parts of this paper are based on work from the research and development project “WIRKsam”, which is funded by the Federal Ministry of Research, Technology and Space (BMFTR) as part of the “Regional Competence Centres for Labour Research” funding measure (funding code: 02L19C617) and supervised by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

Transmitting Volumetric Video via Media over QUIC to Subscribers with Contradicting Spatial and Temporal Preferences

Felicián Németh^{1,3}, István Pelle^{1,3}, and Tamás Lévai^{2,3}

¹*HUN-REN-BME Cloud Applications Research Group, Hungary*

²*HUN-REN-BME Information Systems Research Group, Hungary*

³*Budapest University of Technology and Economics, Hungary*

Corresponding author: nemethf@tmit.bme.hu

Keywords: Virtual Reality, Virtual conferencing, Volumetric video, Multimedia communication, Network protocols, Media over QUIC

Abstract

Real-time volumetric video (VV) is crucial for applications like gaming, healthcare, and education. However, its high bandwidth and low-latency demands challenge traditional transmission methods. We replace WebRTC in an immersive teleconferencing system (De Fré et al., 2024) with Media over QUIC (MoQ), enabling scalable, low-latency VV delivery through a simpler architecture that leverages QUIC's benefits. By organizing point cloud video into multiple layers, bandwidth-limited users receive only a subset of data without the need for explicit bandwidth measurement. A proposed small protocol extension supports users who prioritize spatial accuracy, such as a semantic segmentation microservice requiring all layers, even at lower frame rates. This approach accommodates the needs of resolution and liveness sensitive users within the same call.

Introduction

Real-time volumetric video (VV) is rapidly gaining traction across various domains, including entertainment, healthcare, education, and remote collaboration. Among these, immersive videoconferencing and telepresence stands out as a promising and widely requested application, allowing users to interact in three-dimensional, spatially rich environments. Recent advances have enabled the capture, compression, and transmission of point cloud (PC) video, making volumetric telepresence feasible.

A notable example of this progress is the WebRTC-based volumetric teleconferencing platform (De Fré et al., 2024), which demonstrated the viability of real-time PC video streaming for multi-user conferencing that streams Draco-encoded PC streams (*i.e.*, videos) in which each frame is independent of the others. However, while WebRTC provides robust, low-latency transport, it also introduces architectural complexity that faces scalability challenges, particularly as the number of users and the heterogeneity of network conditions increase.

This work presents a novel approach that replaces the WebRTC transport layer in a Unity-based PC teleconferencing system with Media over QUIC (MoQ). To our knowledge, this is the first demonstration of MoQ being used for PC video delivery, and the first VR video conferencing application leveraging MoQ. Our solution enables a significantly simplified architecture, leveraging MoQ's publish-subscribe model and QUIC's inherent scalability and reliability to support large-scale, real-time volumetric streaming.

Crucially, our MoQ-based system adapts to bandwidth fluctuations more gracefully than prior art thanks to QUIC's congestion control and our semantic segmentation microservice, which clusters PC feature points and assigns priorities based on their relevance to the conferencing context (e.g., prioritizing human faces over less critical scene elements). Unlike previous approaches, our method eliminates the need for explicit bandwidth measurement.

Furthermore, our implementation introduces flexible transmission preferences, allowing users and services to select between spatial and temporal prioritization, thus fine-tuning the viewing experience. We validate our approach with a live demonstration, showcasing the system's ability to support diverse user and service requirements in real-time.

Live Volumetric Video Transmission via Media over QUIC

Media over QUIC is an IETF initiative to enable low-latency, large-scale media transmission (Nandakumar et al., 2025). Its broad use cases include videoconferencing, live streaming, and even futuristic scenarios such as interstellar communication. MoQ maintains a straightforward architecture that includes publishers, subscribers, and relays. Relays are responsible for transmitting the content between the publisher and subscribers. Publishers inject content into the relay network, where the *content mapping scheme* is the following: data is structured as objects (e.g., video frames) grouped into groups (e.g., a group of pictures) and organized into tracks (continuous media streams). A set of tracks is a namespace, which provides an abstraction to bundle multiple streams (e.g., video, audio, and text tracks of a participant in a video conference call). Subscribers retrieve tracks from relays to reconstruct media streams.

In MoQ, prioritization is a fundamental mechanism to ensure efficient and timely delivery of media objects, especially under network congestion. MoQ allows the subscriber and the original publisher to assign priorities to objects, groups, and tracks. When multiple objects are ready for transmission, MoQ prioritizes them by subscriber priority, publisher priority, and order of groups or object IDs within the same track. This prioritization enables critical media data, such as keyframes, or live event updates, to be delivered ahead of less time-sensitive content, optimizing playback quality and minimizing latency. Control messages receive the highest priority to maintain session integrity.

Our implementation builds upon the volumetric video framework established in prior work (De Fré et al., 2024), adapting its core principles to Media over QUIC's transport model. The adaptation involves three key design choices. First, the content mapping is not straightforward; different schemes have different properties. In our scheme, each Draco-encoded PC instance at a discrete timestamp is encapsulated as a single MoQ object. Since these PCs function as independent keyframes (lacking temporal dependencies), we assign each object to its group. This atomic grouping ensures minimal decoding latency and eliminates inter-frame blocking, critical for real-time volumetric conferencing.

Second, unlike the prior WebRTC implementation that multiplexed quality layers, we distribute the PC's different detail levels (LODs) across separate tracks. Suppose the bandwidth between the relay and the subscriber is insufficient to transmit all the tracks. In that case, the relay tries to send the most recent objects from lower priority tracks only until a new higher priority object arrives to it. This mechanism

effectively eliminates the need for explicit bandwidth measurements. This architectural shift results in subscribers implicitly receiving tracks fitting into their network capacity (e.g., low-LOD tracks for bandwidth-constrained users); priority isolation: high-detail tracks can be assigned lower priority without affecting low-LOD delivery; and decoding efficiency: clients can process independent tracks in parallel, avoiding layered decoding bottlenecks.

Third, we implement a dynamic prioritization scheme leveraging MoQ's native priority fields. Priorities are dynamically adjusted per subscriber using SUBSCRIBE_UPDATE messages based on field of view optimization (e.g., deprioritize participant's all PC tracks if not in the viewport).

Each publisher's content is isolated under a unique namespace (e.g., /spirit/{user-id}/), containing an application control track (highest priority), an audio track, and PC tracks. By decoupling LODs into independent tracks (rather than layered substreams), our approach eliminates the need for complex inter-layer synchronization. It enables heterogeneous subscription patterns from bandwidth-constrained mobile users receiving only the low-quality PC track to resource-rich segmentation microservices subscribing to all LODs. If the microservice is bandwidth-constrained, we propose an extension to the current MoQ draft that defines a new subscription property named LINKED_TRACK. With it, a subscriber can join the groups of different tracks together. Once a relay starts to serve an object from a group g of a track, it will not start serving from another group while there are unsent objects from group g of a linked track. This extension supports contradicting spatial and temporal transmission preferences in a single call with the same content mapping scheme.

Showcasing PC Transmission with Semantic Segmentation over MoQ

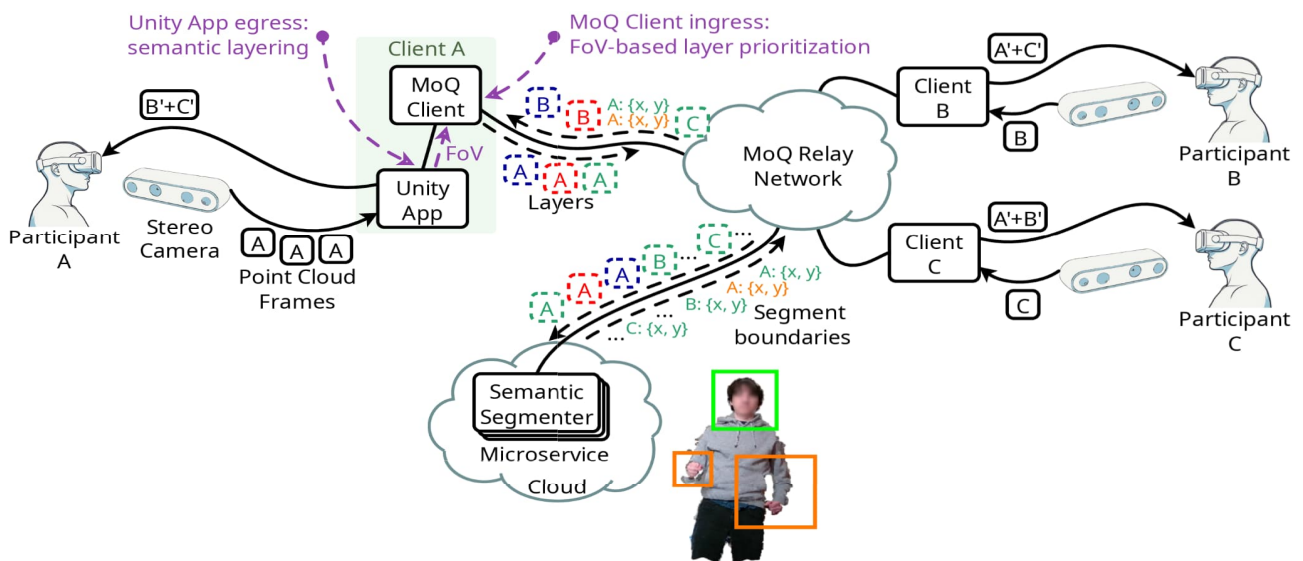


Figure D 8. System architecture.

Figure D 8 shows the high-level system architecture of the demo. The data flow is the following. A stereo camera records the participant's PC frames during the teleconference. Based on the previous segmentation boundaries returned by the Semantic Segmenter, the Unity App splits PC A into separate disjunct sub-PCs (blue, red, and green colored A letters in Figure D 8). The sizes of the sub-PCs are pre-set and independent of the size of the segments. For example, the highest priority sub-PC might contain the head segment and a random subset of the hand segment, or just a random subset of the head segment if the participant leans into the camera. The Unity App then forwards the sub-PCs and FoV info to a special MoQ Client. The Client transmits the disjunct sub-PCs in separate tracks to the relay network. Based on the available bandwidth, Client B can receive just a subset of other participants'

objects in their tracks. For example, it receives A' , the two highest priority sub-PCs from A and C', one of the highest priority sub-PCs from Client C. The Unity App of Participant B then decodes the sub-PCs, places them into its virtual room, and sends the results ($A'+C'$) to Participant B. MoQ Client A periodically re-prioritizes the tracks of other participants, relying on its FoV info. The new subscriber priorities are advertised to the relay network via SUBSCRIBE_UPDATE messages.

The Semantic Segmenter microservice utilizes a standard MoQ client; however, it links the tracks of a participant with the LINKED_TRACK property. Without the linking, it might receive just the highest priority sub-PC. Still, if the participant moves fast, this sub-PC might no longer contain the participant's head. Therefore, it is preferable to receive the full PC less frequently than to receive a portion of the PC more frequently. The semantic segmentation converts the received full-resolution PCs to 2D images using PyVista then infers (head and hand) segments employing MediaPipe (Lugaresi, et al., 2019) Holistic landmarking.

Conclusions

This work demonstrates the feasibility and advantages of using Media over QUIC for real-time volumetric video conferencing. By mapping point cloud video tracks into MoQ objects and leveraging MoQ's publish-subscribe architecture, we enable a simplified, scalable, and low-latency teleconferencing solution well-suited to the high bandwidth and strict timing requirements of immersive 3D communication.

Acknowledgements

This project is sponsored by the 1) SPIRIT project (Grant Agreement No. 101070672)³³; activities related to semantic segmentation were supported by 2) project No. 2024-1.2.4-KDT-2024-00002 provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, under the 2024-1.2.4-KDT funding scheme and 3) the CHIPS JU as part of the EU's Horizon Europe research and innovation programme, SMARTY Project (grant agreement No. 101140087).

References

- De Fré, M., van der Hooft, J., Wauters, T., & De Turck, F. 2024. Demonstrating Adaptive Many-to-Many Immersive Teleconferencing for Volumetric Video. In ACM MMSys '24, 453–458. <https://doi.org/10.1145/3625468.3652192>
- Nandakumar, S., Vasiliev, V., Swett, I., & Frindell, A. 2025. Media over QUIC Transport. Internet-Draft, Internet Engineering Task Force, work in Progress. <https://datatracker.ietf.org/doc/draft-ietf-moq-transport/12/>
- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann. 2019. Mediapipe: A framework for perceiving and processing reality. IEEE CVPR.

³³ <https://www.spirit-project.eu/>

Drivable Group Tele-presence Demo

Nikolaos Zioulis, Nikolaos Kotarelas, Georgios Albanis, Spyridon Thermos,
and Anargyros Chatzitofis

Moverse P.C., Greece

Corresponding author: nick@moverse.ai

Keywords: Markerless Motion Capture, Gaussian Avatars, Tele-presence, Real-time

Introduction

Real-time human representations are the backbone of immersive tele-presence technology. Combined with free viewpoint navigation, they allow us to pursue natural communications beyond contemporary videoconferencing. Initially introduced as the concept of virtualized reality (Kanade et al., 1997) or telepresence (Fuchs et al., 1993) it involved complex and bulky systems that focused on the 3D digitization of dynamic scenes (Cheung et al., 2000; Yang et al., 2005). Yet the use of animated virtual characters was also being explored to allow for lighter-weight systems (Hirose et al., 1996). The latest advances in both presentation technology with head-mounted displays and higher quality 3D digitization as enabled by machine learning are pushing forward visual fidelity and drive the design of new tele-presence systems (Li et al., 2020; Lawrence et al., 2024; Tran et al., 2024). The core challenge of such systems is the time-varying network, that prevents 3D information from being transmitted at the necessary rates and fidelity.

At the same time, markerless motion capture has been evolving and it is now possible to use as the driving signal for animated characters. With real-time engines improving the visual fidelity of characters (e.g. Unreal Engine MetaHuman) this opens a new pathway for tele-presence that exploits low bitrate data to drive expressive user representations. With higher rate motion capture and richer skeleton topologies, the resulting animations can convey the nuances of body language. Gestures and body expressions constitute a very important communication cue as when we interact with others, we are continuously giving and receiving wordless signals (Sapir 1927). This externalization provides a means for multiple participants to synchronize their mind states and achieve effective communication (Licklider and Taylor, 1968).

Recent research has also been acknowledging the necessity for nonverbal communication. The usability of a remote instruction system increased when using a full body avatar (Wang et al., 2019), and participants interacting with highly expressive avatars felt more social presence and attraction and exhibited better task performance (Wu et al., 2021). Embodied 3D avatars have been shown to support nuanced nonverbal behavior and provide a high level of presence (Smith et al., 2018) with conversation patterns that are very similar to face-to-face interaction. Further, higher levels of face and body expressiveness were shown to be important for enhancing perceptions of realism and interaction quality

(Fraser et al., 2022), while harnessing full face and body motion capture can make XR interaction very similar to face-to-face interaction (Rogers et al., 2022).

In this work we describe a system to be demonstrated that will capture multiple users' body motion in real-time without wearing a suit or sensors. The captured body motion will be transmitted to a client that will showcase the use of different representations for the users, spanning from generic avatars to virtual characters and photorealistic reconstructions. The resulting bitstream is low even in highly interactive rates and shows new potential for developing future group communication technology. Same technology can also be used for (remote) virtual production.

System Overview

The markerless motion capture system is designed to capture multiple users across a volume in real-time and comprises a multi-view camera acquisition sub-system and a body fitting sub-system.

We use a distributed network of edge AI cameras that can process images on the camera, offloading computation to allow for real-time processing at high rates (e.g. 60Hz). The cameras are calibrated and synchronized and transmit only color images and processed metadata (e.g. AI model results like bounding boxes and keypoints). The data are further processed on a central computing node that solves the body motion of each actor in real-time. We use a parametric body model (Loper et al., 2023) as the solved representation as it allows us to consolidate various AI-based constraints in a unified manner. The system then streams the solved data that includes the shape and articulation parameters along with the global coordinate position and rotation of each user to downstream applications. Figure D 9 shows a technical diagram schematic of the capturing system and its streaming to downstream applications.

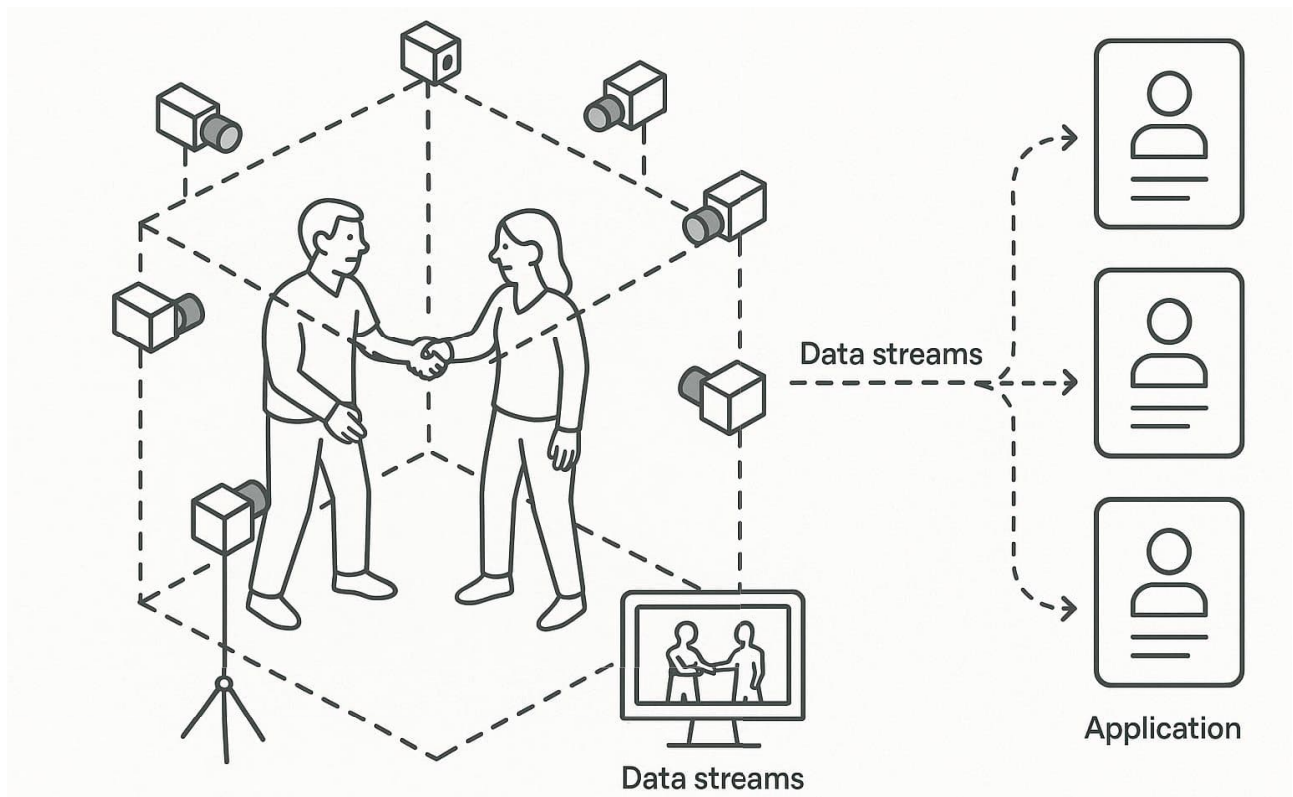


Figure D 9. A multiview camera markerless motion capture system streaming animation data to downstream applications that embody users in different representations.

Real-time Demonstration

Our live demonstration includes setting up the multiview camera markerless motion capture system and three different downstream applications to showcase the different opportunities for integrating live motion capture into various tele-presence systems. More specifically:

1. a generic avatar, as provided by the parametric body model that reconstructs a mesh-based representation of the users,
2. a virtual animated character, a choice that allows for personalization of the avatars, but needs to overcome the challenges of different topologies for both the mesh representation, as well as the skeleton definition, and,
3. a photorealistic representation, in the form of a reconstructed animated Gaussian Avatar captured by the same multi-view system.

The demonstration will be participatory, with the audience being allowed to use the system as its markerless nature makes it very easy and flexible. It will also demonstrate the system's robustness with respect to user appearance and sizes.

The downstream applications are implemented in Unity3D with the following showing exemplary 3D scenes with the live multi-user capturing animating the corresponding representations. Figure D 10 shows the 3 downstream applications that will be demonstrated live.



Figure D 10. The two images on the left show two different nodes that stream data to a shared virtual environment using generic avatars. The middle image shows streaming to virtual animated characters, while the right images show the live driving of Gaussian Avatar.

Acknowledgements

Funded by the European Union (CORTEX2, 101070192). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the EU. Neither the EU nor the granting authority can be held responsible for them.

References

- Cheung, G.K., Kanade, T., Bouguet, J.Y., & Holler, M., 2000. A real time system for robust 3D voxel reconstruction of human motions. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662) (Vol. 2, pp. 714-720). IEEE.
- Fraser, A.D., Branson, I., Hollett, R.C., Speelman, C.P. & Rogers, S.L., 2022. Expressiveness of real-time motion captured avatars influences perceived animation realism and perceived quality of social interaction in virtual reality. *Frontiers in Virtual Reality*, 3, p.981400. <https://doi.org/10.3389/frvir.2022.981400>
- Fuchs, H., & Neumann, U., 1993, October. A vision of telepresence for medical consultation and other applications. In *Sixth International Symposium of Robotics Research* (pp. 555-571).
- Hirose, M., Deffaux, G., & Nakagaki, Y., 1996. Development of an effective motion capture system based on data fusion and minimal use of sensors. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (pp. 117-123).
- Kanade, T., Rander, P., & Narayanan, P.J., 1997. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE multimedia*, 4(1), pp.34-47.
- Lawrence, J., Overbeck, R., Prives, T., Fortes, T., Roth, N., & Newman, B. 2024. Project starline: A high-fidelity telepresence system. In *ACM SIGGRAPH 2024 Emerging Technologies* (pp. 1-2).
- Li, R., Xiu, Y., Saito, S., Huang, Z., Olszewski, K., & Li, H., 2020. Monocular real-time volumetric performance capture. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16* (pp. 49-67). Springer International Publishing.
- Licklider, J.C. , & Taylor, R.W., 1968. The computer as a communication device. *Science and technology*, 76(2), pp.1-3.
- Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M.J. 2023. SMPL: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (pp. 851-866). <https://doi.org/10.1145/3596711.3596800>
- Rogers, S.L., Broadbent, R., Brown, J., Fraser, A., & Speelman, C.P. 2022. Realistic motion avatars are the future for social interaction in virtual reality. *Frontiers in Virtual Reality*, 2, p.750729. <https://doi.org/10.3389/frvir.2021.750729>
- Sapir, E., 1927. Language as a form of human behavior. *The English Journal*, 16(6), pp.421-433. <https://doi.org/10.2307/803262>
- Smith, H.J., & Neff, M. 2018. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-12).
- Tran, P., Zakharov, E., Ho, L.N., Hu, L., Karmanov, A., Agarwal, A., Goldwhite, M., Venegas, A.B., Tran, A.T., & Li, H., 2024. VOODOO XP: Expressive One-Shot Head Reenactment for VR Telepresence. *arXiv preprint arXiv:2405.16204*. <https://doi.org/10.48550/arXiv.2405.16204>
- Wang, T.Y., Sato, Y., Otsuki, M., Kuzuoka, H. & Suzuki, Y., 2019. Effect of full body avatar in augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 1221-1222). IEEE. <https://doi.org/10.1109/VR.2019.8798044>
- Wu, Y., Wang, Y., Jung, S., Hoermann, S., & Lindeman, R.W. 2021. Using a fully expressive avatar to collaborate in virtual reality: Evaluation of task performance, presence, and attraction. *Frontiers in Virtual Reality*, 2, p.641296. <https://doi.org/10.3389/frvir.2021.641296>
- Yang, Z., Nahrstedt, K., Cui, Y., Yu, B., Liang, J., Jung, S.H., & Bajscy, R., 2005. TEEVE: The next generation architecture for tele-immersive environments. In *Seventh IEEE International Symposium on Multimedia (ISM'05)* (pp. 8-pp). IEEE.

Beyond Classical Cooperation: Reimagining Business Meetings through VR Conferencing

Yazid Benazzouz¹, Alexis Lombard², Jean-Pierre Lorré¹, and Gaël de Chalendar²

¹*Linagora GSO, France*

²*Paris-Saclay University, CEA, France*

Corresponding author: ybenazzouz@linagora.com

Keywords: Virtual Reality (VR), Business Meetings, cooperation, immersive conferencing system

Introduction

This demo explores the use of Immersive conferencing in Virtual Reality (VR), designed to connect both virtual and non-virtual participants in a shared collaborative space to extend participation to the widest possible audience. It focuses primarily on business meeting scenarios, while remaining adaptable to other use cases. The aim is to enhance immersive collaboration, fostering stronger engagement and a greater sense of presence, especially valuable in today's remote working environments. In addition, this approach enables interactive co-design on virtual objects and overcomes the limitations of traditional conferencing tools. The solution also brings significant environmental benefits by minimizing travel requirements, reducing carbon emissions and environmental impact. Built on a modular architecture, the application offers scalability and supports the addition of new functionalities. A general view of the immersive room is shown in the following Figure D 11

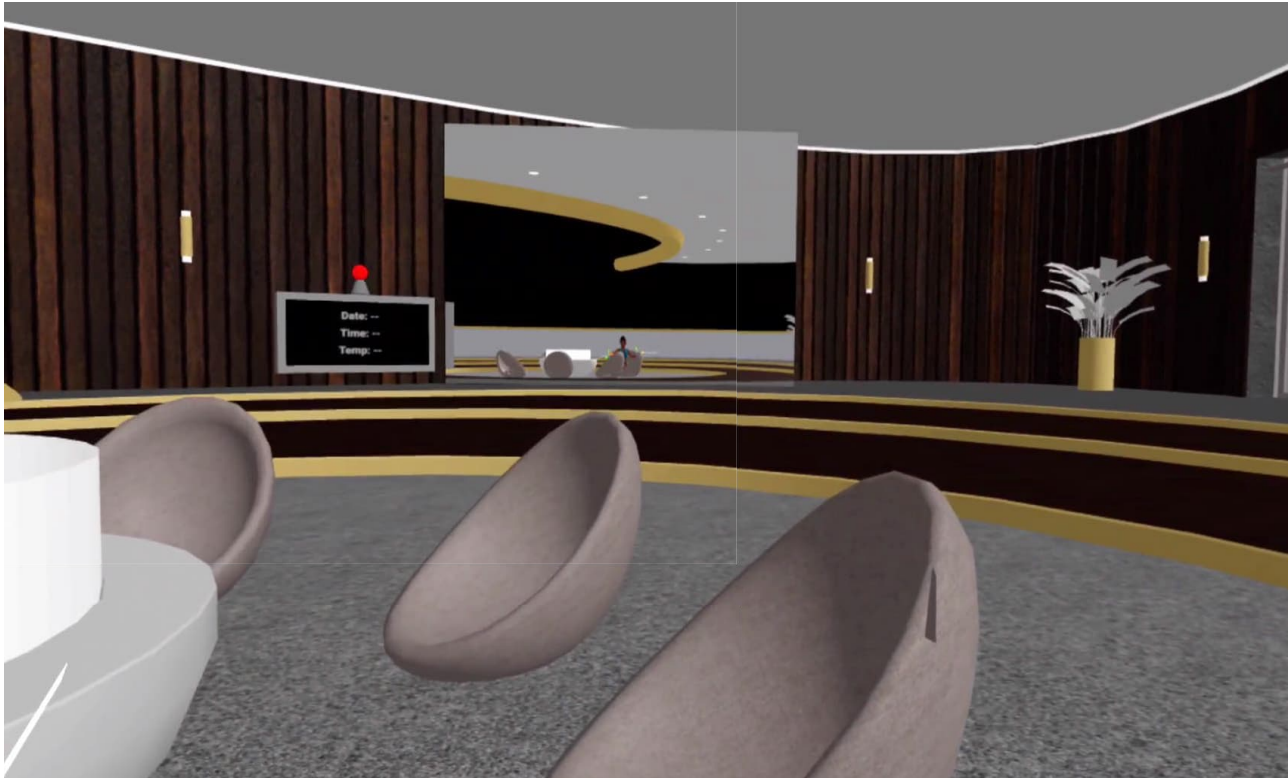


Figure D 11. Business Meeting Immersive room

Design principles

The design of our Virtual Reality (VR) conferencing application for business meetings is founded on a set of complementary, human-centric principles. Together, they build an ecosystem that integrates immersive technology, intelligent features, and ethical AI, summarized in the following key points:

- Hybrid Inclusivity: Enables both VR and non-VR participants to connect.
- Supports various devices and setups to ensure broad accessibility through WebXR.
- Leverages spatial audio, 3D avatars, and a shared virtual environment to create a realistic sense of "being there."
- Functionality for Dynamic Co-Creation: Enables collaboration on 3D models and shared virtual assets in real time.
- Enable Intelligent Assistance: Embeds CoVA, a virtual assistant to actively listen and deliver relevant data, reminders, contextual prompts, and execute actions. Facilitates adoption of voice shortcuts (e.g., "Mute me," "Exit VR") and AI features, including intent recognition for user-defined interactions and a retrieval-augmented generation module meant to broaden the scope of conversation by leveraging shared documents and domain-specific knowledge graphs.
- Integrate IoT Awareness: Brings live data from real-world sensors or machinery directly into the VR workspace.
- Automate Post-Meeting Summaries: Generates summaries from recorded conversations, transforming audio into structured and shareable content.
- Scalable, Modular Architecture: Flexible system design allows new tools and functionalities to be added.

Integrating Trust, Safety, and Accessibility

In developing our virtual reality application for business meetings, we placed a strong emphasis on transparency, user safety, and inclusive design. To support explainability, each AI-generated meeting summary includes an “AI summary disclaimer,” ensuring users are aware of its automated nature. We also addressed the risk of cybersickness by implementing a session timer that prompts users with gentle reminders—such as “Take a break?”—after 30 minutes of continuous use.

Accessibility was a core design principle. We integrated voice shortcuts for common actions like “Mute me” and “Exit VR,” making the interface more intuitive and usable for individuals with disabilities. On the data governance front, we established clear policies: transcripts are processed by default, but storage is disabled unless users actively choose to “save this meeting.” Additionally, sensitive information such as phone numbers and email addresses is automatically redacted from transcripts using pattern-based filters.

To ensure reliability and user comfort, the application’s user guide describes fallback mechanisms that recommend the use of “Stationary” mode during business meetings. This minimizes motion-related discomfort and reduces physical risks, particularly in constrained environments. These measures collectively reflect our commitment to building a secure, inclusive, and user-centered VR collaboration platform.

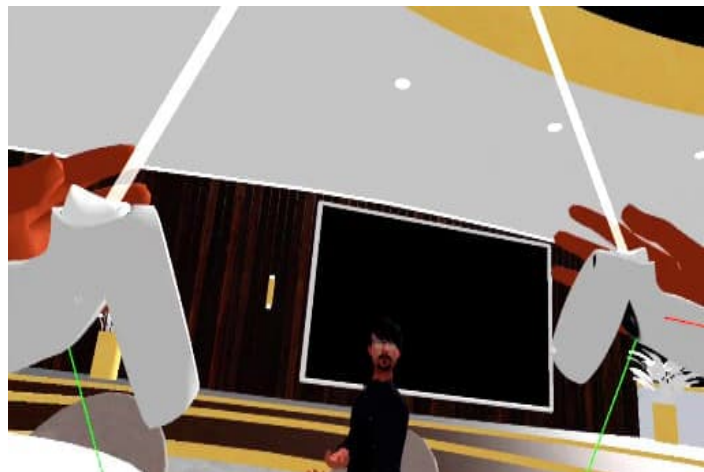
Demo Journey and Insights

The journey begins as the organizer creates a virtual “bubble,” a spatial meeting room designed to host the collaborative session. He defines the participant list and sends out invitations. Participants joining via a VR headset can specify remote IoT devices, such as environmental sensors, to integrate into their individual virtual environments, allowing, for example, real-time access to critical physical-world data.

Before entering the space, each user personalizes their presence by selecting or creating an avatar, either by choosing from a library of customizable options or generating a 3D likeness using their own facial image. To ensure full access to immersive tools, the session must be initiated from VR by the organizer rather than through a standard web interface, setting a clear XR-first goal that encourages meaningful engagement.

Path 1: Immersive VR Participation: Participants with VR headsets enter the bubble and experience:

- Realistic avatars with spatial audio
- Gesture-based interactions and intuitive navigation
- Access to virtual visualization panel, shared documents, and 3D virtual objects

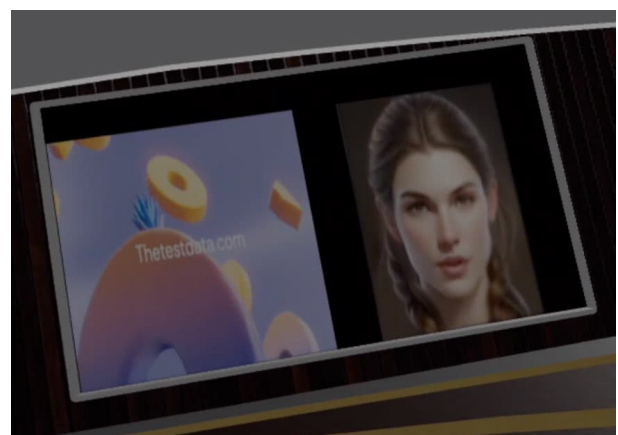
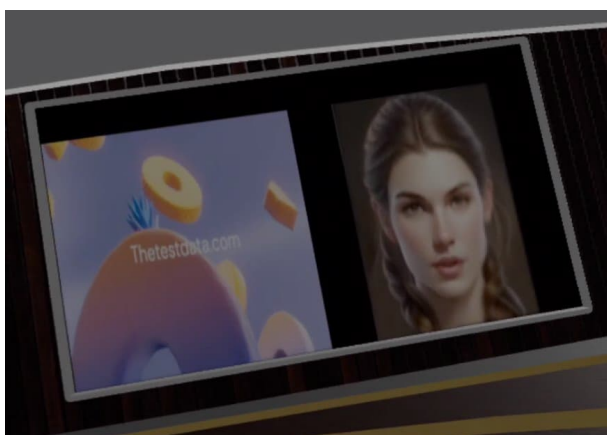


Path 2: Real-Time Awareness & Assistance: Once inside the bubble, users gain access to advanced interaction tools:

- **Engagement Indicators:** At any time, users can check their personal engagement metrics, such as, focus level or interaction frequency via the left controller menu. This helps them stay aware of their participation and adjust accordingly.
- **IoT Integration:** Users can activate real-world IoT devices, such as temperature sensors, to monitor critical environmental data while in VR. This is especially useful in industrial or safety-sensitive contexts, where real-time awareness of physical conditions is essential.
- **Virtual Assistant: CoVA,** A built-in AI assistant is available to answer subject-specific questions, provide contextual help, or guide users through tools and features, making the experience more intuitive and productive. It combines user control with contextual adaptability. User-defined intents enable deterministic responses for predictable and critical interactions, as well as intent-triggered actions (e.g., document summary fetching, XR control, source attribution, web search), with minimal configuration effort required from the user. Out-of-scope queries are addressed by dynamically generating answers grounded in session-specific documents or domain-specific structured graphs. The assistant can be specialized in real time for each conference group through a dedicated configuration interface, allowing users to shape a unique agent suited to their present objectives and working context.

Path 3: Hybrid (2D) Participation: Participants without access to VR join through web or mobile interfaces. They experience:

- Audio and video feeds from within the virtual space
- Shared content via simplified layouts
- Basic collaboration tools like chat, document viewing, and voice
- The platform ensures real-time synchronization between immersive and non-immersive users, enabling equitable participation across formats.
- VCAA (Video Conferencing Alternative Appearance) avatar to appear as a 2D representation during the session.



Path 4: Collaborative work:

- VR users can co-interact with shared virtual objects, Rotate, annotate, and manipulate 3D objects in real time
- Sketch ideas, brainstorm, and iterate together through audio and avatar movements.
- Such interactions are either limited or impossible in traditional video conferencing platforms. This pilot addresses that gap by offering modular, task-specific immersive tools.

Conclusion

This demonstration shows the potential to move towards immersive remote collaboration systems accessible to the general public. It opens the way to new forms of shared experiences. By going beyond the limits of classical cooperation, this virtual reality application offers a promising model for democratizing immersive collaborative environments. It foreshadows a future where professional meetings, family moments, and shopping journeys are reimagined within engaging, intuitive, and deeply human digital spaces.

Acknowledgment

This work was developed as part of the CORTEX² project, which has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101070192.

References

- Lombard, A., Benazzouz, Y., Castillo López, G., de Chalendar, G., Lorré, J.P. 2023. CoVA : A Virtual Assistant for Virtual Reality Conferencing. EuroXR 2024: Proceedings of the 21st EuroXR International Conference. <https://doi.org/10.32040/2242-122X.2024.T432>
- Pagani, A., Minaskan Karabid, N., Javanmardi, A., Xie, Y., Rivier, S., Bailleau, V., Helbert, E., ... , & Garcia-Palacios. A. 2023. CORTEX² Extended collaborative telepresence for future work and education. LMDE, Research Colloquium. Leading & Managing in the Digital Era (LMDE-2023), June 19-20, Athens, Greece.
- Javanmardi, A., Pagani, A., & Stricker, D. 2024. G3FA: Geometry-guided GAN for Face Animation, 35th British Machine Vision Conference 2024, {BMVC}, Glasgow, UK, November 25-28, 2024

Demonstrating V-PCC Bitstream Transcoding

Michael Rudolph, Finn Schnier, and Amr Rizk

Leibniz University Hannover, Germany

Corresponding author: michael.rudolph@ikt.uni-hannover.de

Keywords: Volumetric Video, V-PCC, Transcoding, Point Cloud Streaming

Motivation

Moving from 2D videos to volumetric representations of objects or humans allows users to freely navigate scenes during immersive media consumption. An established format is given by point clouds, that offer a high degree of flexibility for the 3D object as well as low processing overhead, as they can be directly acquired from sensors like RGB-D cameras or LiDAR scanners. However, as most multimedia applications are distributed over a communication network, compression is required to handle the required high data rates that arise from the high sampling density of visually appealing point clouds.

Targeting the compression of point cloud sequences, the MPEG standardized the Video-based Point Cloud Compression Codec (V-PCC). V-PCC leverages the high compression efficiency of 2D video codecs through decomposing each point cloud frame into 2D patches, which are sorted into a sequence of video frames (Schwarz et al., 2018). However, deriving coherent patches in the point cloud requires clustering points into groups according to their normal direction, a process that contributes to the overall high latency of the currently available test model implementation.

Our work envisions an adaptive bitrate streaming scenario, where clients select from a set of pre-encoded representations of a stream segment to achieve high user quality under bandwidth constraints. For this, content is typically encoded according to a set of predefined coding configurations, each offering a different trade-off between rate and quality of the segment. However, for large scale distribution, it is desirable to allow on-the-fly transcoding of the already encoded segments, allowing to dynamically offer new configurations.

In this demonstration, we build on the framework of RABBIT (Rudolph et al., 2023), a method that directly transcodes the underlying video sub-streams with video codecs, bypassing the need for computationally expensive reconstruction of the 3D representations during transcoding. Thus, RABBIT allows on-the-fly transcoding of V-PCC encoded bitstreams with similar quality to the full transcoding using the decoder-encoder stack of V-PCC. In the project, we design a deployable implementation of RABBIT in Python and evaluate its behavior during on-the-fly transcoding when streaming to multiple clients with different bandwidth requirements, which allows to derive resource requirements and scalability metrics of the system in realistic use cases.

Related Work

Building on HTTP adaptive streaming (HAS), the de-facto standard for large scale video streaming, extensions of the Dynamic Adaptive Streaming over HTTP (DASH) protocol to accommodate point cloud data have been proposed in the work by Hosseini et al., 2018. Since users can freely move inside a scene, further work investigated rate allocation schemes that maximize the quality perceived in the user view port by incorporating the user location and viewing direction into the quality adaptation mechanism (Van der Hooft et al., 2019). Further decomposition of a point cloud objects into view-dependent patches (Rudolph et al., 2022) or cuboids (Subramanyam et al., 2020) allows for more fine-grained rate allocation.

Due to the high latency of the currently available test model³⁴, the media on the media server needs to be pre-encoded to a fixed set of representations, thus limiting users to adhere to the pre-defined bitrate ladder. Recently, the first real-time implementations of the V-PCC encoder offer reduced latency (Fréneau et al., 2025) for compressing segments, but it still requires full decoding to 3D space when aiming to transcode an already compressed representation.

Figure D 12. Schematic Overview of our Demo. A point cloud sequence is split into segments and pre-encoded using V-PCC (offline). During streaming, we demonstrate the flexibility of the transcoder by allowing a user to adjust the rate control parameters for attributes (attQP) and geometry (geoQP) to receive the transcoded representation.

Demo Description

The goal of this demo is to showcase the degrees of flexibility achievable with our transcoder implementation for V-PCC encoded content. For this, a pre-encoded point cloud video will be streamed to a receiver device which runs a streaming client to decode and visualize the incoming point cloud. At the receiver, the user will be able to individually select a configuration, i.e. the quality of the geometry and attribute sub-streams, which will be signaled to the sender device for dynamic adjustment and transcoding. The setup is illustrated in Figure D 12.

We implement our transcoder utility in Python. First, the bitstream is parsed and the geometry, attribute and occupancy video sub-streams are extracted. Each sub-stream is then transcoded

³⁴ <https://github.com/MPEGGroup/mpeg-pcc-tmc2>

individually through a video codec. For the video sub-streams of geometry and attribute, rate-control is performed by adjusting the coding configuration for the video encoder according to the desired configuration of the 3D representation. The occupancy video sub-stream, which is coded in lossless manner, can be optionally downsampled to further adjust the quality of the point cloud. After transcoding, we replace the original bitstreams through the transcoded video sub-streams to receive the transcoded V3C bitstream.

The receiver decodes transcoded representation using a real-time decoder implementation³⁵ and visualized in a renderer using the three.js framework.

Conclusion and Future Work

This demo aims at demonstrating the level of flexibility and low latency achievable through on-the-fly transcoding. While we opted here for 1:1 communication between the sender and receiver to showcase the capabilities of the transcoder, we will evaluate the scalability of a DASH-based transcoding system with multiple users that exhibit heterogeneous bandwidth conditions, resulting in parallelization of transcoding operations. We aim at complementing the SPIRIT platform with a framework for volumetric video transcoding and streaming of V-PCC encoded content for future applications aiming at volumetric media distribution and allowing researchers to further explore topics in the field of immersive multimedia streaming.

Acknowledgments



**Funded by
the European Union**

Funded by the European Union (SPIRIT, 101070672). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. The SPIRIT project has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

References

- Fréneau, L., Gautier, G., Mercat, A., & Vanne, J. 2025. uvgVPCCenc: Practical open-source encoder for fast V-PCC compression. In *Proceedings of the 16th ACM Multimedia Systems Conference* (pp. 235-241). ACM. <https://doi.org/10.1145/3712676.3718336>
- Hosseini, M., & Timmerer, C. 2018. Dynamic adaptive point cloud streaming. In *Proceedings of the 23rd Packet Video Workshop* (pp. 25-30). ACM. <https://doi.org/10.1145/3210424.3210429>
- Rudolph, M., & Rizk, A. 2022. View-adaptive streaming of point cloud scenes through combined decomposition and video-based coding. In *Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis* (pp. 41-49). ACM. <https://doi.org/10.1145/3552457.3555732>
- Rudolph, M., Schneegass, S., & Rizk, A. 2023. Rabbit: Live transcoding of V-PCC point cloud streams. In *Proceedings of the 14th Conference on ACM Multimedia Systems* (pp. 97-107). ACM. <https://doi.org/10.1145/3587819.3590978>

³⁵ <https://github.com/benclmnt/tmc2-rs/tree/main>

- Schwarz, S., Preda, M., Baroncini, V., Budagavi, M., Cesar, P., Chou, P. A., ... & Zakharchenko, V. 2018. Emerging MPEG standards for point cloud compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1), 133-148. IEEE. <https://doi.org/10.1109/JETCAS.2018.2885981>
- Subramanyam, S., Viola, I., Hanjalic, A., & Cesar, P. 2020. User centered adaptive streaming of dynamic point clouds with low complexity tiling. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 3669-3677). ACM. <https://doi.org/10.1145/3394171.3413535>
- Van der Hooft, J., Wauters, T., De Turck, F., Timmerer, C., & Hellwagner, H. 2019. Towards 6DoF HTTP adaptive streaming through point cloud compression. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 2405-2413). ACM. <https://doi.org/10.1145/3343031.3350917>

EMPATHI: Emotion-Responsive Avatars for Personalized Cultural Heritage Interaction

Annalisa Mombelli^{1}, Thomas Agostinelli², Andrea Generosi², Marco Luciani³, Giovanni Valente¹,
Francesco Rondani¹, Maura Mengoni², Hillary Pedrizzi¹, and Renato Marchetti¹*

¹RE:LAB srl, Italy

²UnivPM - Università Politecnica delle Marche, Italy

³Emoj s.r.l, Italy

*Corresponding author: annalisa.mombelli@re-lab.it

Keywords: XR, VR, emotion recognition, cultural heritage, real time avatar, immersive storytelling, adaptive interaction

Introduction and motivation

Cultural heritage institutions increasingly seek technologies that create meaningful connections between visitors and historical content. Traditional static presentations often fail to engage diverse audiences effectively, especially younger generations accustomed to interactive digital experiences. EMPATHI addresses this challenge by developing an emotionally intelligent avatar system that adapts to visitor responses in real-time. The SPIRIT project provides an ideal foundation, offering a comprehensive platform for real-time immersive telepresence with modular architecture designed to support third-party enhancements. EMPATHI leverages SPIRIT's Real-Time Avatar Animation use case, extending its capabilities through emotion recognition to create more natural and engaging interactions. This alignment with SPIRIT's vision of scalable, innovative telepresence solutions positions EMPATHI as an exemplar of the platform's potential for transformative applications. The project centers on a photorealistic 3D avatar of Ludovico Ariosto, the renowned Renaissance writer, who guides visitors through the "Atelier Furiose Interazioni" exhibit at Palazzo del Mauriziano in Reggio Emilia, Italy. This technological showcase, launched by TICHE (Italian Technological Cluster for Cultural Heritage), already features six interactive stations exploring Ariosto's Orlando Furioso through innovative digital technologies. EMPATHI enhances this experience by adding an emotional dimension to visitor-avatar interactions.

System architecture and innovation

The avatar's behavior is dynamically adapted through a bidirectional emotional feedback system. A CNN-based module analyzes facial expressions of the producer (avatar controller), while visitors provide simple feedback using a GUI interface. Based on this emotional input, the avatar adjusts tone, storytelling, and facial expressions in real time, enabling adaptive and inclusive engagement, especially for educational purposes. The EMPATHI system introduces several key innovations to the SPIRIT platform:

- I. **Emotion Recognition Module:** At the core of EMPATHI lies a CNN-based emotion recognition system utilizing the VGG13 architecture (Talipu et al., 2019). The module processes video streams captured at 10-15 FPS, analysing facial expressions to classify emotions according to Ekman's (Ekman & Friesen, 1978) six basic categories (joy, surprise, sadness, anger, fear, disgust) plus neutral. The system also computes engagement scores (0-100) and valence values (-100 to 100) to provide nuanced emotional assessment.
- II. **Bidirectional Feedback System and Simplified Emotional Feedback (SEF) GUI:** EMPATHI implements a novel two-way emotional feedback loop through carefully designed interfaces:



Figure D 13. UX/UI moodboard of Producer and Consumer SEF GUI.

- **Producer Side:** The avatar controller's emotions are captured and processed in real-time, influencing the avatar's facial expressions and demeanour. **Emotional GUI (Producer Feedback):** the system visualizes the producer's emotional state through an innovative overlay system. As shown in Figure D 13, emotions are represented through dynamic auras surrounding the avatar that change colour based on detected emotions (Thorstenson et al., 2018); on-screen emoticons that appear contextually; subtle particle effects that enhance the emotional atmosphere; real-time emotion indicators integrated seamlessly into the virtual environment.
- **Consumer Side:** Visitors interact through a SEF GUI, providing intuitive like/dislike responses that inform the controller about audience engagement. **Consumer SEF GUI:** designed following human-centered design principles, this interface enables visitors to provide immediate emotional feedback through large, touch-friendly buttons optimized for both tablet and XR interface; intuitive iconography using universally recognized symbols; color-coded responses following universal associations linked to emotions; minimal cognitive load to maintain immersion.

III Historical Avatar Reconstruction: Ludovico Ariosto's 3D avatar (Figure D 14.) was meticulously crafted based on extensive iconographic research (Ruffini, 2011), primarily referencing Titian's masterful portraits—the "Ritratto dell'Ariosto" (c. 1515, Indianapolis Museum of Art) and "Ritratto di uomo detto l'Ariosto" (c. 1510, National Gallery London) as well as the 1532 Orlando Furioso woodcut engraving to understand his contemporary public image. The 3D model of Ariosto was created based on reference images of the subject. His body structure and clothing were modeled in Blender, then retopologized and optimized in terms of polygon count to ensure the best graphical rendering when imported into Unity3D.



Figure D 14. Ludovico Ariosto

Technical implementation

EMPATHI extends SPIRIT's Kubernetes-orchestrated infrastructure with three specialized Docker containers operating in isolation for enhanced security and maintainability (Figure D 15). The *Emotion Recognition Service* processes video streams through CNN-based analysis to classify Ekman's emotions with temporal filtering. The *Audio Animation Server* extracts visemes (i.e., minimal visual units that represents a specific position of the mouth and face during the pronunciation of a sound or group of sounds) for lip-sync generation, while the *Streaming Server* synthesizes emotional and audio data into real-time 3D animations with complex blend shape interpolations. The *Communication Pipeline* orchestrates bidirectional data flow between Producer and Consumer devices via LAN/5G networks, maintaining sub-100ms latency. Producers stream 720p video and audio to the central server via WebRTC, while Consumers relay spatial positioning data. Server-side parallel processing routes video frames (10-15 FPS) to emotion recognition and audio to animation services. The central orchestrator synthesizes processed metadata into enhanced streams incorporating emotional animations and lip-sync data. Both Producer and Consumer interfaces receive synchronized updates. Producers monitor audience emotions via dashboard for narrative adjustments, while Consumers experience the emotionally responsive avatar seamlessly. This architecture demonstrates successful AI service integration within SPIRIT's existing streaming infrastructure without performance compromise.

Methodology for user testing

The EMPATHI system will undergo comparative evaluation through two experimental conditions at Palazzo del Mauriziano. Condition A presents traditional linear storytelling via live narrator. Condition B features the emotionally adaptive Ariosto avatar, accessible through VR headsets (teens/adults) or tablets (children from age 5). Each 30-minute single-user session targets participants across four age cohorts (5-6 preschool, 7-12 children, 13-17 adolescents, 18-60 adults). Measured outcomes include engagement levels (self-reported Likert scales and behavioral observation), comprehension (age-appropriate quiz assessment), emotional response (self-reported), and interpretive plurality (coded post-experience reflections). Additionally, both end-user experience (interface usability, emotional immersion, personalization satisfaction) and producer experience (narrative control, responsiveness via dashboard monitoring) are evaluated through structured post-session surveys, providing comprehensive insights into the system's effectiveness in enhancing cultural heritage engagement across diverse age groups.

Expected impact and applications

EMPATHI's contributions significantly extend SPIRIT's capabilities and market reach by introducing emotion recognition modules, enhanced avatar animation, and inclusive interface patterns to the platform's component library. The project validates SPIRIT's extensibility through real-world deployment, opening new vertical applications in cultural heritage, education, and tourism. Beyond technical contributions, EMPATHI establishes methodologies for creating historically accurate yet emotionally responsive avatars, frameworks for multi-generational museum accessibility, and personalized guide systems that adapt to visitor engagement, bringing historical figures to life for immersive learning. Post-project, EMPATHI modules will be contributed to SPIRIT's open component library, with comprehensive integration guidelines supporting future emotional computing applications. This symbiotic relationship demonstrates how domain-specific innovations can enrich the broader telepresence ecosystem, fostering continued advancement in immersive technologies across education, heritage preservation, and inclusive digital experiences.

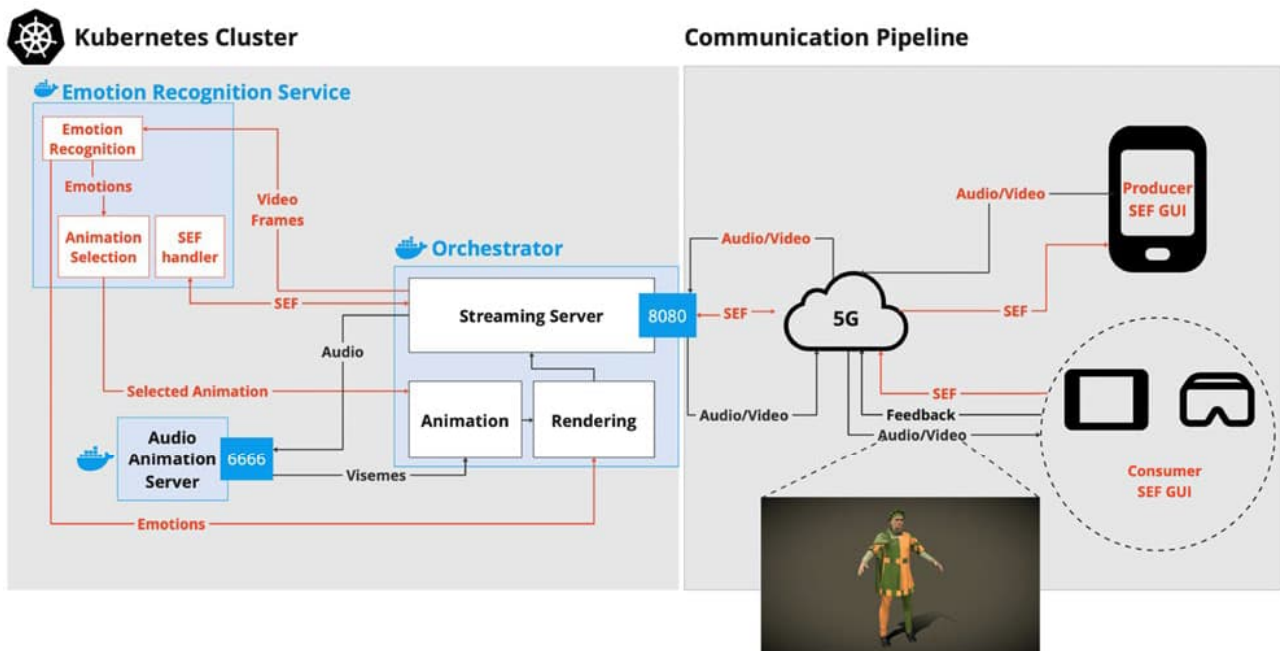


Figure D 15. The high-level proposed pipeline

References

- Ekman, P., & Friesen, W. V. 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press.
- Ruffini, M. 2011. *L'iconografia di Ludovico Ariosto*. Leo S. Olschki.
- Talipu, A., Generosi, A., Mengoni, M., & Giraldi, L. 2019. Evaluation of deep convolutional neural network architectures for emotion recognition in the wild. IEEE ISCT, 25–27.
- Thorstenson, C.A., Elliot, A.J., Pazda, A.D., Perrett, D.I., & Xiao, D. 2018. Emotion-color associations in the context of the face. *Emotion*. 2018 Oct;18(7):1032-1042. <https://doi.org/10.1037/emo0000358>. Epub 2017 Nov 27. PMID: 29172620.

Real-Time Avatar-Guided VR Tours: Integrating Live Telepresence in Cultural Heritage Applications

Eirini Kolokytha, Aimilia Bantouna, Charalampos Korovesis, Maria Kounalaki,
and Eleni Giannopoulou

WINGS ICT Solutions, Greece

Corresponding authors: ikolokytha@wings-ict-solutions.eu

Keywords: Virtual Reality, Telepresence, Cultural Heritage, Real-time Communication, SPIRIT Platform, Tourism Technology

Introduction

Traditional virtual heritage tours rely on static content and pre-recorded narratives, which limit user engagement and fail to provide the personalized experience that live human guides offer. Furthermore, they lack adaptability when content needs to be changed. In such cases, when new content is introduced or existing content needs to be updated, new narratives need to be recorded or updated which is costly and takes time. This paper presents an innovative approach to cultural heritage tourism immersion through the integration of real-time avatar-guided virtual reality experiences, developed as part of the ANNETE project within the EU SPIRIT platform initiative (SPIRIT Consortium, 2024). The proposed work aims to transform static 360° panoramic tours into dynamic, interactive experiences where live human guides appear as realistic avatars within VR environments, enabling real-time communication and personalized cultural storytelling. The implementation focuses on the Corinth Canal heritage site in Greece, where visitors can explore immersive 360° environments (Vincke et al., 2019) while interacting with live guides who provide contextual information, answer questions, and adapt the tour experience based on visitor interests. This approach addresses the growing demand for accessible cultural heritage experiences while maintaining the authenticity and personal touch that human expertise provides.

System Architecture and Implementation

Technical Framework

The proposed application utilizes a three-tier architecture (Figure 1) optimized for low-latency, high-quality VR experiences:

1. **Client Layer:** The client application, accessible via mobile devices and VR headsets, provides an intuitive interface for exploring 360-degree panoramic views of the Corinth Canal. Built using the Unity game engine, the application ensures broad compatibility across devices while maintaining high visual fidelity. The interface includes a. Multi-language support (English and

Greek), b. Spatial audio capabilities for immersive narration, c. Interactive hotspots for accessing points across the Corinth Canal and d. Real-time avatar visualization of the guide.

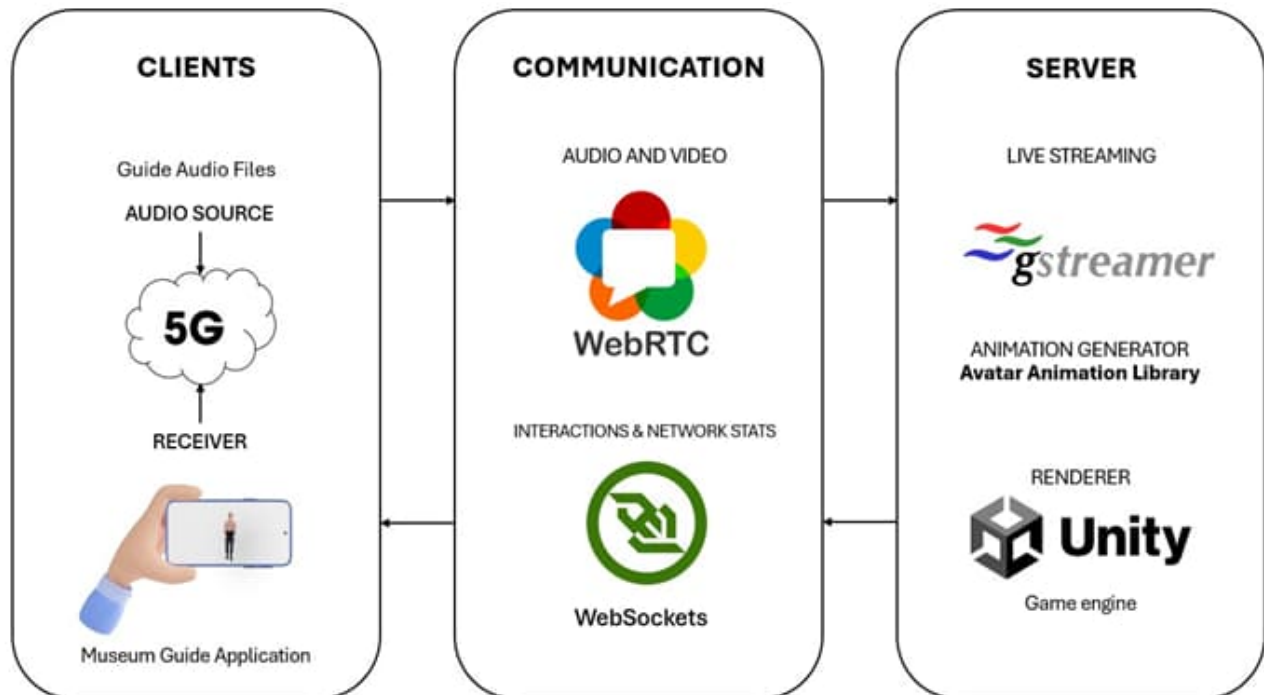


Figure D 16. System Architecture Schema

- Communication Layer:** A hybrid communication approach has been developed using: a. WebRTC for ultra-low latency audio and video streaming between guides and visitors and b. WebSockets for real-time interaction data and network statistics.
- Server Layer:** The backend infrastructure, enhanced by the SPIRIT platform, provides the following capabilities, namely a. GStreamer-based live streaming with adaptive bitrate control, b. Unity-powered rendering engine for avatar animation, c. Avatar Animation Library integration for realistic guide representations (Thies et al., 2020) and d. Dynamic resource allocation through the SPIRIT Diktyo Scheduler.

SPIRIT Platform Integration

As part of the ANNETE project (Enhancing Telepresence: Real-Time Avatars in Tourism and Culture), the proposed application leverages on several key SPIRIT components:

- Avatar Component for Real-Time Rendering:** The Avatar component is essential for delivering real-time animation and streaming of realistic avatars, which are core to our immersive telepresence solution. This component ensures accurate, high-fidelity rendering with minimal latency, enabling users to experience lifelike, interactive avatars during collaborative sessions. To meet the project's requirements, the platform's edge computing capabilities, equipped with high-performance GPUs (e.g., NVIDIA RTX A6000), are utilized for rendering avatars efficiently. This feature is crucial for maintaining smooth, low-latency interactions while ensuring visual consistency across multiple testbed sites.
- Quality of Experience (QoE) Tools for Performance Monitoring:** To ensure optimal user experience, the platform's Quality of Experience (QoE) tools will be utilised, available for both online and offline monitoring. These tools will allow us to measure and fine-tune critical

performance metrics such as latency, frame synchronization, visual quality, and network performance. The integration of QoE tools will enable iterative optimization of the proposed solution, ensuring that real-time communication and holographic streaming meet the highest standards of reliability and user satisfaction. The availability of these tools across the SPIRIT testbeds ensures consistent evaluation under varying conditions.

3. **Security Components for Data Protection:** Given the sensitive nature of holographic and avatar-based telepresence data, integrating the platform's Security components is essential. Trusted Execution Environments (TEEs) and end-to-end encryption will be employed to ensure secure data transmission and processing across testbeds. These components will safeguard user privacy and protect sensitive information from unauthorized access or breaches. The robust security infrastructure provided by SPIRIT ensures compliance with data protection standards, enhancing the reliability and trustworthiness of our solution.

Current Development Status and Expected Results

The ANNETE project is currently in the implementation phase (March-October 2025), with the VR tour application and avatar integration system under active development. The 360° panoramic content of the Corinth Canal (**Figure D 17**) has been captured and is being integrated into the VR framework. Avatar animation systems are being developed and tested, with initial prototypes showing promising results for real-time rendering and synchronization.



Figure D 17. Capture of the 360° panoramic VR app of the Corinth Canal.

The system is designed to support multiple concurrent users (Viola et al., 2023), enabling group tours and collaborative exploration experiences. Next, in collaboration with the respective testbed teams, the configuration of the involved SPIRIT testbeds will be decided and implemented.

Expected outcomes include:

- Successful deployment and execution of the proposed solution on SPIRIT testbeds;

- Data logs for performance metrics and real-time QoE monitoring; and

Impact and Future Directions

As part of WINGS' wi.LIVE Culture / Education product development strategy, the proposed application represents a significant evolution of the wi.LIVE culture platform capabilities. The integration with SPIRIT components will enable quicker development of enhanced immersive reality experiences, strengthening WINGS' position as a pioneer in digital culture innovation. The successful validation through ANNETE will expand WINGS' market presence in the telepresence tourism sector, attracting new clients seeking advanced digital solutions. Furthermore, this aligns with WINGS' mission to enhance citizen engagement and accessibility to Greek culture and heritage through digital tools, particularly focusing on overcoming barriers of geography, language, and mobility while engaging younger audiences through innovative cultural presentation methods.

Future work will focus on expanding the system to support additional heritage sites, incorporating advanced AI-driven personalization features, and developing specialized training modules for cultural heritage guides adapting to avatar-based interactions. The project's outcomes will contribute to the broader understanding of telepresence applications (Fuchs et al., 2014) in cultural tourism and provide practical insights for deploying similar systems at scale.

References

- SPIRIT Consortium. 2024. "SPIRIT Platform (Second Version)," Project "Scalable Platform for Innovations on Real-time Immersive Telepresence", EC grant agreement 101070672.
- Vincke, S., de Lima Hernandez, R., Bassier, M., & Vergauwen, M. 2019. Immersive visualisation of construction site point cloud data, meshes and BIM models in a VR environment using a gaming engine. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 77-83. <https://doi.org/10.5194/isprs-archives-XLII-5-W2-77-2019>
- Thies, J., Elgharib, M., Tewari, A., Theobalt, C., & Nießner, M. 2020. Neural voice puppetry: Audio-driven facial reenactment. In *European conference on computer vision* (pp. 716-731). Cham: Springer International Publishing.
- Viola, I., Jansen, J., Subramanyam, S., Reimat, I., & Cesar, P. 2023. Vr2gather: A collaborative social vr system for adaptive multi-party real-time communication. *IEEE MultiMedia*, 30(2), 48-59.
- Fuchs, H., State, A., & Bazin, J. C. 2014. Immersive 3D telepresence. *Computer*, 47(7), 46-52. <https://doi.org/10.1109/MC.2014.185>

SURE-VFT Remote Sensing: An XR based Educational Application For Enhancing Skills in Land Use Land Cover (LULC) Mapping Through Satellite Imagery

Petros Patias¹, Charalampos Georgiadis¹, Georgios Mallinis¹,
Anastasios. Stamnas¹, Vasiliki Fragkoulidou¹, Themistoklis Roustanis²,
George Patias², Anastasia Makridou², Patrick Bourdot^{3,4}, and Marco Sacco^{3,5}

^a*School of Rural and Surveying Engineering, Aristotle University of Thessaloniki, Greece*

^b*KIKLO - Geospatial Information Technologies, Greece ,54124*

^c*EuroXR (European Association for eXtended Reality), Belgium*

^d*LISN/CNRS, University Paris-Saclay, France*

^e*CNR-STIIMA (National Research Council of Italy), Italy*

Corresponding author: patias@auth.gr

Keywords: Extended Reality, Education, Virtual Field Trips, Remote Sensing

Abstract

The SURE-VFT Remote Sensing is a virtual field trip scenario designed to strengthen the geospatial and remote sensing skills of undergraduate students in the field of Rural and Surveying Engineering. This immersive virtual experience includes two primary tasks: orienting in-situ photos from the Land Use-Land Cover Area Frame Survey (LUCAS) to the surrounding ground features and categorizing it according to a pre-defined Land Use and Land Cover (LULC) classification system based on satellite image subsets and in-situ LUCAS photos. Utilizing a gamified approach, undergraduate and postgraduate engineers familiarize themselves with Mediterranean rural landscape, learn to identify landmarks in the field, and understand how these features are transformed into interpretation elements in satellite imagery. The system supports interactive navigation, provides immediate feedback, and culminates with a questionnaire and data export functionality for further analysis.

Introduction

Understanding and accurately classifying Land Use and Land Cover (LULC) is essential in environmental monitoring, urban planning, agriculture, and natural resource management. Traditionally, this knowledge is developed through a combination of classroom instruction and fieldwork (Martín-Gutiérrez et al., 2017). However, advances in virtual learning environments now enable students to gain

practical experience in LULC interpretation using interactive and immersive technologies (Pedram et al., 2024).

The SURE-VFT Remote Sensing (RS) has been developed as an educational tool aimed at undergraduate students in Rural and Surveying Engineering. It provides a controlled, simulated environment in which users can explore real-world data and imagery to perform spatial analysis tasks. The experience focuses on two core activities: aligning ground-level photographs with cardinal directions and identifying LULC categories based on multisource visual data, including satellite imagery and geotagged field photos. This manual serves as a comprehensive guide to navigating and completing the virtual experience. It outlines the technical controls, describes the virtual scene, and provides step-by-step instructions for performing tasks. By following this manual, students will not only enhance their geospatial reasoning but also gain valuable skills applicable to GIS and remote sensing workflows.

Purpose and Learning Objectives

The SURE-VFT Remote Sensing virtual field trip scenario is designed mainly for undergraduate students in the School of Rural and Surveying Engineering. It aims to:

- Familiarize students with different LULC types (e.g., forest, agriculture, urban, grassland, shrubland).
- Enhance their skills in interpreting remote sensing (RS) data, including satellite imagery and field photos.
- Train them to record geospatial data for further image classification using GIS tools (e.g., supervised or unsupervised classification in GIS or remote sensing software).

Navigation within the virtual environment

Users are given some basic functions so they can move around the virtual scene without the use of additional equipment:

- Movement: Use W, A, S, D or arrow keys; Shift to sprint, Space to jump.
- View Modes: Toggle between 1st-Person (F2) and 3rd-Person (F3) (Figure D 18).
- Lock/Unlock View: Use F4 to toggle the view lock.

Virtual Scene Overview

The experience is set in a virtual scene (room), where four (4) different room walls present different information:

The 1st wall presents introductory information about the training.

The 2nd wall (Figure D 18) includes (I) Sentinel 2-RGB image over the sample plot and (II) the four ground photos, randomly distributed, that must be properly oriented to the four cardinal directions (North, South, East, and West).

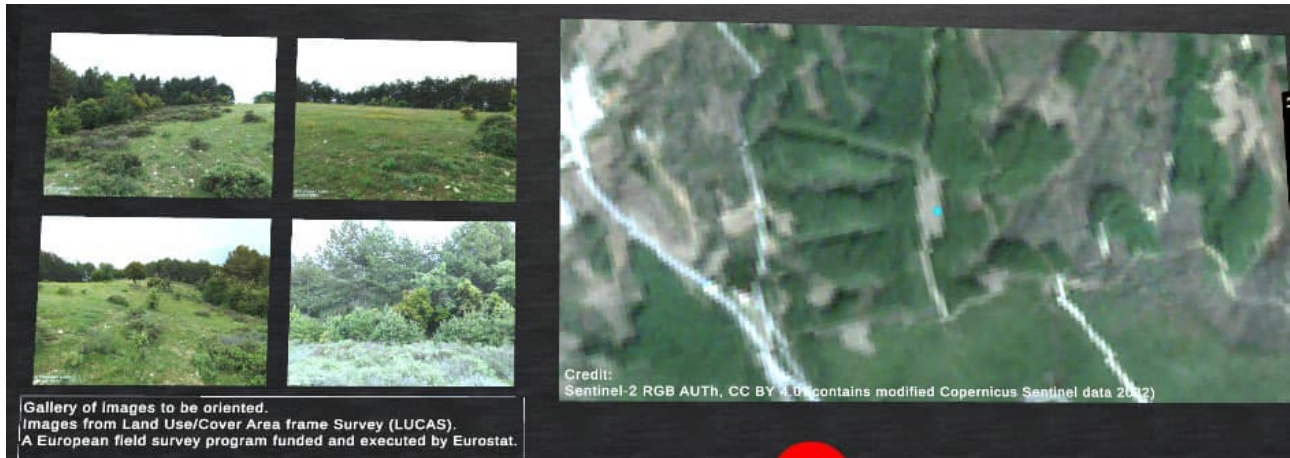


Figure D 18. Ground photos and Sentinel 2-RGB image.

The 3rd wall includes (I) the “active” ground photo, which is the next to be oriented (1 out of the 4) and (II) an orientation board that allows users to track their progress throughout the activity for each plot (whether each cardinal direction has a photo assigned to it).

The 4th wall includes (i) a classification board that allows users to select the LULC type for each point and (ii) a message board to inform users of feedback.

Task 1: Selection of the proper orientation for each ground photo

To complete this process, 5 steps are required:

1. Identification of the photo to be oriented (3rd wall).
2. Rotation of the viewing perspective towards the 2nd wall, where the Sentinel satellite image and the four ground photos are placed.
3. Navigation to the correct 3D red directional marker (N, E, S, W).
4. When correctly matched, the ground photo image is placed next to the corresponding pin (Figure D 19), allowing progression to the next marker. Also, the orientation board now has a green tick mark for the correctly aligned cardinal direction.
5. Rotation of the viewing perspective towards the 3rd wall so that the next ground photo in line to be properly oriented. The same procedure is repeated for all four ground photos.

Task 2: Assignment of the correct LULC class for each plot

The 2nd Task consists of the following steps:

1. Classification of the LULC (e.g., forest, urban, arable, shrubs, heterogeneous) of each plot using the Sentinel image (2nd wall)
2. Immediate feedback is provided, allowing additional attempts to ensure correct classification
3. Upon successful classification, a green exit marker appears.
4. Students then proceed to the next field plot and repeat the same process.

Final Steps: Questionnaire & Data Retrieval

After completing the two tasks over all sample plots, a pop-up prompts users to complete a feedback questionnaire. Upon submission, students receive a certificate and access to the collected data.

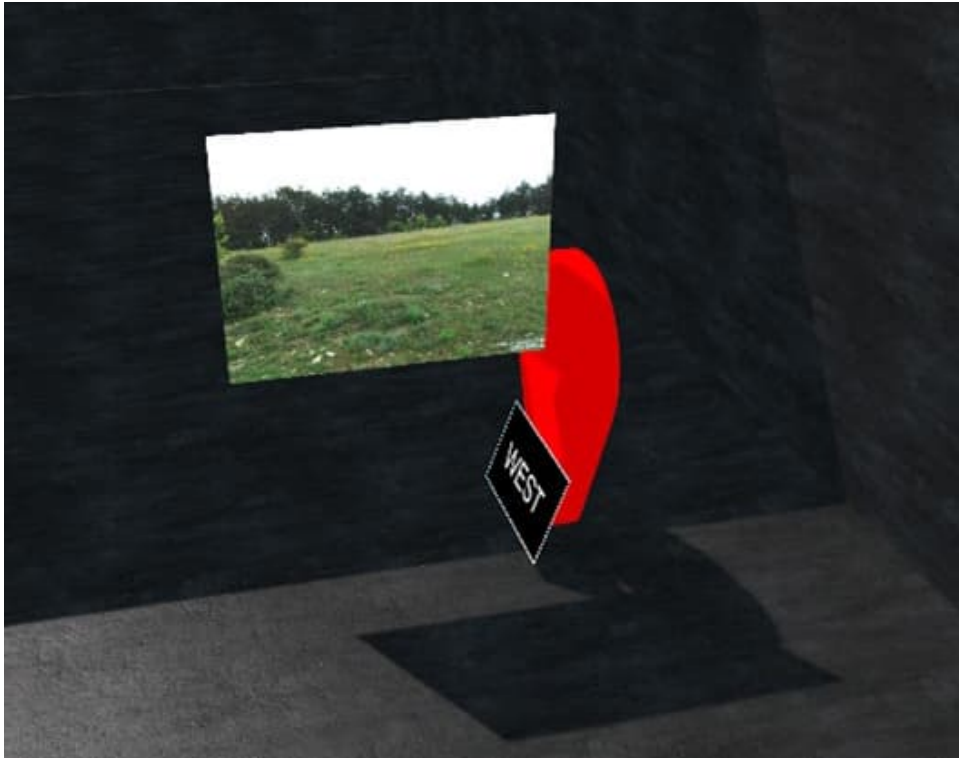


Figure D 19 Correctly matched ground photo to the West pin.

Conclusions

Undergraduate and postgraduate students can learn how to classify land use and land cover using remote sensing data in a thorough and entertaining way with the SURE-VFT Remote Sensing virtual field trip. By combining immersive 3D navigation with real-world geospatial tasks, the SURE-VFT Remote Sensing hosted in the XR4ED36 platform, offers a hands-on learning environment that reinforces theoretical knowledge through real world RS tasks.

Acknowledgements

The SURE-VFT project has indirectly received funding from the European Union's Horizon Europe research and innovation action programme, via the XR4ED – Open Call issued and executed under the XR4ED project (Grant Agreement no. 101093159).

References

- Martín-Gutiérrez, J., Mora, C. E., Añorbe-Díaz, B., & González-Marrero, A. 2017. Virtual technologies trends in education. *Eurasia journal of mathematics, science and technology education*, 13(2), 469-486. <https://doi.org/10.12973/eurasia.2017.00626a>.
- Pedram, S., Kennedy, G., & Sanzone, S. (2024). Assessing the validity of VR as a training tool for medical students. *Virtual Reality*, 28(1), 15. <https://doi.org/10.1007/s10055-023-00912-x>.

³⁶ <https://xr4ed.cellock.com/>

Virtual Reality-Based Exercises for the Rehabilitation of Vestibular Disorders

Edoardo Bonfanti¹, Alessandro Nardi², Marco Sacco², and Sara Arlati²

¹*IRCCS "S. Maria Nascente" - Fondazione Don Gnocchi, Italy*

²*Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing,
National Research Council of Italy, Italy*

Corresponding author: sara.arlati@cnr.it

Keywords: Head Mounted Display, vestibulo-ocular reflex, optokinetic load, retinal slip

Introduction

Acute Unilateral Vestibulopathy (AUVP) is an acute peripheral vestibular syndrome defined by an acute unilateral loss of peripheral vestibular function without evidence of acute central neurological or acute audiological symptoms or signs (Strupp, M. 2022). The annual incidence of AUVP is estimated to range between 3.5 and 15.5 cases per 100,000 population. It represents approximately 7% of cases in specialized dizziness clinics and accounts for 8% of patients presenting with vertigo in general (Adamec, I. 2015). The clinical manifestations of acute vestibular neuritis result from both static imbalance and dynamic dysfunction of vestibular inputs originating from the semicircular canals and otolithic organs. Static imbalance arises from asymmetric tonic activity in the vestibular nuclei at rest, while dynamic disturbances reflect impaired compensatory mechanisms during head movements. (Zee, D.S. 2007). Dynamic signs are characterized by vestibulo-ocular reflex (VOR) asymmetry, indicative of semicircular canal dysfunction, impaired ocular counter-rolling (OCR), and postural instability, reflecting otolithic involvement (Halmagyi, 2010). Despite persistent peripheral vestibular dysfunction, static signs and symptoms typically resolve within weeks. In contrast, dynamic signs may persist indefinitely in the absence of functional recovery, resulting in chronic gaze instability and postural imbalance during head movements directed toward the affected labyrinth (Brandt, T. 2010).

Recently, virtual reality has emerged as a promising technology for implementing interventions to counteract vestibular disorders. VR's immersive and game-like environment can promote greater patient engagement, making rehabilitation exercises less monotonous and improving adherence to protocols (Kwang, H.S. 2024). VR systems provide a straightforward way to deliver more complex stimuli than those typically achievable in conventional rehabilitation settings. Moreover, they offer the advantage of personalization, enabling the creation of digital environments tailored to each patient's specific triggers and rehabilitation goals.

A recent literature review has identified preliminary evidence suggesting that VR may be effective in vestibular rehabilitation - both as a complementary tool and, in some cases, as a standalone

intervention - for improving balance, reducing perceived disability, and enhancing quality of life in patients with peripheral and central vestibular disorders (Le Perf, G. 2025). However, studies are still sparse, and the heterogeneity of the existing studies has not allowed the definition of optimal VR-based protocols and exercises.

In this context, we developed a series of immersive exercises using a VR head-mounted display specifically designed to promote vestibular adaptation and enhance vestibulo-ocular reflex (VOR) gain. These exercises are directly inspired by the foundational VORx1 and VORx2 protocols introduced by Susan Herdman in the early 1990s, and they integrate principles of retinal slip induction and optokinetic stimulation through dynamic visual cues combined with controlled head movements.

Methods

Three different exercises have been designed and developed using Unity and deployed to be used with a Meta Quest 3 Head Mounted Display (HMD).

Exercise 1: Active Head Rotation with Counterphase Stimuli

Our first VR exercise emulates the VORx1 and VORx2 paradigms using a virtual ball as the fixation target, and it is composed of three phases (*static target*, *dynamic target*, *pseudo-random target motion*). In this exercise, to enhance both the magnification effect and the duration of retinal slip stimulation, patients should fixate on a target while performing horizontal or vertical head movements (Herdman, S.J. 1998). Gaze instability is indeed due to the decreased gain of the VOR to head movements, and the most effective stimulus for enhancing VOR gain is the error signal generated by retinal slip (Herdman, S.J. 1998), i.e., the image motion on the retina during head motion.

To perform the VORx1 (*static target* phase), the patient should begin the exercise either seated or standing. A letter or visual target should be placed at eye level, approximately one meter in front of them. The patient is instructed to fix their gaze on this target and maintain visual focus throughout the entire exercise. While keeping their eyes on the target, the patient should voluntarily move the head horizontally (as in a head shake). These head movements must be rapid yet controlled. Studies have shown that the most significant adjustments in VOR gain occur at all the frequencies used during training (Herdman, S.J. 1997); thus, the VR environment frequency functioning automatically adapts to patients' head motion.

In the VORx2 (*dynamic target*) phase, the patient moves and turns the head toward one direction (e.g., to the right) while the target simultaneously moves in the opposite direction (to the left). After returning both the head and the target to the central position, the patient repeats the movement to the other side - head to the left, target to the right - maintaining steady visual fixation on the moving target.

Finally, in the pseudo-random motion phase, the target ball adopts an unpredictable, pseudo-random trajectory while staying within the patient's visual field. This stage elevates task complexity, challenging the patient's ability to stabilize gaze under sudden and irregular visual perturbations.

Exercise 2: Active Head Rotation with Directional Discrimination

The second exercise combines active head movements with visuospatial discrimination. The patient is instructed to maintain gaze on a central fixation point while actively rotating their head in various directions. Upon each movement, a directional arrow randomly appears within the headset display. The patient must verbally identify the direction of the arrow.

This task promotes multisensory integration, requiring the brain to seamlessly combine vestibular and visual inputs, thereby potentially strengthening the VOR pathways. Furthermore, it challenges cognitive skills and rapid visuospatial processing, which are crucial for real-world tasks. This exercise is inspired by the principles of the Functional Head Impulse Test (FHiT). The FHiT is a clinically validated assessment tool in which the individual is required to fixate on a central target while performing rapid, unpredictable head impulses; meanwhile, brief visual stimuli (such as optotypes or symbols) are presented, and the subject is asked to recognize or report them. This setup assesses not only the functional integrity of the VOR but also the ability to integrate visual and vestibular cues during dynamic head movements.

Exercise 3: Visual Scanning under Dynamic Optokinetic Load

Retinal slip can also be triggered by optokinetic visual stimuli (Herdman, S.J. 1998). In particular, training with unidirectional optokinetic input can enhance vestibular responses in the corresponding direction (Shelhamer, M. 1994). The third exercise thus proposes an optokinetic visual stimulus and resembles a standard task in which the patient is asked to fix their gaze on a screen with black and white stripes moving slowly in a horizontal direction. The patient follows the continuous movement of the stripes with their eyes, taking care to keep their head perfectly still. As the stripes reach the edge of the patient's field of vision, their eyes quickly snap back to the opposite side to follow the next movement of the stripes, mimicking the natural optokinetic nystagmus response.



Figure D 20. A screenshot of the third exercise illustrating the graffiti that the patient is required to count as part of the visual task.

We improved upon the standard version by providing a more realistic and immersive exercise. The patient is positioned in front of a slowly moving train covered in graffiti, rendered within the VR environment. The train traverses the visual field laterally, providing continuous optokinetic stimulation. The patient is tasked with visually scanning the train and counting discrete graffiti images.

Conclusions and Future Work

In the near future, a protocol to conduct a first feasibility trial will be defined. Such a protocol will rely on existing recommendations for gaze stability exercises, which foresee 4 to 5 times per day, totaling 20 to 40 minutes daily, supplemented by 20 minutes of balance and gait training (Herdman, S.J. 2007). Such a first trial will also encompass assessing patients' user experience, integrating usability, acceptance, and cyber-sickness investigation on a subjective level.

References

- Adamec, I., Krbot Skorić, M., Handžić, J., & Habek, M. 2015. Incidence, seasonality and comorbidity in vestibular neuritis. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 36(1), pp. 91–95. Available at: <https://doi.org/10.1007/s10072-014-1912-4>.
- Brandt, T., Huppert, T., Hüfner, K., Zingler, V.C., Dieterich, M., & Strupp, M. 2010. Long-term course and relapses of vestibular and balance disorders. *Restorative Neurology and Neuroscience*, 28(1), pp. 69–82. <https://doi.org/10.3233/RNN-2010-0504>.
- Halmagyi, G.M., Weber, K.P., & Curthoys, I.S. 2010. Vestibular function after acute vestibular neuritis. *Restorative Neurology and Neuroscience*, 28(1), pp. 37–46. <https://doi.org/10.3233/RNN-2010-0533>.
- Herdman, S.J. 1997. Advances in the treatment of vestibular disorders. *Physical Therapy*, 77(6), pp. 602–618. <https://doi.org/10.1093/ptj/77.6.602>.
- Herdman, S.J. 1998. Role of vestibular adaptation in vestibular rehabilitation. *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, 119(1), pp. 49–54. [https://doi.org/10.1016/S0194-5998\(98\)70195-0](https://doi.org/10.1016/S0194-5998(98)70195-0).
- Herdman, S.J. 2013. Vestibular rehabilitation., *Current Opinion in Neurology*, 26(1), pp. 96–101. <https://doi.org/10.1097/WCO.0b013e32835c5ec4>.
- Herdman, S.J., Clendaniel, R.A., Mattox, D.E., Holliday, M.J., & Niparko, J.K. 1995. Vestibular adaptation exercises and recovery: acute stage after acoustic neuroma resection., *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, 113(1), pp. 77–87. [https://doi.org/10.1016/s0194-5998\(95\)70148-6](https://doi.org/10.1016/s0194-5998(95)70148-6).
- Herdman, S.J., Hall, C.D., Schubert, M.C., Das, V.E., & Tusa, R.J. 2007. Recovery of dynamic visual acuity in bilateral vestibular hypofunction. *Archives of Otolaryngology--Head & Neck Surgery*, 133(4), pp. 383–389. <https://doi.org/10.1001/archotol.133.4.383>.
- Kwang, H.S. 2024. Clinical application of virtual reality for vestibular rehabilitation. *Research in Vestibular Science*, 23(4), pp. 124–131. <https://doi.org/10.21790/rvs.2024.019>.
- Le Perf, G., Thebault, G., Duflos, C., Herman, F., Cauquil-Gleizes, S., & Laffont, I. 2025. Can virtual reality replace conventional vestibular rehabilitation tools in multisensory balance exercises for vestibular disorders? A non-inferiority study. *Journal of NeuroEngineering and Rehabilitation*, 22(1), p. 86. <https://doi.org/10.1186/s12984-025-01623-x>.
- Shelhamer, M., Tiliket, C., Roberts, D., Kramer, P.D., & Zee, D.S. 1994. Short-term vestibulo-ocular reflex adaptation in humans. II. Error signals. *Experimental Brain Research*, 100(2), pp. 328–336. <https://doi.org/10.1007/BF00227202>.
- Strupp, M., Bisdorff, A., Furman, J., Hornibrook, J., Jahn, K., Maire, R., Newman-Toker, D., & Magnusson, M. 2022. 'Acute unilateral vestibulopathy/vestibular neuritis: Diagnostic criteria', *Journal of Vestibular Research: Equilibrium & Orientation*, 32(5), pp. 389–406. <https://doi.org/10.3233/VES-220201>.

Immersive Exploration of Fluid Dynamics: Developing an XR/VR Platform for CFD Virtual Testing in Education and Social Inclusion

Konstantinos Lyras¹, Dinesh Bhatia¹, and George Papadakis²

¹*MultiFluidX, Greece*

²*National Technical University of Athens, Greece*

Corresponding author: k.lyras@multifluidx.com

Keywords: VR, XR, CFD simulation, 3D modelling, Real-time modelling, Telecooperation, Telepresence.

vScientist part of CORTEX²

Many engineering concepts can be complex and abstract, making it difficult to understand them. Virtual Reality (VR) technology can be a very useful tool in education and training with high-quality, inclusive, and sustainable tools using real-life examples.

The project vScientist combines Computational Fluid Dynamics (CFD) simulations and the next generation telecooperation framework based on VR and AI of CORTEX² project (Figure D 21). NTUA and MultiFluidX developed a straightforward workflow for enhanced viewing of 3D physics-based simulations and associated data in an immersive virtual environment. The project enables real-time data transfer, dynamic visualisation and more efficient remote collaboration in STEM fields. Through this platform, users can visualise, interact, and analyse 3D virtual experiments in an immersive environment, or run their own in seconds. For instance users can explore a virtual wind tunnel and learn wind turbine optimisation (Figure D 22).

The project enhances the learning and accessibility for students and individuals from diverse backgrounds, promoting social inclusion and accessibility.

Over the 12-month implementation period, the vScientist platform successfully developed a state-of-the-art VR/XR-CFD application specifically for training Fluid Dynamics STEM students, deploying three distinct simulated scenarios (virtual wind tunnel with 3D airplane wing, 3D virtual marine propeller testing, and flow around cylinder, Figure D 22 and Figure D 23) and conducting comprehensive validation with many users (undergraduates and postgraduates) in operational classroom environments. The systematic evaluation demonstrated significant educational value, with students rating the platform's potential to improve fluid dynamics learning (Figure D 24). The achievement of TRL7 through classroom testing and positive user feedback confirmed the platform's readiness for broader educational implementation. This advancement fostered more efficient collaboration between

educators, students and researchers promoting social inclusion, accessibility and remote collaboration in STEM fields.

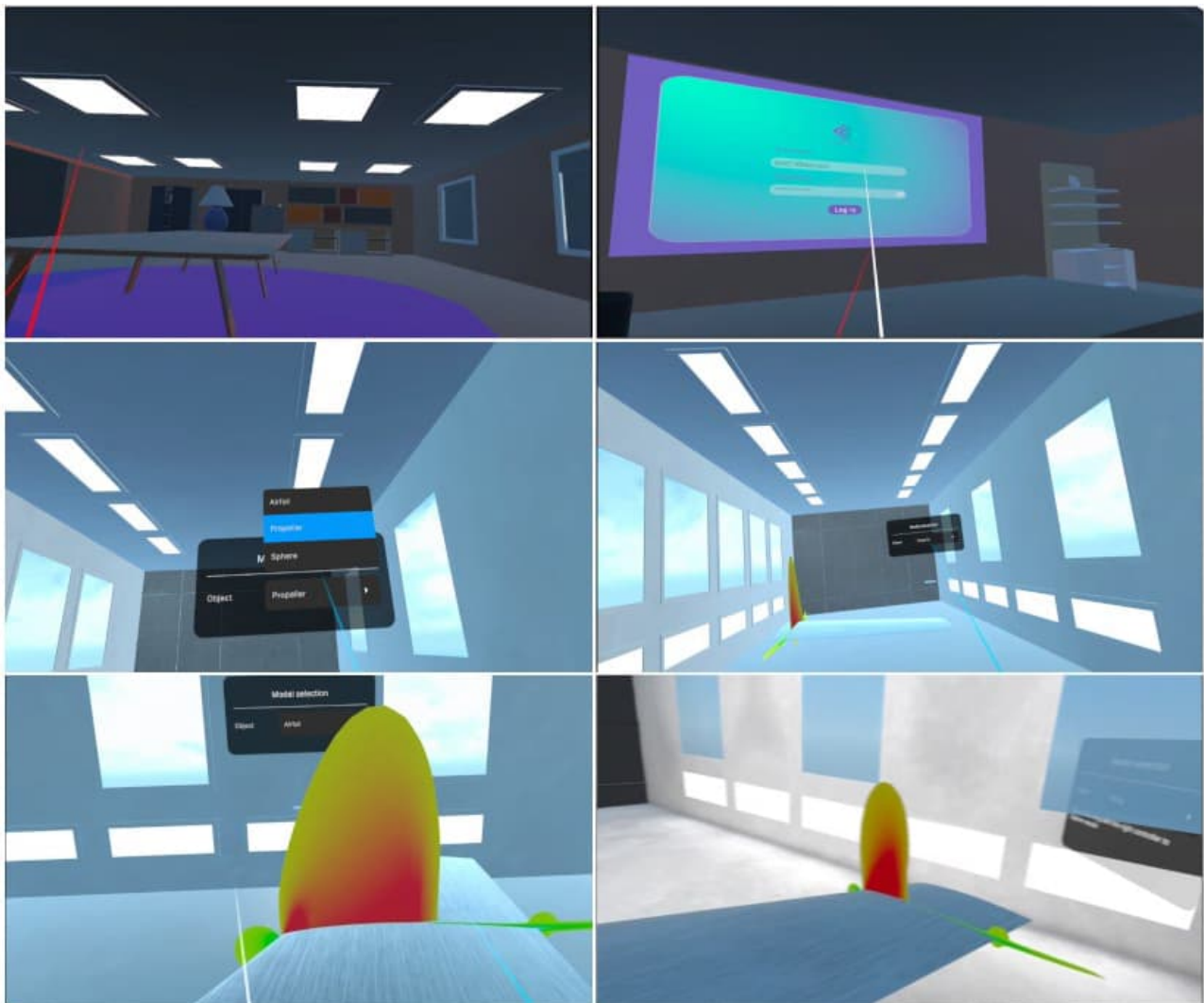


Figure D 21. Enhancing education through extended reality and vScientist in CORTEX². Showcase of the developed CFD-VR workflow for the wind tunnel case. Users can model and analyse shock waves on turbine blades.



Figure D 22. Wind tunnel testing used for our in-classroom demonstration with a 3D model of NTUA's wind tunnel.

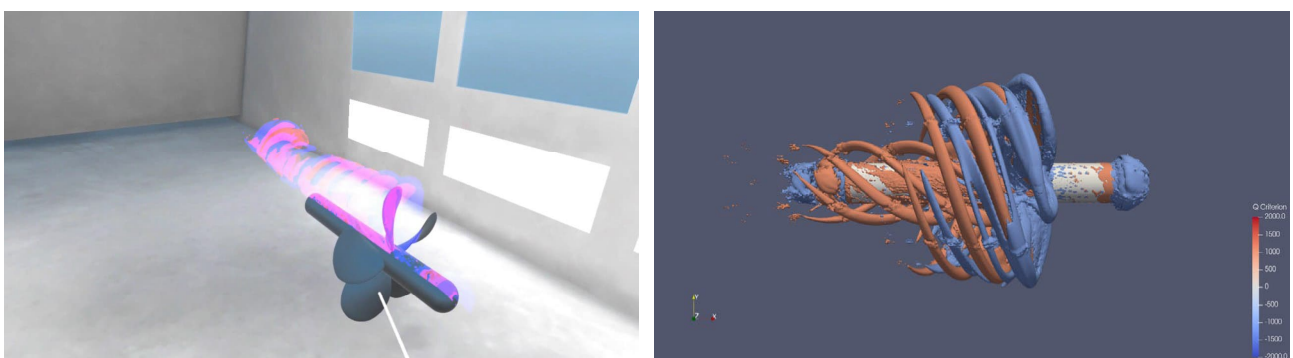


Figure D 23. A ship propeller model simulated using CFD and visualised with our VR-CFD application. The VR model (left) and the CFD simulation (right).



Figure D 24. Responses for evaluating user experience.

References

- Papadakis, G., & Manolesos, M. 2020. The flow past a flatback airfoil with flow control devices: Benchmarking numerical simulations against wind tunnel data. *Wind Energy Science*, 5(3), 911-927. <https://doi.org/10.5194/wes-5-911-2020>
- Bhatia, D., Loukas, J., Cabrera, A., & Lyras, K. 2024. A deep learning computational fluid dynamics solver for simulating liquid hydrogen jets. *Physics of Fluids*, 36(5). <https://doi.org/10.1063/5.0206562>
- Bhatia, D., & Hesse, H. 2023. Enhancing Student Engagement in Engineering and Education Through Virtual Reality: A Survey-Based Analysis. In *TENCON 2023-2023 IEEE Region 10 Conference (TENCON)*, 170-175.
- Lyras, P., Hubert, A., & Lyras, K. G. 2023. A conservative level set method for liquid-gas flows with application in liquid jet atomisation. *Experimental and Computational Multiphase Flow*, 5(1), 67-83.

Author Index and Tables of Figures & Lists of Tables

Author Index

A

Acheampong, Rebecca	246
Ackermann, Philipp.....	68, 70, 71, 179, 183
Adam-Ferri, Lorena	229
Ahmed, Zobaer	243
Aichele, Isabell	193
Albanis, Georgios	152, 274
Alesani, Andrea	22, 26, 58, 62
Almécija Pereda, Ana	139
Alobaid, Maha.....	94
Alonso-Ramos, Victor	117
Amditis, Angelos	63
Ampatzoglou, Apostolos.....	189
Anghel, Vladut-Gabriel	246
Antoniacomì, Alessio	104
Antonietti, Alessandro	165
Arend, Johannes M.	174
Arlati, Sara	299
Atzori, Luigi	144
Azhar, Abraiz.....	37

B

Bafes, Konstantinos	13
Ballesteros, Marta García	135
Banakou, Domna	37
Bantouna, Aimilia	291
Barone, Nunzio	217
Barous, Ioannis	184
Becker, Daniela	221
Benazzouz, Yazid	278
Benc, Wolfgang.....	174
Berretta, Luciana	89
Bhatia, Dinesh.....	303
Bi, Xinyu	225
Bistas, Fotios.....	149, 184
Blázquez Gil, Gonzalo.....	139
Bogdan, Ioana Corina.....	246
Bolierakis, Spyridon Nektarios.....	63
Bolognini, Silvia.....	160
Bonfanti, Edoardo.....	299
Bosta, Athina	98
Bourdot, Alain-Pierre	54
Brescia, Walter.....	217
Bukits, Tamás.....	109

C

Caboni, Renato.....	144
Cabrero Barros, Sergio	109
Cai, Meng	113
Campagnoli, Carlo	225
Carreiro Alonso, Antonio	139
Carreres-Fluixà, Josep	229
Castillo-López, Galo.....	206
Cernigliaro, Gianluca	104, 257
Chalendar, Gaël de.....	278
Chatzigeorgiou, Alexandros	189
Chatzitofis, Anargyros	152, 274
Christofi, Maria	80
Cisternino, Ivan	217
Collazo, Miguel.....	121
Colombo, Vera	160, 165
Concheiro, Antón	121
Cords, Hilko	50
Crevecoeur, Karen	125

Ç

Çöltekin, Arzu.....	234, 251
---------------------	----------

D

Dacal-Nieto, Angel	117
Dadoukis, Aris	63
Danciu, Gabriel.....	246
de Chalendar, Gaël.....	206
De Cicco, Luca	217
de Torres Coll, Patricia	109
Delgado, Angel	121
Díaz-Labrador, Anabel.....	121
Domínguez-Pedreira, Alba	117
Drejer, Julia	221

E

Egglì, Giulia Maria	170
Elosegui, Ander	109
Escrivá, David Millán	135
Evangelou, Iordanis.....	128, 149

F

Fernández	
Pablo	121
Sergi	238
Fernández Graña, Pablo M.	139
Ferreira, Wesley	89
Floricescu, Bogdan-Valentin	246

Floris, Alessandro..... 144
 Fluri, Luca..... 50, 234
 Folina, Maria-Theodora 189
 Fraedrich, Katja..... 262
 Fragkoulidou, Vasiliki..... 54
 Freitas, Carla M.D.S. 89
 Fresnais, Alexandre..... 125
 Fűrnstahl, Philipp..... 45

G

García-Ballesteros, Marta 229
 Gattullo, Michele 76
 Gavalas, Damianos..... 27
 Gazi, Angeliki 27
 Georgiadis, Charalampos..... 54
 Gerth, Tatiana 45, 243
 Getaneh, Mahder Mulusew..... 17
 Giannopoulou, Eleni 291
 Giménez Millán, Alfredo 135
 Gkaravelis, Anastasios 128, 149
 Gómez-Maureira, Marcello A. 157
 González, Diego 238
 Goriachev, Vladimir 26, 58, 62
 Goutsu, Yusuke 213
 Grabner, Helmut..... 45, 243
 Grasnick, Armin 221
 Gunkel, Simon N.B. 104
 Gunkel, Simon N.B. 257
 Guths, Leonardo 89
 Gürçınar, Elif 251

H

Hadjipanayi, Christos 37
 Hayes, Hannah..... 225
 Hazboun, Shana 13
 Heikkinen, Jaakko 174
 Helin, Kaj.....4, 22, 26, 62
 Heurtier, Éric..... 125
 Hoch, Armando..... 45
 Holler, Alessandro..... 179
 Huerta, Ivan 238
 Hüsser, Cloe 50
 Häsler, Steven 179

I

Inamura, Tsenari..... 213

K

Kadlubsky, Alina..... 84
 Karaceylan, Burak 132
 Karadayi, Zafer 132
 Karagiannis, Ioannis..... 63

Karaseitanidis, Ioannis 63
 Karjalainen, Jaakko..... 58, 62
 Kasapakis, Vlasis..... 27
 Kassens-Noor, Eva..... 113
 Katika, Tina..... 63
 Kazanidis, Ioannis..... 13
 Kefala, Anna 184
 Kessler, Martina 179
 Khan, Ali Mehmood 221
 Klunder, Tessa 104, 257
 Kolokytha, Eirini 291
 Korovesis, Charalampos..... 291
 Kotarelas, Nikolaos 152, 274
 Kotsis-Panakakis, Vasileios-Ektoor 149
 Kounalaki, Maria 291
 Kovács, Andor 84
 Kuula, Timo 22, 26

L

Ladjánszki, Balázs 84
 Lampropoulos, George..... 13
 Lanitis, Andreas..... 80
 Lasica, Ilona..... 13
 Laviola, Enricoandrea..... 76, 77, 79
 Leu, Andreas 251
 Lévai, Tamás..... 270
 Levigoureux, Alex..... 125
 Li, Yue..... 157
 Lindahl, Peter 174
 Lokki, Tapio 174
 Lombard, Alexis..... 206, 278
 Lorré, Jean-Pierre..... 278
 Lukic, Yanick X..... 179
 Lyras, Konstantinos 303

M

Maciel, Anderson 89
 Maggio, Antonio Pio 217
 Makridou, Alexandra 54
 Mallinis, Georgios 54
 Mangina, Eleni 13
 Manzke, Michael..... 94
 Marangelli, Luana 76
 Martos, Marc 238
 Mascolo, Saverio..... 217
 Medrano Navas, Samuel..... 135
 Merz, Cédric..... 251
 Meseguer Valenzuela, Andrés 135
 Meyer, Michelle 201
 Meyer-Kahlen, Wolfgang 174
 Michael-Grigoriou, Despina 37, 40
 Miniussi, Glauco..... 104

- Mondellini, Marta 165
Montoya, Sergio 238
Moreno-Renart, Laura 229
Moser, Mirella 262
Mygdakos, Gregory 189
- N**
Nardi, Alessandro 299
Naskou, Kostas 63
Nassar-Kyriakidou, Chalima Dimitra 184
Navas-Medrano, Samuel 229
Nedel, Luciana 89, 93
Negrão, Matheus 89
Németh, Felicián 270
Nikolaidis, Nikolaos 189
Nuss, Julia 221
- O**
Oppermann, Leif 41, 266
Oprea, Ionut Alexandru 246
Ortega-Bataller, Susana 229
- P**
Papadakis, George 303
Papaioannou, Georgios 149
Papathanasiou, Athanasia Maria 184
Patias
 George 54
 Petros 54
Paunovic, Natalija 197
Pelle, István 270
Pereira, Lucas 89
Pérez
 Héctor 121
 Jose Vera 135
Perisé, Ramón 109
Perperidou, Dimitra 189
Perret
 Anna 31
 Jerome 31
Pierdicca, Roberto 160
Pittaras, Iakovos 184
Pons, Patricia 229
Porcu, Simone 144
Prokudin, Sergey 45
Pugliese, Raffaele 160
- R**
Ragonesi, Ruggero 238
Rahim, Galit 257
Regondi, Stefano 160, 165
Reichherzer, Carolin 201
Rekeraho, Alexandre 246
Riedlinger, Urs 41, 266
Rienda-Jimena, Jesús 229
Rizk, Amr 283
Rocamora Montoro, Pablo 135
Rodríguez-González, Oscar 117
Rodríguez-Pérez, María 117
Romano, Sara 76
Rommens, Jochem 211
Rossidis, Vagelis 189
Roustanis, Themistoklis 54
Rudolph, Michael 283
Ruiz, Albert 238
- S**
Sacco
 Marco 165, 299
 Maurizio 54
Santangelo, Gabriele 217
Santos-Torres, Andrés 109
Schmidlin, Stefan 179
Schnier, Finn 283
Schäfer, Wolfgang 170, 193, 197
Seibold, Matthias 45
Seimenis, Aimilios 63
Semmar, Nasredine 206
Serantes, Santiago 121
Socas, Diego 121
Sokratous, Dimitris 37
Soler-Dominguez, José Luis 229
Soler-Domínguez, Jose Luis 135
Stamnas, Anastasion 54
Stathoulas, Konstantinos 63
Stavroulia, Kalliopi-Evangelia 27, 80
Suhre, Niklas 113
Sörös, Gábor 84
- T**
Tamayo, Iñigo 109
Tang, Siyu 45
Terzopoulos, George 13
Theodoridis, Konstantinos 27
Thermos, Spyridon 152, 274
Thomas, Yannis 184
Tikka, Petri 58
Toledo, Leonel 238
Tomás, Patricia Pons 135
Tonetto, Flavio 160
Torchelsen, Rafael 89
Tsaramirsis, Konstantinos 13
Tsertou, Athanasia 63
Tsimiklis, Georgios 63

Tsinakos, Avgustos..... 13
 Tsioutas, Konstantimos..... 184

U

Uzun, Yücel 41, 266

V

Vagianos, Dimitris..... 27
 van den Berg, Stéphanie M..... 157
 Vasilakis,
 Andreas A. 128
 Andreas-Alexandros 149
 Veiga, Diego 121
 Vitsas
 Nick..... 128
 Nikolaos 149
 Volioti, Christina 189
 Vona, Cecilia 160
 Vosinakis, Spyros 98
 Wan, Annie 72

X

Xylomenos, George..... 184

Y

Yildirim, Funda 157

Z

Zank, Markus 201
 Zioulis, Nikolaos 152, 274
 Zorrilla, Mikel..... 109

.

Tables of Figures

FIGURES IN APPLICATION TRACK

Figure 1.	Example: "Distribution of resources ZENODO Learning Scenario & Spatio.io environment activities.	15
Figure 2.	CoXist Plus VR game Leve 1 and 3 scene.....	18
Figure 3.	CoXist Plus VR game Leve 4 and 5 scene.....	18
Figure 4.	CoXist Plus VR game Vernacular architecture and designed modeled assets	19
Figure 5.	CoXist Plus VR game Vernacular architecture interior of rural houses	19
Figure 6.	CoXist Plus VR game testing, hackathon and workshop	20
Figure 7.	CoXist Plus VR game demo at science museum for government officials	21
Figure 8	Tablet (iPad Pro M4 (2024)) interface with snow groomer.....	23
Figure 9.	Head mounted (Apple Vision Pro) interface in slope.	24
Figure 10.	Updated iPad features e.g. a color-coded snow depth visualization	25
Figure 11.	The characters of Maria and Giannis (from the observers' perspective).....	29
Figure 12.	Recording of the drilling progression speed with a Virtuose 6D High-Force device.....	32
Figure 13.	Records of drill penetration speed for the drilling depth with constant force	33
Figure 14.	Record of resistance torque while fastening the screw with rotation speed 60 deg/s...33	33
Figure 15.	Succession of virtual fixtures (left to right)	34
Figure 16.	Second author testing the haptic rendering (without physical prop).	35
Figure 17.	The environment and gameplay of the featured VR application.....	39
Figure 18:	Sorting task performed without AI assistance, illustrating the challenge of distinguishing plastics from metals.....	42
Figure 19:	Overview of the environment.	43
Figure 20.	SurgiScope video interface with integrated Step Browser.....	47
Figure 21.	Same 3D reconstructed step from different viewpoints.	47
Figure 22.	Left: Evaluation mode showing rotation gizmos and target saw. Right: Quiz interface.	48
Figure 23.	Our AR workflow.....	51
Figure 24.	Example screenshots from an active VR session.	52
Figure 25.	The 3D representation of the observatory, the student (as an avatar)	55
Figure 26.	Ground level photos and Sentinel-2 image.....	56
Figure 27.	Satellite Imagery Wall and spectroradiometer screens	57
Figure 28.	Overview of the simulator setup.....	59
Figure 29.	The assistant and the chat panel.	59
Figure 30.	The visual effects coming from the helicopter model.....	60
Figure 31.	The input devices for the MQTT broker.....	60
Figure 32.	Simplified conversation flow.	61
Figure 33.	Overview of the Ergon-XR system architecture	65
Figure 34.	Physical Pose Calibration Flow	65
Figure 35.	Runtime REBA input and results	66
Figure 36.	Results panel with joint-angle plots and video review of the session.....	66
Figure 37.	AR View: the user's augmented reality view	69
Figure 38.	Interactive editor and test environment for spatial rules.....	71
Figure 39.	One of the VR scenarios which trained caregivers to be aware of food details	73

Figure 40.	One of the VR scenarios which provides choices for the trainees.....	73
Figure 41.	A lived experience in training	73
Figure 42.	A practicing nurse in training.....	73
Figure 43.	Left: The INFINIBRICK setup with the storage boxes, Right: A user engaging the endless assembly task following the LEGO cues	78
Figure 44.	InclusiVR@Work immersive experiences reflecting visual impairments.....	82
Figure 45.	Our open globe viewer tool for developers of location-based AR offers visualization of various OGC GeoPose sources.....	84
Figure 46.	Left: localization query. Middle: First-person view of the estimated camera pose.....	86
Figure 47.	The GeoPose streams of two users are visualised in our globe view.	87
Figure 48.	Sending waypoints from the globe viewer directly into a WebXR session on a mobile client.....	88
Figure 49.	Bimanual asymmetric interaction with an endoscope.....	90
Figure 50.	A user trying to press a button that is outside his safety limits.....	91
Figure 51.	Reconstruction of a medical consultation room presenting the environment and position of the subject user and the virtual patient, and appearance of the avatars.	92
Figure 52.	A user visualizing and annotating a previously recorded simulated training session in a doctor’s office. Besides annotations, temporal navigation, path travelled, and avatar with focused field of view are shown.	92
Figure 53.	The trace feature to visualize the user’s input.....	95
Figure 54.	Translation function with real-time OpenGL command in the VR environment.	96
Figure 55.	A capture of the two digital characters inside the mansion.....	99
Figure 56.	The study process (left) and photos of users during the study (right).....	100

FIGURES IN POSTERS

Figure P 1.	Initial demonstration of the use case in action	105
Figure P 2.	Architecture and deployment.....	106
Figure P 3.	Architecture of the system.....	111
Figure P 4.	Simulation setup and equipment.	114
Figure P 5.	Telepresence view of a senior operator teaching a task to a junior operator.....	118
Figure P 6.	SELFEX ² workbench (left) and in its portable cabin (right).....	119
Figure P 7.	HoloLens 2 usage (left) and phantom hands visualization (right).....	119
Figure P 8.	ROTIE architecture	122
Figure P 9.	Testing ROTIE telepresence framework with dogbot Unitree GO 1 and AR glasses Quest 3.	124
Figure P 10.	An AR scene anchored using ARY.....	126
Figure P 11.	Indicative outputs from the VISOR service.....	130
Figure P 12.	Sequence diagram of the SAME-XR workflow.....	133
Figure P 13.	Screenshots from the web interface.....	134
Figure P 14.	Microsoft HoloLens 2 user interface for robot teleoperation.	136
Figure P 15.	Infrastructure deployment at Factor S.L (end user) premises.....	137
Figure P 16.	User validation at Factor S.L. premises (left) and DT premises (right).....	138
Figure P 17.	FLYTEX architecture.....	140
Figure P 18.	FLYTEX widget.....	141
Figure P 19.	FlyThings Developer testing Flytex in XR glasses	142
Figure P 20.	KPIs	143

Figure P 21.	A point cloud displayed in XR mode using the Meta Quest 3.	145
Figure P 22.	Two PCs representing the selected sound stones created by the artist Pinuccio Sciola.	146
Figure P 23.	HOPE pipeline in detail	151
Figure P 24.	Our approach properly skinning the non-linear Gaussian components, the rotation and SH coefficients.	154
Figure P 25.	An efficient Gaussian Avatar pipeline with no learnable parameters apart from the Gaussian components.	154
Figure P 26.	Experimental Design for Animal Perspective-Taking Experience.....	158
Figure P 27.	XRehab simulated 3D environments. The three outdoor landscapes are on the left. On the right there is the simulated home location where the range-separated and the sorting challenges are performed.	161
Figure P 28.	Structure of the user validation protocol (NRS for emotional scenes only).	162
Figure P 29.	a) Pilot participants, b) Pre-Post results of the STAI-Y1 for evaluating the anxiety level, c) Emotional evaluation of the outdoor scenarios (NRS and SAM), d) User experience evaluation (ITC-SOPI, UEQ-S, and SUS).	163
Figure P 30.	Task 1 “Keep it up” on the left; task 2 “Overtakes” on the right	168
Figure P 31.	Conceptual model linking storytelling elements → narrative transportation & presence → four senior benefits. (own illustration).....	171
Figure P 32.	Concept of a holographic call.....	175
Figure P 33.	Sketch of a technical audio rendering system capable of dynamic, room-adapted spatial audio.	175
Figure P 34.	Example of a room impulse response, consisting of direct sound (dark blue) and reflections (light blue).	175
Figure P 35.	Head-worn microphone array used for testing blind room identification algorithms.	175
Figure P 36.	Initial results for JND of changes in DRR. The threshold is the lowest for medium to high DRRs.	177
Figure P 37.	Holographic calling setup with HMD and transparent headphones.	177
Figure P 38.	Athens testbed.....	185
Figure P 39.	Audio latency measurement setup	186
Figure P 40.	Video latency measurement setup.	186
Figure P 41.	Audio latency	187
Figure P 42.	2D video latency.....	187
Figure P 43.	Volumetric video latency.....	188
Figure P 44.	AgriVision with (a) Microsoft Hololens 2 and (b) Vizux M400.	190
Figure P 45.	AgriVision’s 2 versions: (a) Pro Version deployed on Hololens offering real-time enhanced visualization, and (b) Light Version designed for mobile device.	191
Figure P 46.	Immersive Donation Funnel.	194
Figure P 47.	Demonstration of issues with visually presented cognitive stimulus in VE as commonly used.....	203
Figure P 48.	Macro-F1 score of the intent recognition models fine-tuned on the generated utterances.....	209
Figure P 49.	Technical architecture live volumetric video.....	212
Figure P 50.	Holosys Live volumetric video (~10s delay).	212
Figure P 51.	Roadmap of our work (left) and on-going work (right).	213
Figure P 52.	VR ball-throwing system (left) and ball-throwing simulation using generated data (right)	214

Figure P 53.	The overall immersive teleoperation system employed in the demonstration.....	218
Figure P 54.	Example of navigation scenario.	219
Figure P 55.	3D-Scene (left) and Close up (right).....	223
Figure P 56.	The "Leonardo da Vinci" VR Studio Environment.	226
Figure P 57.	PC-to-HMD modality (left), HMD-to-HMD modality (right)	230
Figure P 58.	Overview of our selection handling concept.	235
Figure P 59.	Our four selection techniques.	236
Figure P 60.	User Interface: Main tool Panel on the left, Object Editing Panel in the middle and Inventory Panel on the right.	244
Figure P 61.	Showcase Applications: Ball Toss on the left, Educational Object Labelling in the middle and Planetary Orbit Simulation on the right.	245
Figure P 62.	Genba SaVR system architecture showing containerized microservices, Kubernetes orchestration, Nakama backend services, adaptive learning pipelines, and integrated security monitoring.....	247
Figure P 63.	Adaptive learning in Genba SaVR using MageAI for real-time content personalization.....	248
Figure P 64.	WebRTC latency measurements with Genba SaVR deployed on the SPIRIT platform infrastructure.	249
Figure P 65.	Multi-user collaboration during line inspection, enabling shared evaluation of station-specific components in the virtual environment.	249
Figure P 66.	Three pointer styles (green ball indicates the position of the "pointer").....	252
Figure P 67.	Example of the avatar representations included in the prototyping phase.	253

FIGURES IN DEMOS

Figure D 1.	Three capture and render perspectives for industrial trainings	258
Figure D 2.	System Architecture.	259
Figure D 3.	Stress Scenario.....	263
Figure D 4.	Reflection Phase	264
Figure D 5.	Overview of the functionalities of the dashboard.....	268
Figure D 6.	Screenshot of the dashboard control interface	268
Figure D 7.	Screenshot of the highscore (left) and statistics (right) collected by the dashboard.....	268
Figure D 8.	System architecture.....	272
Figure D 9.	A multiview camera markerless motion capture system streaming animation data to downstream applications that embody users in different representations.....	275
Figure D 10.	The two images on the left show two different nodes that stream data to a shared virtual environment using generic avatars.	276
Figure D 11.	Business Meeting Immersive room.....	279
Figure D 12.	Schematic Overview of our Demo.	284
Figure D 13.	UX/UI moodboard of Producer and Consumer SEF GUI.....	288
Figure D 14.	Ludovico Ariosto	289
Figure D 15.	The high-level proposed pipeline	290
Figure D 16.	System Architecture Schema	292
Figure D 17.	Capture of the 360° panoramic VR app of the Corinth Canal.....	293
Figure D 18.	Ground photos and Sentinel 2-RGB image.....	297
Figure D 19.	Correctly matched ground photo to the West pin.....	298

Figure D 20. A screenshot of the third exercise illustrating the graffiti that the patient is required to count as part of the visual task.....	301
Figure D 21. Enhancing education through extended reality and vScientist in CORTEX ²	304
Figure D 22. Wind tunnel testing used for our in-classroom demonstration with a 3D model of NTUA's wind tunnel.....	305
Figure D 23. A ship propeller model simulated using CFD and visualised with our VR-CFD application. The VR model (left) and the CFD simulation (right).....	305
Figure D 24. Responses for evaluating user experience.	306

Lists of Tables

TABLES IN APPLICATION TRACK

Table 1.	CoXist Plus - Impact and Evaluation testing data	20
Table 2.	Affordance comparison of four commercial VR artmaking applications with the potential of being used as part of a chronic stroke rehabilitation regimen.	38

TABLES IN POSTERS

Table P 1.	Initial Measurements.....	107
Table P 4.	Results summary	138
Table P 3.	Comparison using HUGS (Kocabas et al., 2024) on the NeuMan dataset.....	155
Table P 6.	Comparison using iHuman (Paudel et al., 2024) on the PeopleSnapshot dataset.	155
Table P 7.	Comparison using SplattingAvatar (Shao et al., 2024) on the PeopleSnapshot dataset. ..	155
Table P 8.	Comparison using Animatable 3D Gaussians (Liu et al., 2024) on the GalaBasketball dataset.....	155
Table P 9:	The Core Visual Encoding System.....	180
Table P 10.	The Core Visual Encoding System.....	180
Table P 11.	The Visual Vocabulary for AR Markers	181
Table P 12.	The Visual Vocabulary for detected Instances	181
Table P 13.	The Visual Vocabulary for Augmented instances.	182
Table P 14.	The Visual Vocabulary for User Elements	182
Table P 15.	Studies with working memory tasks in VEs.	202
Table P 16.	Predicted success rate of target hitting and average error between predicted and actual success rates.....	216
Table P 17.	Core Features.....	223
Table P 18.	Summary of Pruning, Spherical Harmonics distillation and Vector Quantization quantitative results.	239

TABLE IN DEMOS

Table D 1.	Initial technical measurements of Z-3PO.....	260
-------------------	--	-----

Title	EuroXR 2025 Proceedings of the Application, Poster, and Demo Tracks of the 22nd EuroXR International Conference
Author(s)	Kaj Helin, Barbara Schiavi, & Electra Tsaknaki (eds.)
Abstract	<p>EuroXR 2025 Proceedings of the 22st EuroXR International Conference</p> <p>We are pleased to present these conference proceedings in the VTT Technology series, which contains the papers accepted for the Application Track and Poster & Demo Track of EuroXR 2025, the 22nd annual EuroXR conference, being hosted by ZHAW (School of Management and Law) from September 3rd to September 5th 2025, in Winterthur, Switzerland.</p> <p>This publication is thus a collection of the application papers (talks) and posters and demos papers (posters and demonstrations) presented at the conference. It provides an interesting perspective into current and future applications of VR/AR/MR.</p> <p>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) and Athens (2024). The focus of the EuroXR conferences is to present, each year, novel Virtual Reality (VR), Mixed Reality (MR) and Augmented Reality (AR) 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.</p> <p>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</p>
ISBN, ISSN, URN	ISBN 978-951-38-8804-6 ISSN-L 2242-1211 ISSN 2242-122X (Online) DOI: 10.32040/2242-122X.2025.T440
Date	August 2025
Language	English
Pages	318 p.
Name of the project	
Commissioned by	
Keywords	XR, virtual worlds, 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 2025

Proceedings of the Application, Poster, and Demo Tracks of the 22nd EuroXR International Conference

EuroXR 2025

Proceedings of the 22nd EuroXR International Conference

The 22nd EuroXR International Conference - EuroXR 2025 will be held on 3 to 5 September, 2025, organized by ZHAW in Winterthur, Switzerland.

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. Past locations include Bremen (2014), Lecco (2015), Athens (2016), Laval (2017), London (2018), Tallinn (2019), Valencia (2020), Milan (2021), Stuttgart (2022), Rotterdam (2023) and Athens (2024).

EuroXR 2025 will bring together individuals from research, industry, and commerce. Participants include technology developers, suppliers, and all those interested in Virtual Reality (VR), Mixed Reality (MR), Augmented Virtuality (AV), Augmented Reality (AR), and 3D user interfaces. The conference aims to facilitate the exchange of knowledge and experiences, showcase new results and applications, and feature live demonstrations of current and emerging technologies. It also provides a platform for forming collaborations for future work.

ISBN 978-951-38-8804-6
ISSN-L 2242-1211
ISSN 2242-122X (Online)
DOI: 10.32040/2242-122X.2025.T440



beyond the obvious