



EuroXR 2024

Proceedings of the 21st EuroXR
International Conference

Kaj Helin | Despina Michael-Grigoriou | Tina Katika
| Barbara Schiavi | Electra Tsaknaki (eds.)

EuroXR 2024

Proceedings of the 21st EuroXR International Conference

Kaj Helin (ed.)

VTT Technical Research Centre of Finland Ltd, Finland

Despina Michael-Grigoriou (ed.)

Cyprus University of Technology, Cyprus

Tina Katika (ed.)

Institute of Communication and Computer Systems (ICCS), Greece

Barbara Schiavi (ed.)

VINCI Construction, France

Electra Tsaknaki (ed.)

Independent Researcher, Greece

Päivi Vahala-Partanen

Technical editing & Graphical Design

VTT Technical Research Centre of Finland Ltd, Finland



ISBN 978-951-38-8796-4

VTT Technology 432

ISSN-L 2242-1211

ISSN 2242-122X (Online)

DOI: 10.32040/2242-122X.2024.T432

Copyright © VTT 2024

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 2024: Athens, Greece

The EuroXR 2024 conference will focus on cutting-edge advancements in Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) technologies. This includes software systems, display technologies, interaction devices, and various applications. Besides papers on the latest scientific results and highlights from many application fields, the EuroXR conference series aims at creating a unique human-dimension framework, interconnecting European and international XR communities, for knowledge cross-fertilisations between researchers, technology providers, and end-users.

27 to 29 November 2024

Athens, Greece

Conference organizers





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 2024, the 21th annual EuroXR conference, being hosted by ICCS (Institute of Communication and Computer System) from November 27th to November 29th 2024, in Athens, Greece. 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.

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-association.org/>) provides a unique platform for exchange between researchers, technology providers, and end users around commercial or research applications.

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

On behalf of the organising committee,



Kaj Helin



Kaj Helin

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

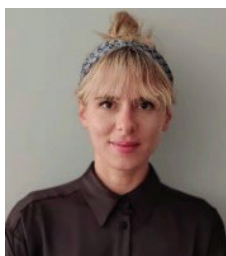


*Despina
Michael-Grigoriou*



Despina Michael-Grigoriou

*EuroXR Application track, Associate Professor at
Cyprus University of Technology, Cyprus*



Tina Katika



Tina Katika

*EuroXR Application track, XR Division Lead at
Institute of Communication and Computer Systems
(ICCS), Greece*



Barbara Schiavi

*EuroXR Poster & Demo Track, XR Research &
Development at VINCI Construction, France*



Electra Tsaknaki



Electra Tsaknaki

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

Application Program Chairs

- ▶ **Despina Michael-Grigoriou**, GET Lab, Cyprus University of Technology, Cyprus
- ▶ **Tina Katika**, Institute of Communication and Computer Systems, Greece
- ▶ **Armin Grasnack**, IU International University of Applied, Sciences, Germany
- ▶ **Domna Banakou**, New York University Abu Dhabi, UAE
- ▶ **Kaj Helin**, VTT, Finland
- ▶ **Jerome Perret**, Haption, France & Germany

Posters and Demos Chairs

- ▶ **Wolfgang Schäfer**, ZHAW School of Management and law, Switzerland
- ▶ **Electra Tsaknaki**, Independent Researcher, Greece
- ▶ **Kayvan Mirza**, President, Co-founder Optinvent, France
- ▶ **Khaled Sarayeddine**, CTO, Co-founder Optinvent, France
- ▶ **Barbara Schiavi**, VINCI Construction

General Chairs

- ▶ **Angelos Amditis**, ICCS, Athens, Greece
- ▶ **Benjamin C. Lok**, University of Florida, USA
- ▶ **Xubo Yang**, Shanghai Jiao Tong University, China

Scientific Program Chairs

- ▶ **Arcadio Reyes-Lecuona**, Universidad de Málaga, Spain
- ▶ **Gabriel Zachmann**, University of Bremen, Germany
- ▶ **Monica Bordegoni**, Politecnico di Milano, Italy
- ▶ **Jian Chen**, Ohio State University, USA
- ▶ **Giannis Karaseitanidis**, Institute of Communication and Computer Systems, Greece
- ▶ **Alain Pagani**, DFKI, Kaiserslautern, Germany
- ▶ **Patrick Bourdot**, CNRS & University Paris-Saclay, France

Organizing Team

- ▶ EuroXR – Patrick Bourdot, Mónica González, Jerome Perret, Gabriel Zachmann and Arcadio Reyes-Lecuona
- ▶ ICCS (Institute of Communication and Computer System) - Angelos Amditis, Ioanna Doutsou, and Iriini Krimpa.

Table of contents

APPLICATION TRACK

Comparison of Metaverse Strategies Worldwide	13
<i>Christoph Paul Runde</i>	
Point Cloud-Based Shared Virtual Spaces for Industrial Use Cases	23
<i>Jaakko Karjalainen, Petri Tikka, Taru Hakanen, Kaj Helin, and Johannes Peltola</i>	
Virtual Reality as a New Frontier For Professional Teacher Training: the VRTEACHER Application.....	27
<i>Kalliopi-Evangelia Stavroulia, Christos Kyrlitsias, Lefteris Ioannou, Giannis Georgiou, Despina Michael-Grigoriou, and Andreas Lanitis</i>	
The Mixed Reality System for Operation, AIT/AIV Support and Reporting in the Space Domain	31
<i>Kaj Helin, Jaakko Karjalainen, Paul Kiernan, and Gianluca Casarosa</i>	
LLM-Driven Human-Robot Interaction in eXtended Reality.....	35
<i>Kaj Helin, Vladimir Goriachev, Petri Tikka, Andrea Alesani, and Jaakko Karjalainen</i>	
Quantum Quadrant: Could We Use XR and AI Together in Education? Or Rather, Should We?.....	41
<i>Yücel Uzun, Andoni Retegui Schiettekatte, Suhana Biswas, and Florian Buchholz</i>	
NoFold: Bridging the Physical and Digital in Board Game Creation.....	45
<i>Christos Lougiakis, Lori Kougioumtzian, Giorgos Ganias, Anastasios Niarchos</i>	
From Stress to Success: Evaluation of a Comfortable VR Environment for Product Engineering	49
<i>Annegret Melzer, Philipp Stiens, Samuel Korb, Franziska Klimant, and Martin Dix</i>	
Real time facial capture and animation to explore the controls neurodiverse people need for their avatars	55
<i>Simon Campion, Radoslaw Dorociak, and Georg Meyer</i>	
AR Assistant for Pruning of Grapevines and Fruit Trees.....	61
<i>Mariia Podguzova, Simon Häring, Jamiu Ojeleye, Stephan Krauß, and Didier Stricker</i>	
A Simple but Effective AR Framework for Human-Object Interaction.....	67
<i>Hong Son Nguyen, Andrew Chalmers, DaEun Cheong, Myoung Gon Kim, Taehyun Rhee, and JungHyun Han</i>	
Using Augmented Reality and Machine Learning for Captioning in Theatrical Experiences	73
<i>Olga Chatzifoti, Nick Achilleopoulos, Spiros Borotis, Moonisa Ahsan, and Eleni Oikonomou</i>	
Cognitive Load in Mixed Reality: Technology Order and Prior Experience in Eye-Tracking	77
<i>Paula López, Ana M. Bernardos, and José R. Casar</i>	
Augmented telecommunication in factory setting.....	83
<i>Narek Minaskan, Bastian Kraye, Alain Pagani, and Didier Stricker</i>	
Bridging Realities: A Gap Analysis of XR Integration in Education and Training for Sustainable Future Skills	87
<i>Electra Tsaknaki, Anders Östman, Konstantinos Smagas, Marco Sacco, and Patrick Bourdot</i>	
XR technologies in the engineering sector – A gap analysis	91
<i>Anders Östman, Electra Tsaknaki, Patrick Bourdot, Son Dang, Aikaterini Stamou, and Vasilis Petrou</i>	

A Mixed Reality Eye-Tracking Study on Visual Influences in Dietary Choices.....	97
<i>Nikolaos D. Tantaroudas, Konstantinos Koukoudis, and Tina Katika</i>	
XR-based Human-Robot Collaboration along a Conveyor Picking Line	105
<i>Spyridon Nektarios Bolierakis, Kostas Naskou, Konstantinos Kokkalis, Fotios Konstantinidis, Tina Katika, Georgios Tsimiklis, Ioannis Karaseitanidis, and Angelos Arditis</i>	
The VESPA 2.0 System: a new frontier of Rehabilitation	109
<i>Emanuele Maria Merlo, Angelo Belfiore, Gian Paolo Donnarumma, Marco Salvatore Carmelo Falciglia, Francesco Generali, and Salvatore Marco Pappalardo</i>	
Power to the operators: Transdisciplinary co-design of XR applications in off-highway vehicle work domains.....	113
<i>Manuel Kulzer, Tanja Brodbeck, and Michael Burmester</i>	
CoVA : A Virtual Assistant for Virtual Reality Conferencing	117
<i>Alexis Lombard, Yazid Benazzouz, Galo Castillo López, Gaël de Chalendar, and Jean Pierre Lorré</i>	
POSTERS	
XR4Human: The Equitable, Inclusive, and Human-Centered XR Project	123
<i>Panagiotis Kavouras, and Vana Stavridi</i>	
LordofChange: A Haptic Sliding Controller with Changeable Tactile Textures.....	125
<i>Hongrun Wang, Tianrui Hu, Yuze Gao, and Wenjun Hou</i>	
The Unfaithful Copy: Performed AI “Personhood” in Mixed Reality Platforms.....	131
<i>Kathleen Bryson, Jessica Sutherland, and Kathleen Richardson</i>	
A pilot experiment on user individualization for sound localization within interactive virtual environments	
<i>Daniel Gonzalez-Toledo, Pablo Gutierrez-Parera, Carmen Soto-Romero, Maria Cuevas-Rodriguez, and Arcadio Reyes-Lecuona</i>	
Evaluation of content for the integration and education of virtual and augmented reality in the field of additive manufacturing.....	143
<i>Torsten Döhler, Moritz Heinrich, Christopher Kracht, Klarissa Koppusch, Fabian Grund, Frank Briesemeister, Eckart Wolf, Andreas Foitzik, Rene Krenz-Baath, and Andrea Böhme</i>	
Towards User-Centered XR Solutions for Smart Agriculture: A Work-in-Progress Design Approach..	147
<i>Alexandra Plexousaki, Asterios Leonidis, Maria Korozi, and Constantine Stephanidis</i>	
AR-enabled digital twin framework for automated shotcrete planning and operation	151
<i>Rehan Khan, Christian Martinez, Mohith Sunkara, and Rahul Tomar</i>	
Enhancing XR Accessibility: Automated Lip Reading and Captioning with Audio-Visual Speech Recognition Integration in Augmented Reality	155
<i>Hrishikesh Mulay, Sam Redfern, and Eleni Mangina</i>	
Psychometric Properties of the Greek Versions of VIMSSQ, FMS, and SSQ for Assessing Cybersickness.....	159
<i>Vasiliki Stefanouli, Dimitrios Diamantis, Eirini Cholopoulou, Chrysoula Lazou, Asimakis Kanellopoulos, Dimitris Iakovidis, and Nikolaos Strimpakos</i>	
eXtended Reality Content Prototyping through On-Device Solutions: An Experimental Platform based on Google’s MediaPipe Face-Tracking and Unity	163
<i>Stavroula Zoi, and George Caridakis</i>	

Dynamic City-Scale Digital Twins for Extended Reality.....	167
<i>Tariqul Islam, Jyothi Pudota, and Rahul Tomar</i>	
Voice-based Interactions in Immersive Applications – Cognitive and Affective Dimensions of User Experience	171
<i>Olga Chatzifoti, and Maria Roussou</i>	
Impact of group synchrony, postural stability and agency on cybersickness symptoms	175
<i>Julia Ayache, Camille Pistre, Pierre Jean, Anna Katharina Hebborn, Didier Stricker, and Benoît G. Bardy</i>	
Concept for a short-term training program to introduce the process of additive manufacturing in the space of Extended Reality (xR)	179
<i>Moritz Heinrich, Torsten Döhler, Christopher Kracht, Klarissa Koppusch, Fabian Grund, Frank Briesemeister, Eckart Wolf, Andreas Foitzik, Rene Krenz-Baath, and Andrea Böhme</i>	
A multimodal AR System: Enhancing Hiking Experiences.....	183
<i>Leyi Hong, Wenjun Hou, Qian Li, and Meiyu Lv</i>	
Immersive Cocooning – Competition between Control and Convenience	187
<i>Wolfgang G. Schäfer, and Gioia Volkmar</i>	
Architecture proposal of a VR operating procedure system for construction training based on BIM and QHSE rules.....	193
<i>Barbara Schiavi, Vincent Havard, Karim Beddiar, and David Baudry</i>	
XR and haptic feedback for electrician training: Ferrexpo case study	199
<i>Maryna Tymchenko</i>	
Beyond human imagination: The art of creating prompt-driven 3D scenes with Generative AI.....	201
<i>Giulio Federico, Fabio Carrara, Giuseppe Amato, and Marco Di Benedetto</i>	
WebAR for Construction Visualization	207
<i>Urs Riedlinger</i>	
Augmenting controls for tracked vehicles with force and tactile feedback: a work in progress	211
<i>Thomas Howard, Benjamin Geslot, and Jérôme Perret</i>	
Ethical impact of Artificial Intelligence in Extended Reality	215
<i>Irma Pöder</i>	
A Virtual Reality Evaluation Testbed for External Human-Machine Interfaces in Autonomous Vehicles	219
<i>Yiran Zhang, Guoqin Zang, and Cédric Bach</i>	
Exploring Boundary Visualization in AR for Co-located Collaborative Workspace Management	227
<i>Inoussa Ouedraogo, Huyen Nguyen, and Patrick Bourdot</i>	
AI-Augmented Mixed Reality for Hand Motion Assistance: Bridging the Gap in Parkinson’s Disease Rehabilitation	231
<i>Xinjun Li, and Zhenhong Lei</i>	
Integrating XR Technologies into Education: Bridging the Gap Between Educators & Developers.....	237
<i>Karoline Wirbatz</i>	
More Than One Milan: Multisensory Experience with an Olfactory Map Application.....	239
<i>Lai Pin Jia, Marco Rossoni, Marina Carulli, and Monica Bordegoni</i>	

Integrating Extended Reality (XR) in Teaching Sustainability: A Qualitative Study with Student Teachers across the Disciplines of STEM, Humanities, and Languages	243
<i>Rachel Farrell, Eileen Bowman, Marelle Rice, and Mark Baldwin</i>	
In- and outside Hololens 2 AR application for visualization of architecture 3D project data.....	249
<i>Lukas Koch, Christian Carthoff, Gerhard Wittfeld, and Ingrid Scholl</i>	
DEMOS	
"VoiVR": A Comfortable, (A-) Synchronous Collaborative VR Tool for Product Engineering.....	257
<i>Philipp Stiens, Samuel Korb, Annegret Melzer, Franziska Klimant, and Martin Dix</i>	
Pneumatics-Case: Enhancing Learning through Augmented Reality and Digital Twin Technology	261
<i>Rostyslav Boychuk, Ivan Symotiuk, Daniil Röbnikov, Vladimir Kuts, Kashif Mahmood, Simone Luca Pizzagalli, and Tauno Otto</i>	
Extended Reality simulator to evaluate remote control interactions for reach stackers.....	267
<i>Andrea Alesani, Kaj Helin, Vladimir Goriachev, and Timo Kuula</i>	
Industrial remote cooperation	273
<i>Bastian Krayer, Narek Minaskan, Alain Pagani, and Didier Stricker</i>	
The XR4ED Intelligent XR Tutor.....	277
<i>Fotis Liarokapis</i>	
Haptic Directional Awareness in Virtual Reality.....	281
<i>Jean Botev, Emanuel Dias da Silva, and Ningyuan Sun</i>	
An Open-Source Authoring Solution for Cultural Heritage Augmented Reality Applications.....	287
<i>Mathis Bodin, Jean-Philippe Farrugia, Fabrice Jaillet, and Bruno Tellez</i>	
Exploring the use of experiential marketing and extended reality technology to support the consumer learning process of sustainability policies and principles.....	291
<i>Rumbidzai Gandiwa, Rose Leahy, and Pio Fenton</i>	
5D/3D Virtual Reality for an effective Cognitive Rehabilitation: the VESPA 2.0 Project.....	295
<i>Emanuele Maria Merlo, and Salvatore Marco Pappalardo</i>	
UnityMol-Lite: Cross-platform Collaboration and Interactive Methods for Molecular Modeling in XR.....	301
Author Index	307
Tables of Figures.....	313
Lists of Tables.....	317

Application track



Comparison of Metaverse Strategies Worldwide

Christoph Paul Runde

Virtual Dimension Center (VDC), Fellbach, Germany

Corresponding author: christoph.runde@vdc-fellbach.de

Keywords: Metaverse, Virtual Worlds, Strategy, Economy, Global, Japan, China, South Korea, Finland, Europe, Dubai

SCOPE

The metaverse (or Virtual Worlds) is currently attracting a great deal of attention in business and society (Council of EU, 2022; European Parliament, 2022; Siemens, 2023). At the center of the often-controversial discourse is the next development stage of the internet in three-dimensional space. The metaverse is currently associated with virtual environments with always-on 3D experiences that combine physical and virtual reality using various technologies. This enables social interactions and economic transactions. The impact on the economy and society is considered to be substantial: corporate consultancies forecast metaverse markets with volumes of up to USD 394 billion by 2025 (BCG) (Bobier et al., 2022) and USD 5 trillion by 2030 (McKinsey & Company) (Elmasry et al., 2022). A metaverse study from 2023 by Ernst Young (EY) and NOKIA clearly shows (Douin et al., 2023 p. 28, Figure 5.1) that it will not only be the big tech companies (1st place) and existing metaverse platform (2nd place) that will drive future metaverse developments, but also industrial equipment suppliers (3rd place), manufacturers of enterprise software (4th place) and game developers (5th place). Such huge impact projections and various ways to contribute to and to benefit from the metaverse motivated a number of countries and the European Commission to work out respective metaverse strategies to position themselves (Kshetri, 2023). This paper compares these strategies, giving also an outlook on what to expect in the future.

Active countries and regions

A number of countries want to lead the way with dedicated metaverse strategies. Some of these are coordinated by state actors, while others are industry consortia.

China has published a three-year plan "Three-Year Action Plan for the Innovative Development of the Metaverse Industry (2023-2025) in China". Although China focuses on the industrial metaverse, applications in the entertainment industry are also mentioned. China wants to promote the Yuan metaverse ecosystem, new technologies and also the digital infrastructure. China's strategy aims to create global metaverse champions and metaverse celebrities. A secure and trustworthy industrial governance system shall be established. China has launched its own metaverse standardization

initiatives (MIT, 2023; Pascual, 2022; Liu and Ryder, 2022; Baptista, 2022; Kshetri, 2022; Hernandez, 2022; Knight, 2022; SCMP, 2022; Terrill, 2005; Kalathil, 2003, Yang, 2001).

POLITICO

Enter keyword



EXPLORE

NEWSLETTERS & PODCASTS

POLITICO PRO

FROM POLITICO PRO

Beijing is coming for the metaverse

Proposals reviewed by POLITICO show China wants to assert state control over virtual environments.



Experts that reviewed the proposal at the request of POLITICO said it risks violating principles of privacy and freedom | Jade Gao/AFP via Getty Images

BY GIAN VOLPICELLI

AUGUST 20, 2023 | 4:00 PM CET | 5 MINUTES READ

Figure 1. Politico.eu reports about China's metaverse ambitions in August 2023 (Volpicelli, 2023).

The **Dubai** Metaverse Strategy seeks to foster innovation, enhance the metaverse's economic contributions through R&D collaborations, and promote advanced ecosystems utilising accelerators and incubators. It also aims to foster talent and invest in future capabilities by providing the necessary support in metaverse education aimed at developers, content creators and users of digital platforms in the metaverse community. The Dubai Metaverse Strategy aims to turn Dubai into one of the world's top 10 metaverse economies as well as a global hub for the metaverse community. The strategy aims to build on Dubai's achievement of attracting more than 1,000 companies in the fields of blockchain and metaverse. It also promotes Dubai's ambitions to support more than 40,000 virtual jobs by 2030. This would further boost Dubai's economy and support the UAE government's vision of increasing the number of blockchain companies by five times the present number. (Dubai's Higher Committee, 2023;

Cabral, 2022; UAE Moments, 2022; Abdulmalak, 2022; Handagama, 2022; O'Malley, 2022; Al-Monitor, 2022; Hanafi et al., 2021)

The **European Union (EU)** published "*An EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition*". Core elements of Europe's strategy are to foster a supportive business environment and to support SMEs and start-ups; a talent pool of virtual world specialists is to be built; children are to be empowered and protected in virtual worlds. The EU further mentions to build its metaverse solutions based on interoperability, standardization, open standards and open interfaces. There shall be support for the open-source community. European solutions shall profit from the EU's robust legislative framework and European measures for IP protection. The goals of Europe's strategy are to boost the EU's technological capabilities and to accelerate the uptake of new business models. Europe seeks to support access to finance to support creators and to scale up innovative business models. Furthermore, societal progress shall be supported, and public services be improved (EC, 2023).

The "*Metaverse Initiative by the Finnish Ecosystem. Virtual Potential into Real-World Impact*" aims to establish **Finland** as a leading architect and enabler for the Metaverse. In addition, the Finland society shall be considered as a benchmark metaverse society. The Finish strategy shall attract major foreign direct investments in all parts of the Metaverse value chain. Finland's metaverse ecosystem expresses its will to avoid domination by a single company or country. The Finish metaverse strategy is based on a Finish vision, Finish values and strengths. Core elements of the Finish metaverse strategy are to boost economic growth, to foster cross-boundary collaboration, to support education for the Metaverse, to advance metaverse research, and to market Finish solutions on a global scale (DigitalFinland, 2023).

The **Japan** metaverse strategy is twofold. On the one hand side, many Japanese tech and finance giants, including Mizuho Financial Group, Sumitomo Mitsui Financial Group, and Mitsubishi UFJ Financial Group, have signed an agreement to form the Japan Metaverse Economic Zone. This metaverse, called Ryugukoku, will be operated by JP Games and will serve as a digital asset platform for member companies to share their technology with users. The platform will allow for the transfer of items and personal data between metaverses using unique identities and wallet services. On the other hand, the Japanese government considered what risks are expected from the metaverse, and how governments and private businesses should respond to these risks. In particular, the "*Study Group on Utilization of Metaverse, etc. for the Web3 Era*" established by the Ministry of Internal Affairs and Communications, and the "*Public-Private Partnership Conference on Dealing with New Legal Issues Concerning Content on Metaverse, etc.*" established by the Cabinet Office, examined a number of legal risks. Issues, among others, were: the relationship between the real world and users, the state of avatars (actions, identity in the virtual space), and the relationship between users and operators (JCB et al., 2023; Japan Ministry of Internal Affairs, 2023).

In **Saudi Arabia**, the metaverse is key part of nation's "*Vision 2030*". The metaverse also plays a valid role in the national "*Gaming and e-Sports Strategy*". The goal of Saudi Arabia is to grow the gaming and e-sports market to US\$ 13.3 billion of gross domestic product (GDP) and to 39.000 jobs by 2030. Saudi Arabia is investing into these businesses. One element of Saudi Arabia's metaverse strategy is to create a metaverse twin of the futuristic megacity NEOM (Ellyatt, 2022; Uddin, 2022; Al-Monitor, 2022; Middle East Eye, 2022; BeyondGamez, 2022; Robertson, 2022; Gideon, 2022; Nureldine, 2022; Panhans et al., 2021; Geschwindt, 2022; Fitch, 2020).

South Korea's ambitious plans included to take the upper hand in the global metaverse market (5th largest market share by 2026), to train metaverse experts (total of 40,000 people by 2026), to nurture metaverse service providers (220 companies by 2026), and to discover metaverse best practices (total of 50 cases by 2026). Key elements of the South Korean strategy to reach these goals are to nurture

specialized corporations that could lead the metaverse industry, train key players in the metaverse era, and to create and support a metaverse ecosystem. Finally, Korea aims to build a world-class metaverse platform; as lighthouse showcase, South Korea has chosen to create virtual Seoul as an exemplary metaverse world. Further showcases, marketing campaigns and diverse types of funds underline South Korea's ambitions. (MSIT, 2022; Davies and Jung, 2022, Cureton, 2022; Park, 2022a; Park 2022b).



MAGAZINE BY COINTELEGRAPH

FEATURES JOURNEYS CRYPTO CITY GUIDES 6 QUESTIONS HODLER'S

Inside South Korea's wild plan to dominate the metaverse

by Andrew Fenton | 11 min | December 8, 2022



South Koreans are obsessed with technology. More than half the population plays video games, and crypto adoption is high — both of which are very promising signs that its plan to dominate the Metaverse will succeed.

Share X Share Share Share

"Many years ago, it was AI. Now, it's metaverse," he says. "From the government's perspective, [...] as long as you don't have a coin itself, they're willing to support a lot of these new technologies" — Doo Wan Nam from StableNode

South Korea: The land of the metaverse

If you had to pick the one country that's most primed to take advantage of the opportunities offered by the metaverse, South Korea would be high on the list.

It's a technology-obsessed country that eagerly adopts new products, where 98% of

Figure 2. Magazine by Cointelegraph reports about South Korea's metaverse ambitions in December 2022 (Fenton, 2022)

Global Metaverse Strategies Comparison








	strategy goals	target groups	strategy key elements	discrete measures
 China	<ul style="list-style-type: none"> world-leading industrial mature metaverse ecosystem Metaverse as an important growth pole of the digital economy establish three to five companies with global influence establish a number of specialised SMEs establish three to five industrial development clusters establish yuan / renminbi universe secure and efficient governance system for the metaverse 	<ul style="list-style-type: none"> industry public administration society 	<ul style="list-style-type: none"> upgrade key technologies Promote an industrial ecosystem enable the industrial metaverse strengthen the industrial base (incl. standardization) 	<ul style="list-style-type: none"> metaverse business parks MV infrastructure MV standardization agencies create MV best practices create MV applications for digital life
 Dubai	<ul style="list-style-type: none"> turn Dubai into one of the world's top 10 metaverse economies turn Dubai into a global hub for the metaverse community attract more than 1,000 companies support more than 40,000 virtual jobs by 2030 foster innovation; promote advanced ecosystems foster talent invest in future capabilities create new governmental work models 	<ul style="list-style-type: none"> tourism education government services retail and real estate health care 	<ul style="list-style-type: none"> XR and MV technologies foster MV innovation and economic contribution cultivate MV talent through education and training develop MV use cases and applications in Dubai government Adopting and scale globally Dubai as the "crypto-capital" of the world UAE as a "test bed" for new technologies 	<ul style="list-style-type: none"> over fifteen initiatives and strategies were announced in the 2022 assembly MV virtual embassy Emirates Airlines training in the MV
 European Union	<ul style="list-style-type: none"> boost the EU's technological capabilities accelerate the uptake of new business models and solutions support access to finance to support creators and to scale up innovative business models support societal progress and improve public services 	<ul style="list-style-type: none"> industry creative industry society / consumers 	<ul style="list-style-type: none"> foster a supportive business environment; support SMEs, start-ups interoperability, standardisation, open standards and interfaces support open-source community build a talent pool of virtual world specialists empower and protect children in virtual worlds EU's robust legislative framework IP protection 	<ul style="list-style-type: none"> skills development for MV technologies (Digital Europe programme) Guiding principles for virtual worlds through citizens' panel research on impact of MV on people's health through Horizon Europe develop a MV toolbox for the general public create European initiative on Virtual Worlds create EU partnership on virtual worlds analysis of XR & MV standards landscape
 Finland	<ul style="list-style-type: none"> establish Finland as a leading architect and enabler for the Metaverse Finland to be considered as a benchmark society attract major foreign direct investments in all parts of the Metaverse value chain avoid domination by a single company or country 	<ul style="list-style-type: none"> society healthcare industry 	<ul style="list-style-type: none"> Finish vision, values and strengths boost economic growth cross-boundary collaboration education for the Metaverse marketing and sales on a global scale advance metaverse research 	<ul style="list-style-type: none"> Metaverse in Action Program (however just recommendations): <ul style="list-style-type: none"> support technology enablers create / support business Networks showcase Industrial Metaverse showcase Metaverse society showcase Metaverse healthcare create Metaverse economic zone „Ryugukoku“ media conference
 Japan	<ul style="list-style-type: none"> respond to rapidly growing demand for digital transformation (DX) for enterprises and government agencies enable interoperability and collaboration between Metaverse platforms serve as new social infrastructure for enterprise DX and employee experience (EX) transformation promote metaverse business expand the "Japan Metaverse Economic Zone" globally respond to MV risks 	<ul style="list-style-type: none"> society / consumers government services enterprises 	<ul style="list-style-type: none"> agreement of numerous companies using their respective mature technologies to create an open Metaverse infrastructure, supporting authentication, digital twins, NFTs, etc. analysis by the "Study Group on Utilization of Metaverse, etc. for the Web3 Era" established by the Ministry of Internal Affairs and Communications 	<ul style="list-style-type: none"> the US\$ 500 billion futuristic megacity NEDM shall have its own metaverse investment in metaverse gaming
 Saudi Arabia	<ul style="list-style-type: none"> gaming and e-sports market: US\$ 13.3 billion to gross domestic product (GDP) and 39,000 jobs by 2030 create digital MV twin of futuristic megacity NEDM 	<ul style="list-style-type: none"> e-games, e-sports real estate society / consumers 	<ul style="list-style-type: none"> MV is key part of Vision 2030 launch of national „Gaming and e-Sports Strategy“ 	
 South Korea	<ul style="list-style-type: none"> take the upper hand in the global metaverse market (5th largest market share by 2026) train metaverse experts (total of 40,000 people by 2026) nurture metaverse service providers (220 companies by 2026) discover metaverse best practices (total of 50 cases by 2026) 	<ul style="list-style-type: none"> society / consumers government services enterprises 	<ul style="list-style-type: none"> create world-class metaverse platform create and support MV ecosystem train key players in the metaverse era nurture specialized corporations that could lead the metaverse industry build an exemplary metaverse world 	<ul style="list-style-type: none"> metaverse platform of the city of Seoul diverse types of funds to be provided "Expanded Virtual World" marketing campaign K-Metaverse pavilion: Further showcases

Figure 3. global metaverse / virtual worlds strategies comparison

The United States of America did not develop a metaverse strategy until now. The reason can presumably be seen in the fact that the US government sees this as a task for the economy.

Comparison

The analyzed metaverse strategies differ greatly in terms of many criteria.

The topic of **skilled people and talents** is mentioned by China, Dubai, the EU, Finland and Korea, but not by Japan and Saudi Arabia.

Only in Japan and Finland was it a **consortium of companies** that formulated the strategy. In all other countries, it was state actors that drew up the strategy.

It is therefore to be expected that the **financing** of the implementation of the strategies will also be handled very differently. Although the Finnish metaverse ecosystem is planning the "Metaverse in Action Programme", it will not be able to raise any significant funds itself. In Japan, it is estimated that the large corporations involved will be able to do so. All other countries (China, Dubai, Saudi Arabia, Korea) are launching funding programs.

The structure of the funding programs depends directly on the **specific nature of the planned projects**. Japan is planning the Japan Metaverse Economic Zone "Ryugukoku", Korea with the Seoul Metaverse and Saudi Arabia with the NEOM Metaverse. Japan is still in the process of realizing a platform, while Korea and Saudi Arabia are already working on concrete applications in the contexts of urban planning, citizen participation, public services and tourism.

This also makes it clear that the **target groups** of the various metaverse strategies differ. All countries address the digital society, the consumers. All countries except Dubai and Saudi Arabia address industry. China, Dubai, Japan and Korea see public services in the metaverse. Dubai, Saudi Arabia and Korea are planning tourism applications in the metaverse. Dubai and Saudi Arabia see real estate applications in the metaverse. Dubai explicitly mentions retail applications, Saudi Arabia explicitly mentions e-sports and gaming applications.

China, Dubai and Japan address **payment systems** for the metaverse. China wants to promote the Yuan metaverse ecosystem. Dubai explicitly emphasizes its role as a global crypto stronghold. Japan envisions a digital asset platform that enables member companies to share their technology with users. For example, Mizuho Financial Group will provide its expertise on payments and metaverse tokens, and Mitsubishi UFJ Financial Group, which has its own digital asset platform, will support Web3 functions and global expansion plans.

Only the European and Chinese strategies explicitly mention **metaverse standardization**. Europe and China have already taken initial steps in this area. Europe is planning an analysis of the status of metaverse standardization and the relevant standardization system. After the topic of eXtended Reality was explicitly mentioned in the Chinese standardization strategy "China Standards 2035", the Chinese Ministry of Industry and Information Technology (MIIT) has announced the establishment of a "Metaverse Standardization Working Group".

Only the European Commission Metaverse / Virtual Worlds strategy provides for a later **review and adaptation of the strategy**.

References

- Abdulmalak, L. (2022) UAE Ministry of Economy Launches Office in metaverse. ITP. Accessed: Oct. 15, 2022, <https://www.itp.net/emergent-tech/uae-ministry-of-economy-launches-office-in-metaverse>
- Al-Monitor (2022). Saudi Arabia to invest billions in Metaverse, blockchain technology AI. Accessed: Oct. 15, 2022, <https://www.al-monitor.com/originals/2022/02/saudi-arabia-invest-billions-metaverse-blockchain-technology>
- Al-Monitor. (2022). UAE goes deeper into metaverse. Accessed: Oct. 15, 2022, <https://www.al-monitor.com/originals/2022/10/uae-goes-deeper-metaverse>
- Baptista, E. (2023). Analysis: A metaverse with Chinese characteristics is a clean and compliant metaverse. Reuters. Accessed: Oct. 15, 2022, <https://www.reuters.com/markets/funds/metaverse-with-chinese-characteristics-is-clean-compliant-metaverse-2022-01-25/>
- Bobier, J. F., Merey, T., Robnett, S., Grebe, M., Feng, J., Rehberg, B., ... & Hazan, J. (2022). The Corporate Hitchhiker's guide to the metaverse. Boston Consulting Group, 1-15. <https://web-assets.bcg.com/89/23/ae97d6e74e149e5951543b1202d3/the-corporate-hitchhikers-guide-to-the-metaverse-apr-2022-r.pdf>
- Cabral, R. (2022) Dubai plans to start new economic metric called 'gross metaverse product'. The National. Accessed: Oct. 15, 2022. <https://www.thenationalnews.com/business/technology/2022/09/28/dubai-plans-to-start-new-economic-metric-called-gross-metaverse-product/>
- China Ministry of Industry and Information Technology (MIIT). (2023): Three-year action plan for the innovative development of the Metaverse Industry (2023-2025) in China. Accessed: Nov. 15, 2023, www.gov.cn/zhengce/zhengceku/202309/content_6903023.htm
- Council of the European Union. (2022, March) Metaverse - Virtual World, Real Challenges. Council of the European Union, <https://www.consilium.europa.eu/media/54987/metaverse-paper-9-march-2022.pdf>
- Cureton, D. (2022). The welcome backs South Korea's metaverse plans. XR Today. Accessed: Oct. 15, 2022, <https://impactfestglobal.com/the-welcome-backs-south-koreas-metaverse-plans/>
- Davies, C., & Jung-a, S. (2022). Asia's largest metaverse platform Zepeto ramps up global expansion. Financial Times. <https://www.ft.com/content/14c88e84-f3c8-485e-a9df-31ead34e48f0>. Accessed: Oct. 15, 2022, <https://www.ft.com/content/14c88e84-f3c8-485ea9df-31ead34e48f0>
- Douin, V., Kindt, S., et al. (2023, June). The metaverse at work. Ernst Young EY/Nokia: London, UK, <https://www.nokia.com/metaverse/industrial-metaverse/the-metaverse-at-work-research/>
- Dubai's Higher Committee of Future Technology Development and Digital Economy (2022). Dubai Metaverse Strategy. Assembly Outcomes Report. Accessed Nov. 15, 2023, <https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/strategies-plans-and-visions/government-services-and-digital-transformation/dubai-metaverse-strategy>
- Ellyatt, H. (2018). Saudi's \$500 billion mega-city NEOM is attracting 'overwhelming' interest from investors. CNBC. Accessed: Oct. 15, 2022, <https://www.cnbc.com/2018/05/10/saudis-500-billion-mega-city-neom-is-attracting-overwhelming-interest-from-investors.html>
- Elmasry, T., Khan, H., Yee, L., et al. (2022) Value Creation in the Metaverse. The real business of the virtual world. McKinsey, <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/value-creation-in-the-metaverse>
- European Commission. (2023). An EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition. Accessed Nov. 15 2023, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3718
- European Parliament. (2022, June) Metaverse Opportunities, risks and policy implications. European Parliament, [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2022\)733557](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2022)733557)
- Fenton, A. (2022, December). Inside South Korea's wild plan to dominate the metaverse, accessed January 13, 2023; online: <https://cointelegraph.com/magazine/inside-south-korea-wild-plan-dominate-metaverse/>
- Finnish Metaverse Ecosystem (2023). Virtual Potential into Real-World Impact. Accessed Dec. 1, 2023, <https://www.digitalfinland.org/>
- Fitch, A. (2020). BLAST terminates deal with Saudi city NEOM. Esports Insider. Accessed: Oct. 15, 2022, <https://esportsinsider.com/2020/08/blast-neom-terminated>

- Geschwindt, S. (2022). Saudi Arabia's Neom invests \$1bn in AI and metaverse. Build in Digital. Accessed: Oct. 15, 2022, <https://buildindigital.com/saudi-arabias-neom-invests-1bn-in-ai-and-metaverse/>
- Gideon, M. (2022). What's next for Saudi Arabian esports? Business of Esports. Accessed: Oct. 15, 2022, <https://thebusinessofesports.com/2022/09/28/whats-next-for-saudi-arabian-esports/>
- Hanafi, M., Kshetri, K., & Sharma, R. (2021). Economics of artificial intelligence in the gulf cooperation council countries. Computer, 54 (12), 92–98, <https://doi.org/10.1109/MC.2021.3113094>.
- Handagama, S. (2022). Dubai unveils metaverse strategy, aims to attract over 1,000 firms. CoinDesk. 2022. Accessed: Oct. 15, 2022, <https://www.coindesk.com/policy/2022/07/19/dubai-unveils-metaverse-strategy-aims-to-attract-over-1000-firms/>
- Hernandez, O. (2022, July). Shanghai allocates \$1.5B to metaverse development fund. Blockworks. Accessed Oct. 15, 2022, <https://blockworks.co/shanghai-allocates-1-5b-to-metaverse-development-fund/>
- Japan Ministry of Internal Affairs and Communications. (2023). Interim summary of the study group on utilisation of the metaverse Towards Web3 era. Accessed Nov 15, 2023, https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2023/2/10_01.html
- JCB Co., Ltd., Mizuho Financial Group, Inc., Sumitomo Mitsui Financial Group, Inc., Mitsubishi UFJ Financial Group, Inc., Resona Holdings, Inc., Sampo Japan Insurance, Inc., Toppan Inc., Fujitsu Ltd., TBT Lab Inc. (2023). Creating an Open Metaverse Infrastructure from a Role-Playing Game Perspective. Accessed Nov. 15, 2023; <https://www.fujitsu.com/global/about/resources/news/press-releases/2023/0227-02.html>
- Kalathil, S. (2003). China's new media sector: Keeping the state in. The pacific review, 16(4), 489-501. <https://doi.org/10.1080/0951274032000132227>.
- Knight, O. (2022). Shanghai plans to cultivate \$52 B metaverse industry by 2025. Accessed Apr. 7, 2023) CoinDesk: <https://www.coindesk.com/business/2022/07/12/shanghai-plans-to-cultivate-52b-metaverse-industry-by-2025/>.
- Kshetri, N. (2022). Metaverse and developing economies. IT Professional, 24(4), 66-69. <https://doi.org/10.1109/MITP.2022.3174744>.
- Kshetri, N. (2023). National Metaverse Strategies. Computer. 56. 137-142. <https://doi.org/10.1109/MC.2022.3227681>.
- Liu, Z., & Ryder, B. (2022). China says no to speculative all-virtual metaverse. Jing Daily. Accessed: Oct. 15, 2022, <https://jingdaily.com/china-metaverse-real-economy-industrial-power/>
- MEE Staff (2022). Saudi Arabia's Neom invests \$1bn in Metaverse. Middle East Eye. Accessed: Oct. 15, 2022, <https://www.middleeasteye.net/news/saudi-arabia-neom-invests-1bn-metaverse>
- New America. The United States and the Metaverse. Accessed: Oct. 10th 2024, <https://www.newamerica.org/future-security/reports/the-forgotten-emerging-technology/the-united-states-and-the-metaverse/>
- Nureldine, F. (2022). With a gamer prince and oil billions, Saudi turns to eSports. AL-Monitor. Accessed: Oct. 15, 2022, <https://www.al-monitor.com/originals/2022/09/gamer-prince-and-oil-billions-saudi-turns-esports>
- O'malley, R.(2022). Dubai to become a global metaverse hub: Everything you need to know. Arabian Business. Accessed: Oct. 15, 2022, <https://www.arabianbusiness.com/gcc/uae/dubai-metaverse-strategy>
- Panhans, D., Joniškis, P., Tamer, F., & Saunier, F. (2021). Gaming & Esports: Media's next paradigm shift. BCG Global. Accessed: Oct. 15, 2022, <https://www.bcg.com/2021/gaming-and-esports-sector-are-the-next-shift-in-media>
- Park, D. (2022). Crypto Bull wins presidency: What it means for South Korea. Forkast. Accessed: Oct. 15, 2022, <https://forkast.news/crypto-bull-presidency-south-korea/>
- Park, D. (2022). South Korea's industrial city to reinvent itself in the metaverse. Forkast. Accessed: Oct. 15, 2022, <https://forkast.news/south-koreas-industrial-city-to-reinvent-itself-in-the-metaverse/>
- Pascual, M. G. (2022, September). China's Metaverse aims to use high-tech to suppress subversion. EL PAÍS. Accessed: Oct. 15, 2022. <https://english.elpais.com/science-tech/2022-09-29/chinas-metaverse-aims-to-use-high-tech-to-suppress-subversion.html>
- Rees, B.L. (2022). Saudi Arabia launches gaming and esports strategy. BeyondGames.biz. Accessed: Oct. 15, 2022, <https://www.beyondgames.biz/26876/saudi-arabia-launches-gaming-and-esports-strategy/>
- Robertson, D. (2022). Introducing the 'gross metaverse product. Politico. Accessed: Oct. 15, 2022, <https://www.politico.com/newsletters/digital-future-daily/2022/10/05/gross-metaverse-product-dubai-korea-00060514>

- SCMP (2022). Shanghai to pursue metaverse development as part of China's economic recovery efforts. South China Morning Post. Accessed: Oct. 15, 2022. <https://www.scmp.com/tech/policy/article/3184781/shanghai-ramp-support-metaverse-development-low-carbon-projects-smart>
- Siemens (2023). Siemens to invest €1 billion in Germany and create blueprint for industrial metaverse in Nuremberg metropolitan region, <https://press.siemens.com/global/en/feature/siemens-invest-eu1-billion-germany-and-create-blueprint-industrial-metaverse-nuremberg>
- South Korea Ministry of Science and ICT. (2023). Korea's Pan-Governmental Strategy on Metaverse. Accessed Nov 15, 2023, <https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&pageIndex=&bbsSeqNo=42&nttSeqNo=621&searchOpt=ALL&searchTxt=>
- Terrill, R. (2005). What does China want? *Wilson Quarterly*, 29(4), 50-61.
- UAE Moments (2022). UAE to use gross metaverse product (GMP) to measure economic success. Accessed: Oct. 15, 2022, <https://www.uaemoments.com/uae-to-use-gross-metaverse-product-gmp-to-measure-economic-success-509680.html>
- Uddin, R. (2021). Saudi Arabia launches 170km city built in a straight line. *Middle East Eye*. Accessed: Oct. 15, 2022, <https://www.middleeasteye.net/news/saudi-arabia-the-line-neom-megacity>
- Volpicelli, G. Beijing is coming for the metaverse (2023, August), <https://www.politico.eu/article/china-beijing-designing-metaverse-proposal-social-credit-system-un-itu/>
- Yang, D. L. (2001). The great net of China. *Harvard International Review*, 22(4), 64.

Point Cloud-Based Shared Virtual Spaces for Industrial Use Cases

Jaakko Karjalainen, Petri Tikka, Taru Hakanen, Kaj Helin, and Johannes Peltola

VTT Technical Research Centre of Finland Ltd, Espoo, Finland

Corresponding author: jaakko.karjalainen@vtt.fi

Keywords: XR, Point Clouds, Multi-Platform, Multi-User, Unity 3D

Introduction

Extended Reality (XR) technologies offer significant benefits for industrial applications by enhancing visualization, training, and operational efficiency. By encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), these technologies enable immersive simulations and real-time data overlays, facilitating improved decision-making and reduced operational risks. For instance, VR can simulate remote and hazardous environments for training purposes without requiring travel or exposing individuals to physical danger, while AR can overlay critical information onto real-world equipment, aiding in maintenance and repair tasks. The growing adoption of XR in industries such as manufacturing, construction, healthcare, and public safety underscores its potential to streamline processes, enhance safety, and foster innovation.

Creating virtual representations of physical spaces for VR can be achieved in multiple ways, including the use of CAD-based 3D models or other pre-existing 3D data. However, when pre-existing data is unavailable, the accurate modelling of the space is often considered too laborious, making the use of VR technologies unfeasible.

This study focuses on the multi-platform utilization of point clouds as the common ground representations of physical spaces. Point clouds enable the capture of detailed physical environments efficiently, facilitating the creation of accurate and dynamic virtual spaces. This approach is particularly advantageous for environments that undergo frequent changes, such as construction sites.

This paper presents implementation of a Unity 3D-based XR application platform supporting multiple device platforms, ranging from VR to AR, with multi-user collaboration capabilities. The platform's versatility and robustness make it suitable for various industrial applications, demonstrating the practical benefits and scalability of point cloud-based virtual spaces.

Methodology

Point clouds can be captured using various techniques, such as photogrammetry and LiDAR. In this study, we utilized a Trimble X7 3D laser scanning system. The Trimble X7 captures an accurate 3D point cloud, supplemented with RGB data from the device's cameras, resulting in a naturally coloured, 1:1 3D point cloud of the environment with millimetre accuracy. The captured data was processed using Trimble Real Works software, where unwanted artifacts were removed, and the point cloud was subsampled to a resolution optimal for XR development.

The XR development was conducted in the Unity 3D environment. Point cloud data was imported using the Pixyz Plugin for Unity, which facilitated the automatic segmentation of the point cloud model, generation of level-of-detail (LOD) groups, and creation of a simplified 3D mesh to serve as a collider. The coordinate system of the imported point cloud was maintained as the ground truth for all multi-platform data.

Achieving multi-platform compatibility between Windows desktop-based VR and AR devices (Apple Vision Pro, Magic Leap 2, and Microsoft HoloLens 2) was accomplished using the Unity 3D platform, supplemented with additional libraries that consolidate multiple lower-level platform-specific libraries under a shared code base. The Unity XR Interaction Toolkit was crucial in enabling consistent user interface elements and interactions across multiple platforms.

OpenXR was employed for desktop VR, allowing support for various VR devices, though it limited the use of certain platform-specific functionalities on devices like Varjo XR3. For AR, where OpenXR support was more restricted, the implementation relied on platform-specific libraries combined with Unity AR Foundation libraries.

The presented implementation includes basic multi-user capabilities built on the Unity NetCode for GameObjects library. The Dissonance Voice Chat plugin was used to enable voice chat functionality, facilitating real-time communication among users within the virtual space.

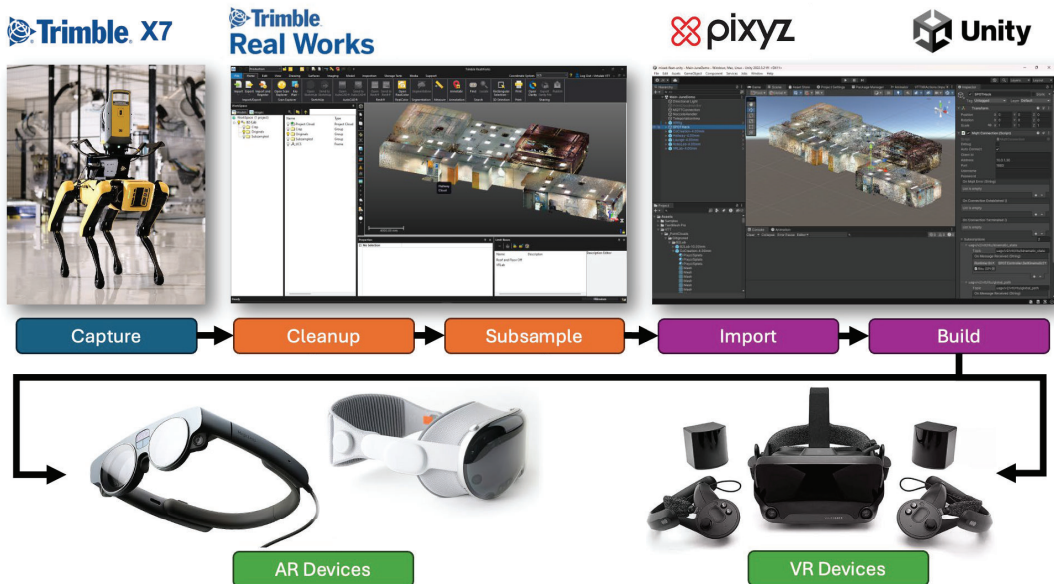


Figure 4. Point cloud capturing, processing and visualization pipeline

Multi-User Collaboration

The aim was to study the multi-user collaboration in industrial setting utilizing different XR technologies. The core concept was built upon the scenario of a worker on the field using AR supported by a remote expert operating in VR. VR environment contained the point cloud representation of the physical space the AR operator was working in. AR operator is unconcerned about the point cloud visualization.

Both VR and AR operators shared various types of 3D spatialized data. Users were represented in the 3D space as simplified avatars that displayed the position and orientation of their heads and hands. This setup allowed users to see each other's movements and communicate in real-time. The application also supported the placement of various types of annotations, including different symbols and both basic 2D and 360-degree photographs. These features provided the necessary components for basic telepresence and cooperation.

The study included also the visualization of different types of mobile robots in real-time. A Boston Dynamic SPOT robot was visualized with 1:1 mapping based on the detailed state data offered by the platform APIs. Other robots were visualized based on data provided by Noccela ultra-wideband (UWB) tracking system visualized as simple 3D models or avatar symbols.

Multi-Use Applications

The developed multi-user collaboration framework was tested in industrial use-cases with expert end-users from various fields including automotive industry, aircraft maintenance, crime scene inspection and mobile robotics. The core framework was kept the same and the use case specific adaptation was made by using different user interfaces best suited for the case at hand.

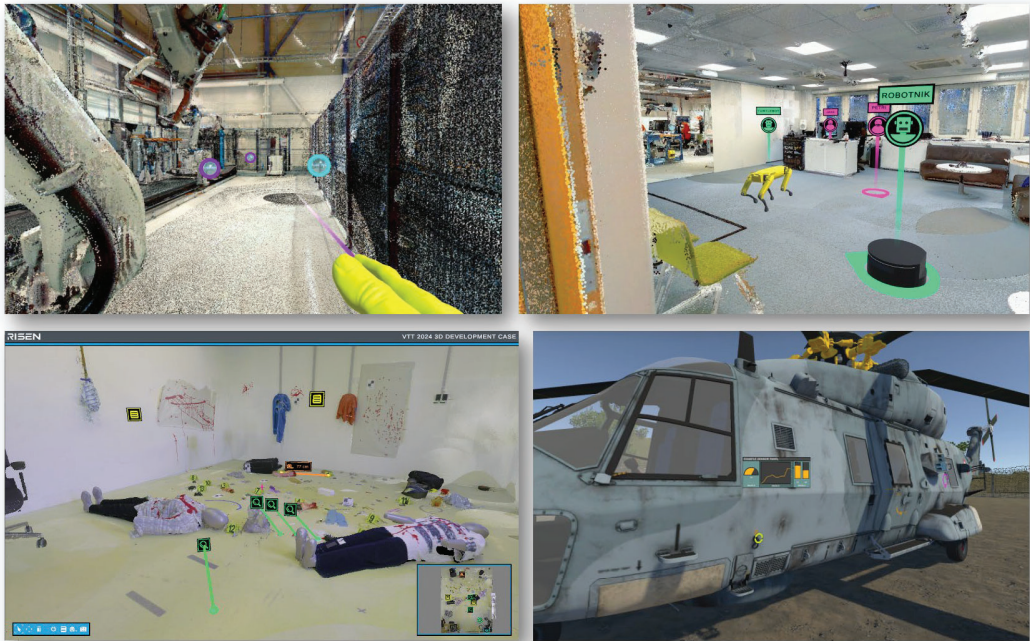


Figure 5. Tested application use cases. Top left: robot cell evaluation for automotive industry, top right: cooperation with mobile robots, bottom left: crime scene inspection and bottom right: aircraft maintenance.

Conclusion

The multi-user collaboration framework developed in this study significantly enhances the sense of presence, situational awareness, and operational coordination in industrial settings. This framework demonstrates the potential of XR technologies to improve industrial workflows through immersive and interactive virtual environments. The point cloud processing pipeline proved effective for creating detailed 3D visualizations of various industrial environments within Unity 3D. Point cloud-based visualization offered a higher level of environmental detail compared to CAD-based visualizations and provided up-to-date representations of dynamic environments.

This study also showcased the practicality and scalability of using point cloud data to support multi-platform XR applications. The implementation of basic multi-user capabilities and real-time visualization of mobile robots highlighted the versatility of the developed platform in various industrial use cases, including robot cell evaluation, aircraft maintenance, crime scene inspection, and cooperation with mobile robots in indoor environments.

Future directions include exploring near real-time point cloud visualization, enabled by advancements in scanning and visualization technologies, and further improving multi-platform support through initiatives like OpenXR. These advancements will continue to drive the adoption of XR technologies in the industrial sector, fostering innovation and operational efficiency.

Acknowledgments

This study has received funding from research project Mixed Fleet funded by Business Finland, VTT and the participating companies.

This study has also received funding research project RISEN funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883116. 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.

Virtual Reality as a New Frontier For Professional Teacher Training: the VRTEACHER Application

Kalliopi-Evangelia Stavroulia^{1,2}, Christos Kyriltsias², Lefteris Ioannou²,
Giannis Georgiou², Despina Michael-Grigoriou², and Andreas Lanitis^{1,2}

¹*CYENS Centre of Excellence, Nicosia, Cyprus*

²*Cyprus University of Technology, Limassol, Cyprus*

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

Keywords: VRTEACHER, Virtual Reality, Teacher Training

Introduction

This paper presents a novel Virtual Reality (VR) training approach for teacher empowerment using a VR training application developed under the VRTEACHER project. The VRTEACHER project was conceived to address the urgent need for innovative approaches to teacher training within the constantly evolving educational landscape, the unprecedented challenges for educators brought by the pandemic worldwide and the lack of hands-on training experiences for teachers. The project aimed to revolutionize teacher education and strengthen the profile of the teaching profession by offering an alternative training paradigm based on VR technology, offering immersive training experiences for educators with virtual scenarios that replicate real-world classroom situations as documented by teachers and education experts through focus groups. Emphasis was given to teachers' empathy-building through the ability to change the user's perspective and experience the same incident both from the eyes of the teacher and from the eyes of students, enabling trainee teachers to empathize and understand their student's viewpoint. Experiencing the scenarios from the student's viewpoint was crucial for the effectiveness of the VR training fostering teacher trainees' emotional connection and understanding increasing their empathy (Hadjipanayi et al. 2024) and compassion (Ventura et al., 2020; Stavroulia & Lanitis, 2023a; Wang et al., 2024).

VRTEACHER adds a unique dimension to teacher education, offering a hands-on, experiential training tool that complements theoretical knowledge and classroom practice. By pushing the boundaries of traditional teacher training and leveraging the potential of VR technology, VRTEACHER contributes to the ongoing innovation and advancement of teacher education practices. More information and the application are available on the project's official website <https://www.vrteacher.eu/>.

VRTEACHER application

Technical specifications

The VRTEACHER multilingual application (in English, Greek and Spanish) runs on smartphone VR headsets in conjunction with any Android mobile device with Android version 7.0 (API level 24) or newer and requires 604 MB of device storage. The Android phone should also meet the minimum requirements for VR which mainly include the existence of a gyroscope sensor. The use of smartphone-based VR headsets minimizes the hardware cost, allowing the widespread deployment and use of the application. The application was developed using UNITY (version 2020.3.22).

The main menu of the VRTEACHER application consists of four sub-categories (Figure 6) “Scenarios”, “Instructions”, “About the VRTEACHER project” and “Settings”. By pressing the “scenarios” tab users are instructed to insert their smartphone in a VR headset so that they can visualize the three scenarios that have been developed.



VRTEACHER

Figure 6. (i) VRTEACHER application main menu. (ii) Settings tab. (iii) Scenarios tab. (iv) Instructions to place the phone into the VR headset.

Scenario cases

The VR application offers three scenarios (Figure 7), co-developed with educators who were involved in the design cycle. The user can experience each scenario both from the teacher's perspective and the virtual student's perspective (Figure 8: 1-7), gaining a deeper understanding of the problem and the student's viewpoint. Additionally, the inner voice technique was used to refer to teachers' internal thoughts and reflections during the time that the incidents occurred in class to enhance the user's immersive experience (Figure 8: 1). At the end of the scenario a mentor appears providing feedback to the teacher trainee (Figure 8: 8).

Scenario 1 Distance education and domestic verbal abuse	Scenario 2 Student phobias related to COVID and panic attack	Scenario 3 Refugee Student
<ul style="list-style-type: none"> The scenario simulates a student who is connected to the lesson by distance and experiences verbal abuse from his mother at home. 	<ul style="list-style-type: none"> The scenario simulates a student who is experiencing a panic attack and anxiety related to COVID-19. 	<ul style="list-style-type: none"> The scenario simulates a student who does not speak the instruction language and does not understand the teacher.

Figure 7. The three scenarios developed under the VRTEACHER project.



Figure 8. (i) Through the eyes of the teacher for all three scenarios and teacher's inner voice. (ii) Student avatar in scenario 1. (iii) Through the eyes of the student in scenario 1, in his room with his mother. (iv) Student avatar in scenario 2. (v) Through the eyes of the student seeing particles of COVID-19 in the class. (vi) Student avatar in scenario 3. (vii) Through the eyes of student refugee. (viii) Avatar of a mentor providing feedback to the trainee.

Interaction within the virtual classroom is gaze-based. Gaze input indicates where the user is focusing, with a cursor indicating the direction of the gaze. Eye gaze input requires minimal effort from the user, as it enables interaction with the virtual world without the need for physical movements or additional hardware. During the VR experience, users are required to use eye-gaze to initiate actions (Figure 9:1) and/or answer multiple-choice questions (Figure 9:2,3) that assess teachers' readiness in appropriately handling various situations depicted in the implemented scenarios.

VRTEACHER Application Impact

The VRTEACHER application was implemented across the partner countries training more than 300 in-service and pre-service teachers (Stavroulia et al., 2023b). The implementation process followed an experimental protocol that was approved by National Bioethics committees in the respective partner countries. The results demonstrated that the VR implementation significantly improved participants' attitudes towards remote and blended learning, inclusive education, and cultural intelligence. Metrics were assessed at three points: before implementation, immediately after, and long-term, four weeks post-implementation. The results also demonstrated a statistically significant difference between the three points related to perspective change. Equally important, the results indicated that the VR intervention had a long-lasting impact on the participants, as the positive change in attitude not only persisted over time but continued to improve.

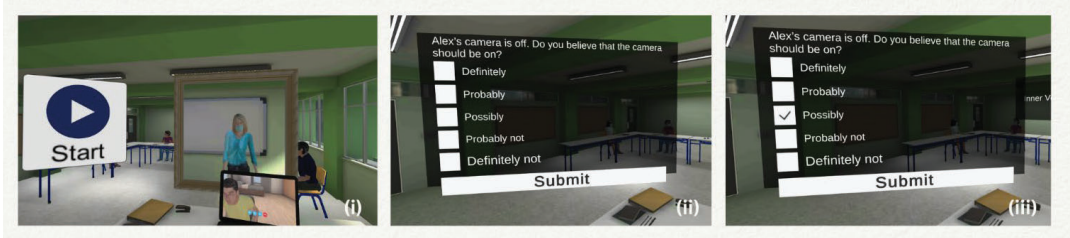


Figure 9. (i) Start button to begin the scenario. (ii) Unanswered question during the scenario. (iii) Selection of answer after looking for some minutes.

Conclusions

The VRTEACHER project addressed the need for novel technology-based hands-on training experiences in teacher education contributing to advancing innovative VR-based teaching practices. The VR application was implemented and evaluated providing evidence of its added value as part of teacher training curriculum. The VRTEACHER application is accessible for use, aiming to inspire new initiatives in the field. Furthermore, this work is ongoing, and the team aims to update the application in line with current trends such as Artificial Intelligence (AI) for more dynamic and adaptive experiences that respond to user behavior and enrich it with new scenarios reflecting real-life in-class incidents experienced by educators.

Acknowledgement

This project was supported by the Erasmus+ programme of the European Union through the project VRTEACHER Virtual Reality-based Training to improvE digitAl Competences of teachERs, Grant Agreement number: 2020-1-CY01-KA226-SCH-082707.

References

- Hadjipanayi, C., Christofi, M., Banakou, D., & Michael-Grigoriou, D. (2024). Cultivating empathy through narratives in virtual reality: a review. *Personal and Ubiquitous Computing*, 1-13. <https://doi.org/10.1007/s00779-024-01812-w>
- Stavroulia, K.E. & Lanitis, A. (2023a). The role of perspective-taking on empowering the empathetic behavior of educators in VR-based training sessions: An experimental evaluation. *Computers & Education*, 197, p.104739. <https://doi.org/10.1016/j.compedu.2023.104739>
- Stavroulia, K.E., Pyrini, A., Álvarez Testillano, I., Keegan, S., Debattista, M., Zarraonandia Ayo, T., Roussias, C., Sanz Prieto, M., Díaz Pérez, P., & Lanitis, A. (2023b). VRTEACHER project: a revolutionary approach to teacher education using virtual reality. *Proceedings of ICICTE 2023*.
- 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), pp.667-676. <https://doi.org/10.1089/cyber.2019.0681>
- Wang, Y., Chen, C., Nelson, M.R., & Sar, S. (2024). Walk in my shoes: How perspective-taking and VR enhance telepresence and empathy in a public service announcement for people experiencing homelessness. *New Media & Society*, 26(7), pp.3931-3950. <https://doi.org/10.1177/14614448221108108>

The Mixed Reality System for Operation, AIT/AIV Support and Reporting in the Space Domain

Kaj Helin¹, Jaakko Karjalainen¹, Paul Kiernan², and Gianluca Casarosa³

¹*VTT Technical Research Centre of Finland Ltd, Tampere, Finland*

²*SKYTEK Ltd, Dublin, Ireland*

³*ESA – European Space Research and Technology Centre (ESTEC), Noordwijk, The Netherlands*

Corresponding author: kaj.helin@vtt.fi

Keywords: Mixed Reality, Space Domain, HoloLens 2, mobiPV, Assembly, Integration, Testing and Verification

Introduction

This paper presents an approach towards digitalizing and enhancing operational procedures within the Test Centre of the European Space Agency (ESA) by exploiting Mixed Reality (MR) systems. The first phase, known as "VirWAIT - Virtual Workplace for AIT & PA Training and Operations Support", focuses on developing the VirWAIT MR system, which was built on the MS HoloLens 2 mixed reality platform. The second phase, the project called "DPIAR-V1 - Digitalisation of Procedure and Introduction of Augmented Reality (Step 1)", [represents a continuation of the ESA's MR use cases and development. The project outcome was a complete end to end solution for the authoring (off- and on-site), deployment and usage of MR to support execution of manual procedures and operations within the Test Centre of the European Space Agency.

The first phase - VirWAIT

The operational VirWAIT MR-systems includes three main parts (Helin et al., 2021):

1. Content authoring
 - a. Desktop mode (off-site authoring)
 - b. Mixed reality mode (on-site authoring)
2. Mixed reality based procedure execution
3. Automatic reporting and as-build 3D model generation

VirWAIT's Use cases

The system were tested in two main use cases at the Test Centre premises:

1. MR supported installation of thermocouples on an Heat Plate used as GSE for the Solar wind Magnetosphere Ionosphere Link Explorer – SMILE (see Figure 10 - left)
2. MR supported phase 2 sensor installation on TEDY (TEst Dummy) for a vibration test campaign on the Hydra facility (see Figure 10 - right)

The VirWAIT MR system was preliminary tested for the configuration of the JUpiter ICy moons Explorer's (Juice) Network Data Interface Unit (NDIU) in preparation to the Thermal Vacuum test campaign held at ESTEC.



Figure 10. VirWAIT use cases. Left: Thermocouples installation to SMILE. Right: Sensors installation to TEDY.

The second phase - DPIAR-V1

Building upon the success of VirWAIT MR, the second phase, the project called DPIAR-V1, represents a continuation of the ESA's MR use cases and development. The project outcome was a complete end to end solution for the authoring (off- and on-site), deployment and usage of MR to support execution of manual procedures and operations within the Test Centre of the European Space Agency. The operational DPIAR-v1 MR-systems and authoring components can be found in Figure 11. System includes:

1. Microsoft HoloLens 2 with DPIAR-v1 MR-player app
2. mobiPV – server for all Operations Data File (ODF) content, AR annotation and 3D models with animation
3. mobiPV's web interface, which also allows to the user interact with systems
4. STAMP sensor data server
5. Authoring environment, which is including on-site and off-site authoring.
6. Remote observation via web access
7. Reporting

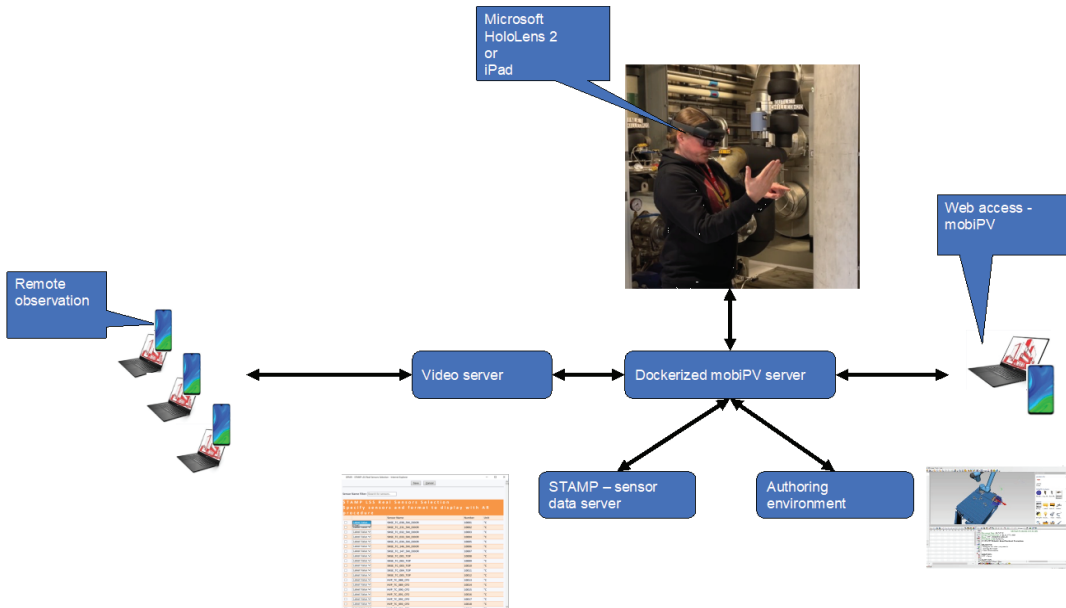


Figure 11. DPIAR-v1 system set-up

DPIAR-v1's Use cases

To verify the main project objectives, two most relevant use cases were selected. They were use cases at the Test Centre premises.

- 1) Large Space Simulator – LSS Basement procedure (see Figure 12)
- 2) Vacuum Test Chamber - VTC1.5 Operating Procedure and Pre-operation

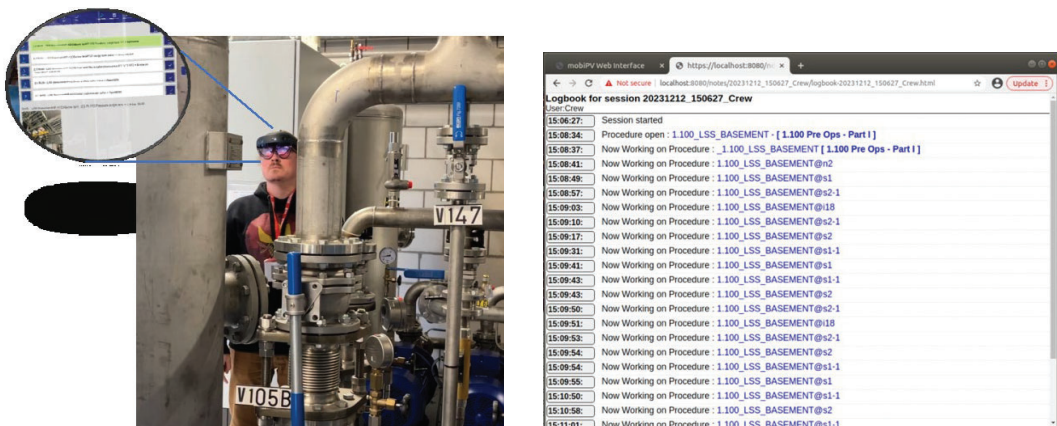


Figure 12. The MR systems in use during procedure execution (LSS- Basement procedure) and its automatic reporting.

Both use cases were authored by leveraging off-site and on-site authoring. At the same time, the semi-automatic off-site authoring templates were utilized and further developed. On-site authoring was employed to add the point of interests (POIs) in the correct locations due to the lack of 3D models. The

authoring process was carried out in collaboration with partners and ESA personnel. The MR system, remote observation and reporting was tested, particularly in the LSS Basement procedure.

Conclusion

The DPIAR-v1 system offers a notable advantage due to its optimized operational workflow, which enables the reduction of the MR procedure authoring time. This is achieved through a combination of semi-automated off-site procedure template and intuitive on-site authoring methods via HoloLens 2. These features streamline the preparation and modification of procedures, thereby playing a crucial role in facilitating the deployment of MR technology in daily operational activities.

Through the continuation by VirWAIT and DPIAR-V1 projects, the Test Centre Division of the European Space Agency highlights its strong interest in applying MR technology to maintenance and operation procedures of Test facilities, but also to training of operators and where possible to AIT (assembly, integration and testing) tasks. Videos documenting the development and outcome can be found in the Appendix.

Acknowledgment

This study has been funded by ESA under two contracts: 4000129549/19/NL/BJ "Virtual Workplace for AIT & PA Training and Operations Support" & 4000140015/22/NL/MG "Digitalisation of Procedure and Introduction of Augmented Reality (Step 1)"

References

SMILE, Solar wind-Magnetosphere-Ionosphere Link Explorer, <https://www.cosmos.esa.int/web/smile>, last accessed 2024/06/21.

JUICE, Jupiter Icy Moons Explorer, https://www.esa.int/Science_Exploration/Space_Science/Juice, last accessed 2024/06/21.

Helin, K., Karjalainen, J., Kuula, T., Kiernan, P., Casarosa, G., & Oliveira, D. M. (2021). *The Mixed Reality System for AIT/AIV in Space Domain – The Remote User Test of the First Integrated Prototype*. Abstract from 4th International Conference of Aerospace and Aeronautics, Valencia, Spain.

Helin, K., Karjalainen, J., Kiernan, P., & Casarosa, G. (2023). Digitalisation of Procedure and Introduction of Augmented Reality. Abstract from ESA- AR/VR for Space Programmes 2023, Noordwijk, Netherlands.

APPEDIX

Video during VirWAIT development (2021): https://youtu.be/gefrH8EJXWU?si=2B7dchjDbz44T_3i

Video of DPIAR-v1 (2024): https://www.linkedin.com/posts/kaj-helin-8523a62_spacetech-xr-ar-activity-7157658812429086721-nABe?utm_source=share&utm_medium=member_desktop

LLM-Driven Human-Robot Interaction in eXtended Reality

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

VTT Technical Research Centre of Finland Ltd, Finland

Corresponding author: kaj.helin@vtt.fi

Keywords: LLM, HRI, XR, Assistant

Introduction

This abstract presents a proof of concept (PoC) of Large Language Models (LLMs)-driven Human-Robot Interaction (HRI) in eXtended Reality (XR). Our LLM assistant is capable of holding multimodal conversations with users through voice, gestures, emoji reactions, and Mixed Reality (MR) hints, all while simultaneously communicating with the UR10e cobot. This to make it assist the user in the right way at the right time. As this is just a proof of concept, the underlying architecture holds great potential and can be replicated in several industrial use cases. While big industry players focus on enhancing raw LLM performance, our goal is to make these models more context-aware, particularly in real-world applications. The specific task chosen for this PoC is the assembly of an IKEA step stool, which consists of 8 steps. The system guides the user through each step, providing real-time assistance to ensure accurate and efficient assembly. We chose IKEA step stool assembly as the use case because there is a variety of data and 3D models available online, and many users are already familiar with IKEA products. Appendix includes video of the proof of the concept.

System architecture

The core application was developed in Unity 2022.3 LTS. It makes use of OpenAI's APIs to communicate with GPT-4o, while a traditional TCP socket communication is used to communicate with the UR10e collaborative robot (cobot). The application was built to run on the Microsoft HoloLens 2 platform, and it makes use of its AR features, Wi-Fi and microphone peripherals (see Figure 13).

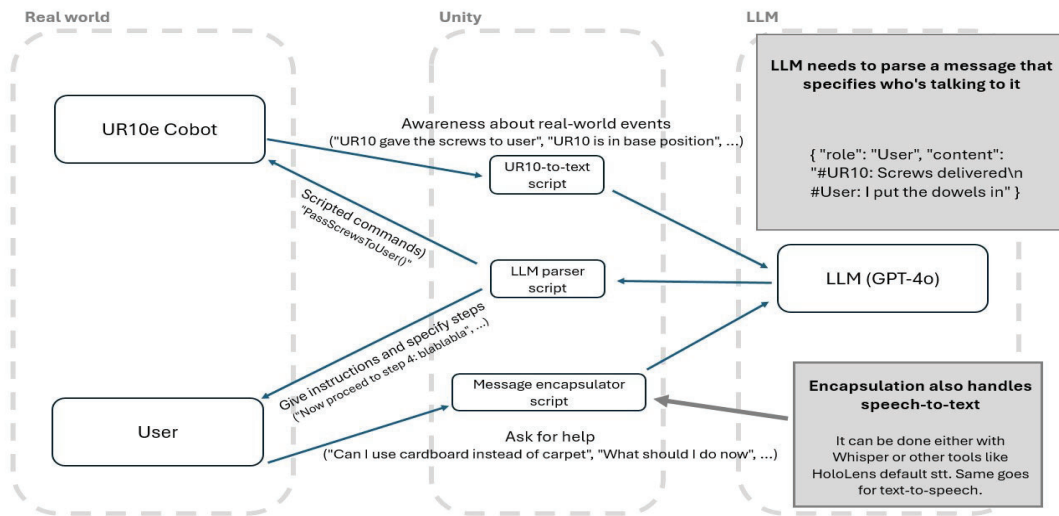


Figure 13. System architecture of the application. The communication to the LLM is mediated by parser scripts that convert different types of input (i.e., from TCP sockets or microphone) into conversational text. Likewise, another parser script takes the LLM's output and redirects the different messages to the right recipients.

LLM-based assistant

Voice assistants such as Alexa and Siri have demonstrated how a software agent can interpret natural language and provide information through voice generation. However, general voice assistants have limited capabilities to follow or provide multi-step instructions. Large-language models provide the means to enhance Natural Language Processing (NLP) and enable more thorough insight on the subject especially if the initial system message or context is provided. (Guan, Y. 2023)

Interaction design

The presented application is based on multimodal communication. The virtual assistant is able to guide the user through the procedure of assembling a piece of furniture by voice instructions (through OpenAI's text-to-speech APIs), as well as AR hints and emoji reactions which are described in more detail in the next sections. On the other hand, the assistant can also receive multiple types of input, namely voice clips from the user and message codes from the UR10e cobot, that are converted into plain, natural text before being fed to the LLM (see Figure 14).

One major limitation of LLMs is their reactive nature - they only respond when prompted, lacking proactive assistance. To overcome this, the application was developed modularly, allowing different entities to communicate with the LLM, enabling it to engage with users without being directly prompted. Research (Sonnino, 2024) showed GPT-4 could handle conversations with multiple entities by parsing prompts with a "sender" and a message field. In this prototype, the cobot is the only added entity, but future expansions could involve agents like sensors, computer vision, or IoT devices. Our results show that multi-entity conversations are feasible, with the LLM generating multiple responses for different entities (robot, user, emoji) within the same message.

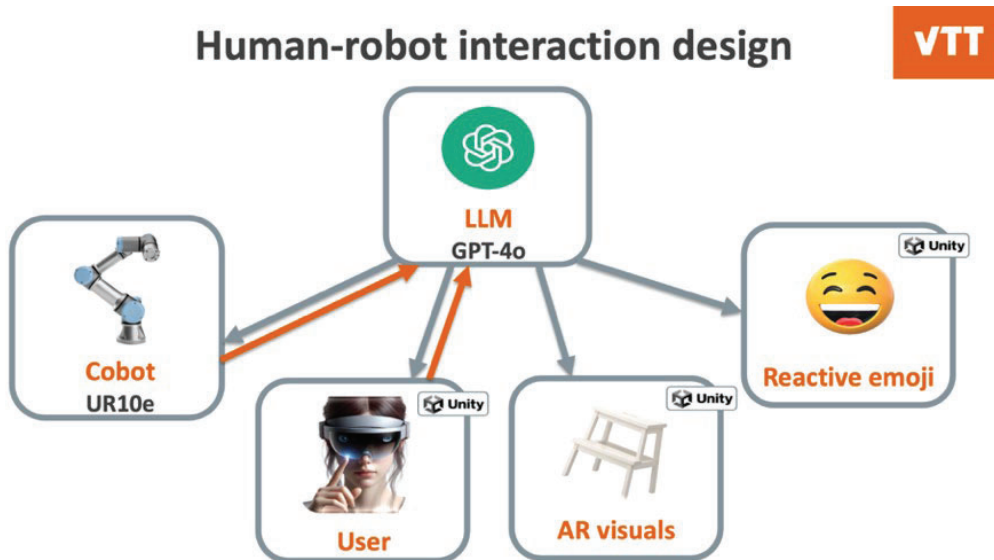


Figure 14. Interaction design graph. The LLM assistant's output affects four different recipients, namely the cobot, the user (through voice), AR hints and the reactive emoji. As input, the assistant is prompted both by the user and the cobot.

Prompt engineering and LLM output

The different responses mentioned above focus on producing a natural output for user interactions and a more structured one for cobot instructions. Research has already demonstrated that structured output for robot instructions (Izzo et al., 2024) can be achieved with appropriate prompt engineering. Our goal was to ensure that the LLM could adapt its communication style depending on its audience, using natural language for users and structured formats for cobots. A parser script was implemented to correct minor errors from the LLM, such as improper capitalization, punctuation, or conversational language in the structured messages.

To guide the assistant in delivering the right behavior, we created an initial system message that outlines the steps the user must perform and how to communicate them. This message is based on the instructions for assembling an IKEA Bekväm step stool, derived from the official manual with seven total steps. The assistant is instructed to stay true to the assembly process and divide longer steps into substeps for easier communication.

UI implementation

The UI elements (Figure 15) can be split into two main groups: 3D assembly instructions and an AI assistant interface. Each UI group can be interacted with directly (by hand) or indirectly (with far pointers). Users can place each UI group anywhere in the environment to keep it out of the way, and the directional arrows are displayed when those elements are out of the user's FOV. 3D assembly instructions can be hidden completely, while AI assistant UI is always visible. 3D assembly instructions are meant to be interacted with by hand while AI assistant UI mostly uses hands-free gaze and speech interactions.

The 3D assembly UI features a 3D product model with the product name and current step displayed. Users control the assembly via a hand menu, offering toggles like "3D Model" (show/hide model), "1:1 Scale" (real-world size), "Exploded View" (show part connections), and "Step Highlight" (focus on current parts). Navigation is through "Previous Step," "Next Step," and "Current Step" buttons, while the AI advances steps based on user progress.

The AI assistant interface uses animated 3D emoji and a chat system, with user questions in green and AI responses in blue. Animated emoji reflect the AI's state, enhancing interaction and emotion. The system has ten states, including Sleeping, Idle, Attentive, Listening, Disappointed, Processing, and Speaking. For example, after inactivity, the AI shifts from Idle to Sleepy, then Sleeping. If the user looks at the AI, it becomes Attentive and may enter the Listening state.

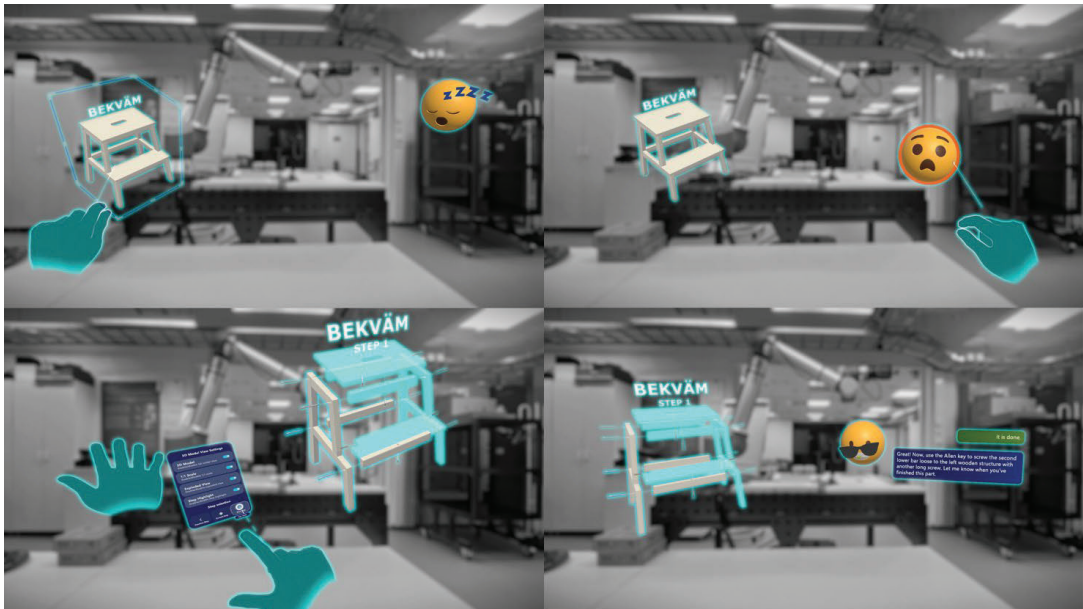


Figure 15. (top left) 3D assembly instructions manipulation; (bottom left) assembly step selection; (top right) AI assistant manipulation; (bottom right) AI assistant's response to the user.

The Listening state uses speech detection with a circular timer and volume bar. If no speech is detected, the AI shows Disappointment and returns to Idle. If successful, it acknowledges, processes, and responds, then reverts to Idle. For the Speaking state, the AI will generate a response to the user's spoken request. Additionally, based on the generated response, AI will select an appropriate animated emoji and generate positive, neutral, and negative automated responses for the user. The Moving state handles hand interactions, and Error appears when server issues occur.

Robot implementation

Universal Robots 6-axis UR10e cobot was utilized for the presented application (presented in Figure 16). The cobot performs a traditional pick and place task, where the individual components to be handled are located on a jig-plate. Component locations are known to the program and no machine vision is required at this stage. Handling of the components is apprehended with a Robotiq vacuum gripper.

Communication between the cobot and the LLM wrapper is established using TCP/IP sockets. In this setup, the server machine hosts the LLM, while the cobot functions as the client. The communication follows a command-based protocol aligned with the step-by-step instructions for assembling the stool previously described. As the operator progresses with the assembly, the LLM sends commands to the cobot to provide the necessary components for the next step. The messages exchanged between the server and client are strings formatted in big-endian. The cobot program interprets these strings and executes the corresponding step. After completion, the cobot program notifies the LLM. The LLM then acknowledges this by informing the operator via generated voice output.

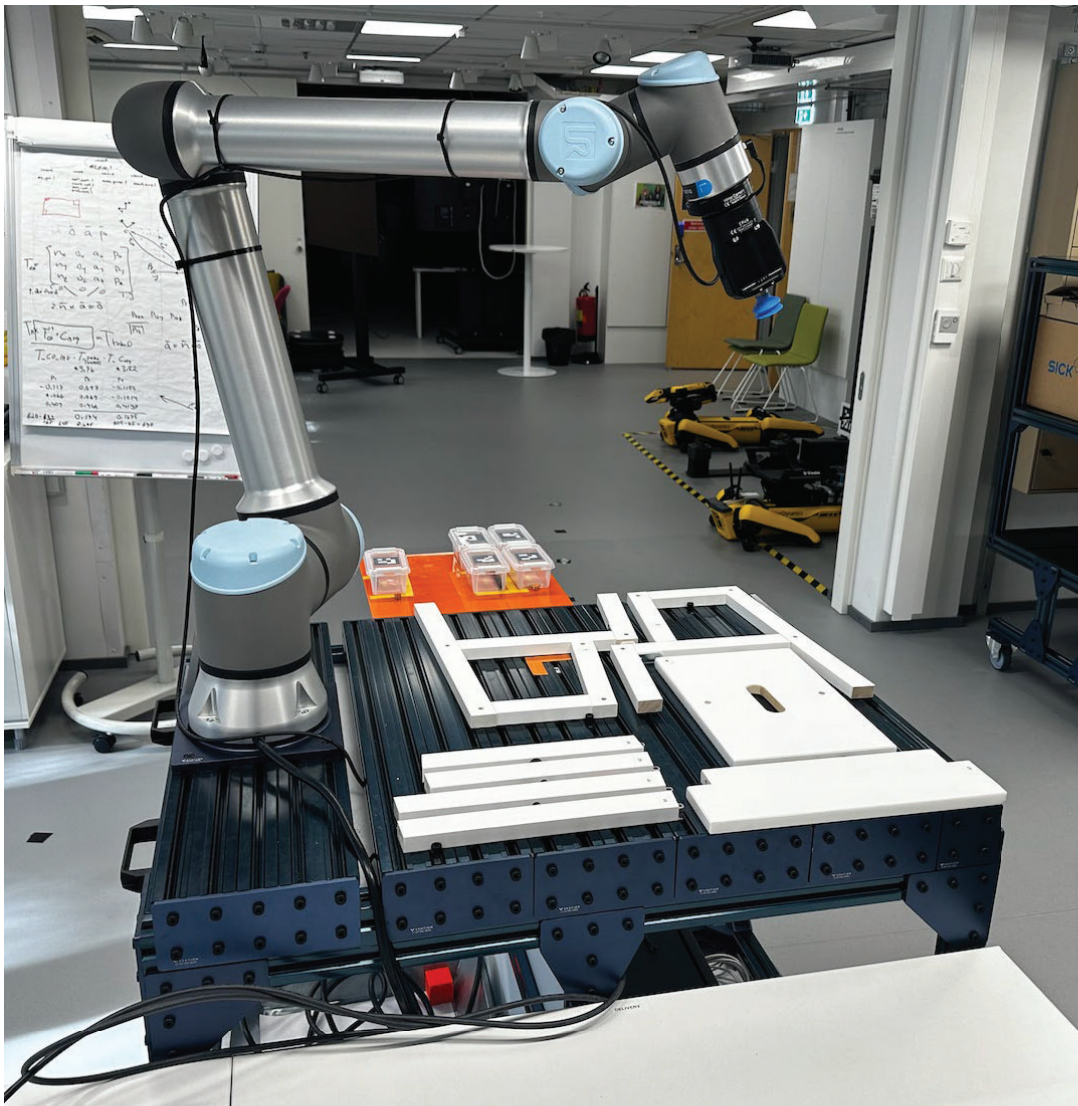


Figure 16. UR10e cobot cell for IKEA step stool assembly.

Conclusion and next steps

In conclusion, our project has successfully demonstrated the transformative potential of LLMs when applied to practical, real-world scenarios. Starting with limited resources, we developed an assistant capable of engaging in dynamic conversations and interacting seamlessly with the UR10e cobot to provide timely assistance. This proof of concept showcases not just the viability of LLMs in industrial settings but also their ability to be context-aware, an area where we see significant opportunity for innovation.

As we move forward, the next steps involve refining the architecture and exploring its applicability across various industries. We should focus on gathering domain-specific data to improve the assistant's decision-making and contextual understanding. Multi-modal systems, which process and integrate text, images, and audio, will enable more complex functions and richer feedback. These interactions make the assistant more flexible and accessible, enhancing the user experience. Partnering with real-world environments for testing will be crucial. By continuing to develop this technology, we can unlock new possibilities for LLMs, driving innovation and practical benefits across sectors.

References

- Sonnino, N. (2024). Large Language Models as smart space aware social conversational agents. Politecnico di Milano. <https://hdl.handle.net/10589/219872>
- Izzo, R., Bardaro, G., & Matteucci, M. (2024) BTGenBot: Behavior Tree Generation for Robotic Tasks with Lightweight LLMs. <https://doi.org/10.48550/arXiv.2403.12761>
- Guan, Y., Wang D., Chu, Z., Wang, S., Ni, F., Song, R., Li, R., Gu, J., & Zhuang, C. (2023). Intelligent virtual assistants with llm-based process automation, <https://arxiv.org/abs/2312.06677>, arXiv preprint arXiv:2312.06677

APPEDIX

Video of the PoC implementation: https://youtu.be/RzetidigL_0?si=peEkdYuxpz-lxrk6

Quantum Quadrant: Could We Use XR and AI Together in Education? Or Rather, Should We?

Yücel Uzun, Andoni Retegui Schiettekatte, Suhana Biswas, and Florian Buchholz

Fraunhofer FIT, Sankt Augustin, Germany

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

Keywords: Extended Reality, Generative AI, Education, Collaboration, Soft Skills

Introduction

Soft skills, such as communication, problem-solving, and teamwork, are essential components of children's education, contributing significantly to their overall development and learning efficiency. While Extended Reality (XR) and Artificial Intelligence (AI) have been widely utilized across various educational contexts (Gua et. al. 2021, Yim et. al. 2024), their combined use in formal education, particularly for building soft skills, remains underexplored (Ouchen et. al. 2022).

Recognizing this gap, we developed a collaborative XR application, named Quantum Quadrant, with text-generative-AI-driven feedback during the XRHack 2024¹ event in Cologne, Germany. The initial concept was to explore how XR and generative AI could be used together to improve both physics understanding and soft skills development in elementary schools. The hackathon environment, with its time constraints and competitive nature, compelled us to focus on core features and rapidly create a minimum viable product. Currently, the application is being refined as a proof of concept, with a focus on addressing the design and technical challenges associated with co-located, collaborative XR experiences, and the usage of generative AI. This paper presents an overview of the application, discusses its unique design and technical challenges, and evaluates the potential and limitations of using XR and AI to support soft skills development in educational environment setting.

Overview of the Application

The main goal of the application is to enhance both physics understanding and soft skills among elementary school students through an interactive, collaborative XR experience. The application is designed for four co-located players using the Meta Quest platform and combines a puzzle game with generative AI-driven feedback to improve communication, teamwork, and problem-solving abilities.

¹ <https://www.xrhack.com>

At the start of each session, four players are assigned individual objects that only they can see and interact with: a laser, a mirror, a cannon, and a magnet. To solve the puzzle, players must coordinate their actions to correctly place their objects within the room. The task is designed to include different physical concepts and basic information about these are also directly presented to user as a text and images. The solution requires directing the laser beam from its source, reflecting it off the mirror, activating the cannon, and finally attracting the fired cannonball with the magnet. Additionally, the laser must reflect off the mirror before it can activate the cannon. Players must clearly communicate in order to complete the task successfully. When the laser is activated, all players can see the entire setup, but objects can no longer be moved. If the solution is incorrect, the game resets, prompting the team to try again.

When the laser is activated, the complete transcript of the players' conversation is sent to OpenAI², where it is analysed to provide feedback on communication skills. The AI generates personalized suggestions for each player, highlighting areas for improvement in their collaborative and communication behaviours. This AI-driven feedback is intended to promote reflection and development of soft skills in an engaging, real-time context.

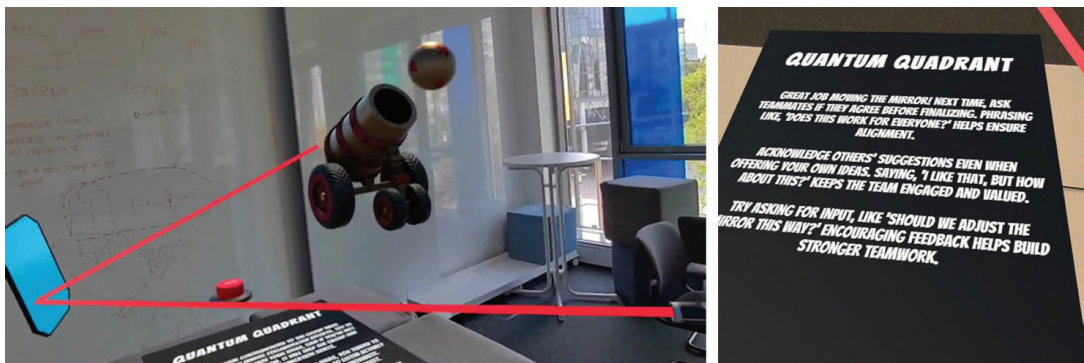


Figure 17. Left: Part of the puzzle, where laser triggers the cannon via a mirror. Right: Visualization of AI feedback.

Design and Technical Challenges

The design of the application centered on integrating physics concepts with soft skills development without overwhelming the players. The structure of the game encourages iterative problem-solving, as players must repeatedly attempt to solve the puzzle through trial and error, refining their strategies and communication with each attempt. This iterative approach not only reinforces physics concepts but also improves the communication and teamwork skills by requiring continuous dialogue and coordination.

From a technical perspective, one of the major challenges was achieving a co-located multiplayer experience on the Quest 3, which lacks support for image marker or object detection as a built-in feature and does not allow direct camera access, preventing for these capabilities to be implemented independently. However, the colocation features of Meta's Presence Platform helped us to position users accurately within the same physical environment relative to each other. Our approach was based on establishing a shared anchor point between the Quest glasses and synchronizing the positions of other objects relative to this reference point. However, it is important to note that the cloud-based

² <https://platform.openai.com/>

synchronization is not always reliable, particularly when users attempt to use the application in different environments.

Another challenge was adapting the application to dynamic environments, ensuring that the puzzle remained solvable across different settings. The current implementation assumes a central table that divides the room into four quadrants in which each player can operate. However, this setup may not work in all environments, so the complexity of the puzzle was reduced to ensure reliable operation.

Performance limitations of the Quest 3 glasses also influenced the design choices. The initial concept presented at the XRHack event included a more complex puzzle with other interactive elements like a pulley and water system. However, the Quest 3 struggled with the realistic physics simulations required for these components, prompting us to simplify the design. While custom implementations and further optimizations tailored to the Quest hardware could potentially address these issues, current off-the-shelf libraries do not offer sufficient performance for these more complex simulations.

Integrating generative AI posed also predictable challenges, particularly in designing appropriate “prompt engineering” techniques to manage the AI’s responses. We have observed that certain prompt wordings, such as “criticize”, often lead to overly harsh feedback despite the rest of the prompt, which could negatively impact young users. Crafting prompts that encourage constructive and age-appropriate feedback requires careful tuning to ensure that the suggestions remained supportive and aligned with the educational goals of the application.

Discussion

Our initial and unstructured observations with the application suggest that the design choices, particularly those involving data scarcity and iterative problem-solving, positively influence communication and group dynamics. By allowing each player to see and interact with only their assigned object, participants are compelled to engage more actively in dialogue and use physical gestures to coordinate their actions. This setup causes a more balanced group interaction, where each player’s contribution is necessary for successful puzzle completion. In contrast, tests with a demo mode where all objects were visible to everyone often resulted in one participant taking the lead, reducing the collaborative nature of the game, and diminishing the soft skills training aspect.

Adapting the application to dynamic environments remains a significant challenge. While schools in controlled environments could provide the necessary space and layout consistency, future iterations should explore more adaptive spatial mapping techniques for puzzles that better accommodate different room layouts, enhancing the robustness and accessibility of the application. Alternatively, the application could include a creative mode, similar to existing authoring tools (Horst et. al. 2022) that allows teachers to design and arrange their own puzzles, adapting the gameplay to fit the physical environment available in their classrooms.

The integration of generative AI in an educational XR setting shows substantial potential, but also raises important ethical and privacy concerns. The use of AI for feedback involves transmitting conversation transcripts to a third-party server, which raises issues of data privacy and consent, particularly when working with children. In addition, XR technologies that track body and hand movements could inadvertently introduce new forms of student monitoring, potentially leading to privacy violations. To address these concerns, we recommend exploring locally processed data solutions, such as on-premises generative AI models and no-log local servers for XR applications, to better protect user data.

During testing, OpenAI generally provided relevant and contextually appropriate feedback; however, the AI also displayed tendencies toward hallucinations, where the AI model generates incorrect or misleading information, and occasionally failed to adhere strictly to the provided prompts. This unpredictability underscores the need for an approval mechanism, allowing teachers or other professionals to review AI-generated feedback before it is shown to students. Such oversight ensures that the feedback remains constructive, accurate, and appropriate for the educational context.

Furthermore, while our team includes a qualified teacher, the AI-generated feedback should also be independently evaluated by other experts to ensure it is applicable and beneficial to school-aged children, particularly in providing constructive feedback within a reflection-on-action context. Establishing clear guidelines for the use of generative AI in schools would be crucial to ensure that AI-driven feedback aligns with educational standards. Incorporating this feedback into design of the application could help guide students to follow suggested improvements and reinforce learning goals in a more structured way.

Finally, this application serves as a proof of concept that demonstrates the feasibility of integrating XR and AI to support both physics learning and soft skills development in an educational context. However, formal scientific studies using a mixed method approach are needed to validate the efficacy of this approach, assess its long-term impact, identify any other unforeseen challenges and address privacy and ethical concerns. Future research should aim to quantify the benefits, refine the AI feedback mechanisms, and explore broader applications in diverse educational settings.

References

- Guo, X., Guo, Y., & Liu, Y. (2021). The Development of Extended Reality in Education: Inspiration from the Research Literature. *Sustainability*, 13(24):13776. <https://doi.org/10.3390/su132413776>
- Horst, R., Naraghi-Taghi-Off, R., Rau, L., & Doerner, R. (2022). Authoring with virtual Reality nuggets—lessons learned. *Frontiers in Virtual Reality*, 3. <https://doi.org/10.3389/frvir.2022.840729>
- Ouchen, L., Tifroute, L., & El Hariri, K. (2022). Soft skills through the prism of primary school teachers. *European J. Educ. Res.*, 11(4), 2303-2313. <https://doi.org/10.12973/eu-jer.11.4.2303>
- Yim, I.H.Y., & Su, J. (2024). Artificial intelligence (AI) learning tools in K-12 education: A scoping review. *J. Comput. Educ.* <https://doi.org/10.1007/s40692-023-00304-9>

NoFold: Bridging the Physical and Digital in Board Game Creation

Christos Lougiakis, Lori Kougioumtzian, Giorgos Ganias, Anastasios Niarchos

National and Kapodistrian University of Athens

Corresponding author: chrislou@di.uoa.gr

Keywords: Human-Computer Interaction, Extended Reality, 3D Interaction, Board Games, Playtesting

Introduction and Motivation

Board game creators face numerous challenges during the design process, particularly in the time-consuming and iterative playtesting phase. Traditional playtesting requires physical presence, physical prototypes and components, as well as considerable time and a dedicated space, often leading to excessive costs and logistical hurdles (Kougioumtzian et al., 2023). Digital tools offer a potential solution by facilitating remote playtesting, streamlining component management, and enabling easier modification. However, according to relevant articles, forums and blog posts, members of the board game community have found that traditional digital platforms often fall short in replicating the nuances of physical interaction, social cues, and the overall tabletop experience. Additionally, they might have a learning curve or limitations when it comes to the design and creation process (Loewen, 2020, Backe, 2020, Reddit, 2022, Kinne, 2024). These limitations can impact the accuracy of playtesting conclusions and influence design decisions.

Recent research has highlighted the potential of eXtended Reality (XR) to bridge this gap. XR environments foster social presence, enable natural gestures and on some occasions facial expressions and support physics-based interactions that closely mirror the real world. These advancements of the medium, combined with effective, user-centric experience design approaches, are capable of creating new opportunities in multiple different fields (Stanney et al., 2021). This potential led us to develop NoFold, a next-generation web/XR hybrid platform designed to cover the needs of the board game community. Utilizing the immersive capabilities of XR and the flexibility of web technologies, NoFold aims to empower board game creators to design, playtest, and share their games seamlessly.

NoFold Platform

NoFold is a web/XR hybrid platform designed to address the limitations of existing tools, offering the accessibility of digital playtesting, while also maintaining the richness of physical interaction. By leveraging immersive XR technologies, we bridge the gap between the physical and digital worlds, providing a comprehensive solution for board game creators.

NoFold's web application serves as the creative hub, where users can create and manage game components, design cards, and define their properties. The XR application, compatible with Virtual Reality (VR) and Mixed Reality (MR) headsets, provides an environment for realistic playtesting. By combining these two complementary platforms, NoFold offers a comprehensive end-to-end solution for the entire board game development lifecycle. The game used in the following images for demonstration purposes is an existing card game called *Love Letter* and apart from a physical copy we do not claim to own the rights of the game in any way.

Web Application

The NoFold web application offers an intuitive interface for game creation and component management. Users can create a personal account and access their games through the dashboard. Currently focusing on card games, the platform allows for bulk uploading of card images, customization of card properties (e.g., dimensions, quantity), and auto-save functionality for seamless workflow. Future developments prioritize features to improve the card game workflow, such as component grouping, bulk editing, and free placement on the virtual table. Games created in the web app are instantly accessible in the XR app through a virtual library of games, enabling rapid iteration and testing.

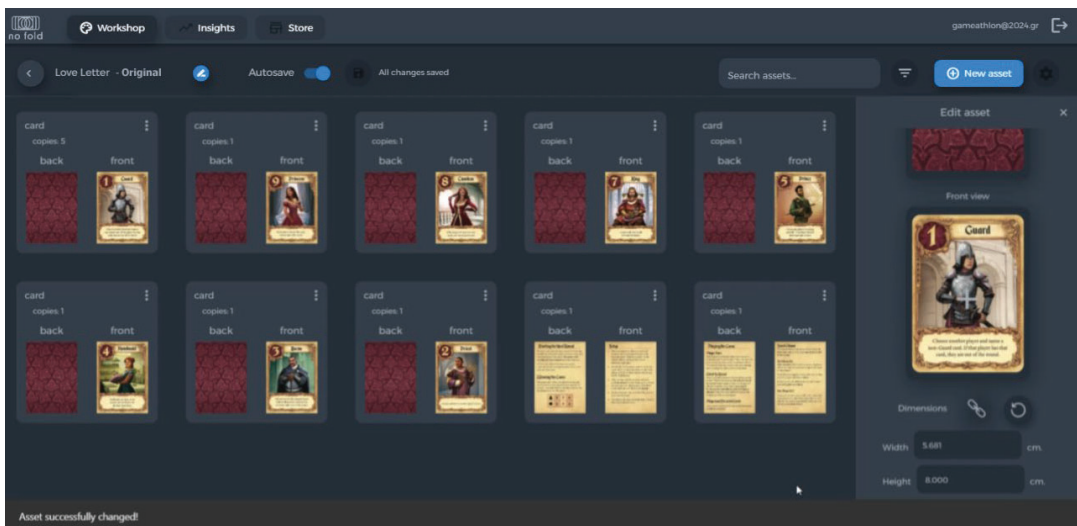


Figure 18. Editing a card game in the game creation interface of the web application.

XR Application

Leveraging the power of VR and MR to create an immersive playtesting environment, users can be co-located or connect with others remotely, interact naturally with virtual game components, and experience the social dynamics of in-person playtesting. They have the option to select avatars and connect to virtual rooms, fostering social presence and enabling voice chat communication.

The XR application brings the created games to life in two ways: (1) in VR where users immerse themselves in a virtual environment, (2) in MR where the physical world gets blended with the virtual. The integration of scene understanding in MR allows NoFold users to seamlessly interact with virtual components in a real-world setting, ultimately transforming any tabletop into a dynamic game board. Additionally, physics-based manipulation of game elements facilitates a more natural and immersive testing process. A virtual library provides access to games created using the web application, which can

be loaded onto a virtual table for playtesting. Providing hand tracking and controller input options offers flexibility for players, while physics interactions allow for intuitive manipulation of virtual objects (Kim and Park, 2015, Höll, et al., 2018, Lougiakis et al., 2024), resembling the naturalness of real-life interactions. Currently, the platform focuses on card games, supporting physics-based interactions like grabbing, moving, and throwing cards. Future updates will introduce more card interactions, group manipulation, a shuffling mechanism, avatar customization, and support for 3D models, game pieces, and dice.



Figure 19. Left: Playing a card game on a virtual table in VR . Right: Playing a card game on a real table in MR³.

Addressing the Needs of Board Game Creators

One of our recent studies exploring the needs of board game creators through targeted questionnaires highlighted the strengths and weaknesses of current digital tools, as well as the specific needs board game creators have for the game design process (Kougioumtzian et al., 2023). Additionally, we are currently exploring their needs further by conducting interviews with individual creators, to get more refined and detailed feedback (study has not been published yet). The results from the questionnaires and interviews show the potential of XR in facilitating social presence, natural interactions, and nuanced player dynamics, addressing the limitations of the existing digital tools.

NoFold's development is driven by user research highlighting the advantages and limitations of current digital tools, while also drawing from the specific requirements for board game creation that were brought to light through it. Our design process is focused on integrating social interaction, natural hand interactions, and physics-based interactions into a unified approach that enables nuanced player interactions and conclusions that can be directly applied to physical game design. Our philosophy aligns with these requirements, prioritizing features that enhance the collaborative and interactive aspects of playtesting.

³ Video overview of the MR version can be found at <https://youtu.be/17LOY8y9zA?si=KC52ITDqsbhUZRGV>

Conclusion and Future Directions

NoFold represents a promising approach towards advancing board game creation tools, addressing gaps in current digital solutions, though future iterations and studies are needed to fully realize its potential and validate the platform's benefits over existing tools. By combining a web-based authoring tool with an immersive XR playtesting environment, we aim to provide a comprehensive platform that supports the entire game creation process, empowering creators to bring their ideas to life and gather valuable feedback. The platform's current focus on card games allows for in-depth feature development and optimization of the core workflow. Future plans include expanding support to other game components such as 3D models, dice, and game pieces, further enhancing the platform's versatility. The platform will also continue to evolve based on feedback received from comprehensive user studies and evaluations assessing all different aspects and the overall user experience, maintaining its value as a tool for the entire board game community. Our goal is to provide a platform that not only streamlines the design and playtesting process, but also fosters a thriving community of creators and players, pushing the boundaries of board game innovation.

References

- Backe, C. (2020). Making board games virtually: A review of 5 tools. Entro Games. <https://www.entrogames.com/2020/04/making-board-games-virtually-a-review-of-5-tools/>
- Höll, M., Oberweger, M., Arth, C., & Lepetit, V. (2018, March). Efficient physics-based implementation for realistic hand-object interaction in virtual reality. In 2018 IEEE conference on virtual reality and 3D user interfaces (VR) (pp. 175-182). IEEE.
- Kim, J. S., & Park, J. M. (2015, May). Physics-based hand interaction with virtual objects. In 2015 IEEE International Conference on Robotics and Automation (ICRA) (pp. 3814-3819). IEEE.
- Kinne, O. (2024, March 5). Digital advantage: Online gameplay experience versus in-person – Topic discussion. Tabletop Games Blog. <https://tabletopgamesblog.com/2024/03/05/digital-advantage-online-gameplay-experience-versus-in-person-topic-discussion/>
- Kougioumtzian, L., Lougiakis, C., & Katifori, A. (2023, April). "Show your cards!": What do creators need for the game design process?. In Proceedings of the 18th International Conference on the Foundations of Digital Games (pp. 1-6).
- Lege, R. (2024). A social presence benchmark framework for extended reality (XR) technologies. Computers & Education: X Reality, 4, 100062. <https://doi.org/10.1016/j.cexr.2024.100062>
- Loewen, M. (2020, December 15). Is Tabletop Simulator the future of playtesting? Cloudfall Studios. <https://www.cloudfallstudios.com/blog/2020/12/15/is-tabletop-simulator-the-future-of-playtesting>
- Lougiakis, C., González, J. J., Ganias, G., Katifori, A., & Roussou, M. (2024, March). Comparing Physics-based Hand Interaction in Virtual Reality: Custom Soft Body Simulation vs. Off-the-Shelf Integrated Solution. In 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR) (pp. 743-753). IEEE.
- Reddit. (2022, September 14). Is Tabletop Simulator a good alternative to physical playtesting? Reddit. https://www.reddit.com/r/boardgames/comments/xcd5gv/is_tabletop_simulator_a_good_alternative_to/
- Stanney, K. M., Nye, H., Haddad, S., Hale, K. S., Padron, C. K., & Cohn, J. V. (2021). Extended reality (XR) environments. Handbook of human factors and ergonomics, 782-815. <https://doi.org/10.1002/9781119636113.ch30>

From Stress to Success: Evaluation of a Comfortable VR Environment for Product Engineering

Annegret Melzer¹, Philipp Stiens¹, Samuel Korb¹, Franziska Klimant¹,
and Martin Dix^{1;2}

¹*Chemnitz University of Technology, Institute for Machine Tools and Production Processes,
Chemnitz, Germany*

²*Fraunhofer Institute for Machine Tools and Forming Technology IWU, Chemnitz, Germany*

Corresponding author: annegret.melzer@mb.tu-chemnitz.de

Keywords: Human Centered Design, User Experience, Virtual Reality, Evaluation, Spatial Design

Introduction

Saving resources, the ability to work in distributed teams and the shortening of development processes are drivers for the use of virtual reality (VR) in the product life cycle. Here collaborative tools for remotely designing products are often praised as one of the most important applications of VR in industry. However, such complex applications that are not intuitive to use and cause more frustration than pleasure can be associated with negative effects such as cognitive load (CL) and stress (Souchet et al., 2023). Thereby, the functional requirements cannot be reduced in order to reduce complexity or improve the quality of use, as they are necessary for task fulfillment. Therefore, the collaborative VR application "VoirVR" was designed and implemented between 2021 and 2023. The application is intended to make users feel comfortable when using VR through targeted user experience (UX) design and the spatial design of the virtual environment. This study investigates the positiveness, stress-freeness, naturalness and intuitiveness of VoirVR, demonstrating that comfortable collaborative VR applications are achievable.

Related Work

UX is understood as the entirety of the (subjective) perception, reactions and emotions of users when using a product or application (DIN EN SO 9241-11:2018-11). For a high UX users have to feel comfortable and satisfied. In order to ensure that UX is as positive as possible, future users should be involved in the development process by applying the user-centered design (UCD) process (DIN EN ISO 9241-210:2020-3). This paper focuses on the fourth and final phase of the process, which involves iterative testing, analyzing and assessment of the evaluation object (LaViola et al., 2017). To measure the overall perception of an application, many UX questionnaires with different notions are available,

each targeting various aspects of user experience (Schrepp, 2020). Among the factors these questionnaires assess, both cognitive under- and overload are significant contributors to poor UX. Chandler and Sweller (1991) distinguish three forms of cognitive load in their cognitive load theory (CLT): Intrinsic Cognitive Load (ICL), Extraneous Load and Germane Cognitive Load (GCL). Besides CL, emotions play an important role in the UX. Relying on environmental psychology, the perception of an environment can have a direct influence on the emotion (Russell & Lanius, 1984).

The VoirVR Application

The demonstrator used for the evaluation is the result of the third iteration of a user-centered design process. The application supports the collaborative configuration of a product using the example of a robot. The UI design is based on the neumorphism style, a fusion of skeuomorphism and flat design. This design focuses on an organic, clear and natural aesthetic that is intuitive and emotionally appealing. The main element of the application is an interactive table, as shown in Figure 1 (left). A miniature view of the product is placed on this table to create a central basis for discussion and to examine and configure the product. A special feature is the ability to teleport to the miniature table for a 1:1 view. There are interactive stations around the table for product configuration. Each user has a personal inventory with tools such as a camera, microphone, pen and notepad. The Avatars have an upper body, head and abstract arm representation with controllers as hands. By tilting and holding the joystick on the controller, the user can navigate to specific locations within the VR environment.



Figure 20. Rendering of the VoirVR environment and the interaction concept

Measurements

The "VoirVR" application is designed for collaborative product configuration. The aim of this study is to evaluate the overall design in terms of design goals for a pleasant environment to avoid stress and to provide easy-to-use interaction concepts to enable intuitive interaction and balanced cognitive load. In order to measure the three dimensions of CL the Cognitive Load Scale (Klepsch et al., 2017) was selected for this study. The scale includes two items for ICL, three for ECL, and three for GCL, each rated on a 7-point Likert scale (1 = 'absolutely wrong' to 7 = 'absolutely right'), where higher numbers indicate a higher cognitive load. The Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) was chosen to measure the effect of the "VoirVR" environment on user's emotions in the dimensions Valence, Arousal and Dominance by a picture-oriented 5-point scale with figures that represent from unhappy to happy for valence, sleepy to excited for arousal and from small to large for dominance with large representing full control over a situation. The UEQ-short (Schrepp et al., 2017) and the INTUI (Ullrich & Diefenbach, 2010) were selected as UX questionnaires. The UEQ-S measures the general, hedonic and pragmatic quality of a product on a 7- Point Likert scale (-3 to +3). An Excel tool for data analysis is provided by

Schrepp and his colleagues⁴. The INTUI (Ulrich & Diefenbach, 2010) was used to assess the four sub-components of intuitive interaction: Effortlessness, Verbalizability, Gut Feeling, and Magical Experience and additional the global rating of Intuitiveness on a 7-point Likert scale. The higher the rating the higher is the level of the sub-components examined.

Procedure

The study took place at Chemnitz University of Technology (approved by local ethics board). Due to the higher experimental control, the evaluation was performed with an assumed asynchronous collaboration instead of multiple users simultaneously. Before the actual study, the participants received written instructions detailing the study's progression, functionalities, interaction concepts, and tasks. The evaluation was carried out using a Meta Quest 3 head-mounted display (HMD) in AirLink mode. In the beginning, a VR tutorial video was presented to the participants, demonstrating the task and controller assignments followed by a free explorations phase. The study supervisor provided clarifications if necessary. When the participants felt confident in using the application, the study started.

They were asked to press a virtual buzzer, which unlocked a prepared task. The task for the study was to recreate a given configuration on the interactive table. A total of six configuration variants were developed, which were counterbalanced and assigned to the test subjects. The participants were shown the virtual board, which depicted the individual components in the desired configuration. In addition, the participants had to perform small prototypical design review tasks such as writing a text message, taking a picture or marking a specific point on the robot. For this purpose, the subjects received messages in the VR scenario, which were also assigned to the tasks and participants in a counterbalanced manner. The messages served as stimuli to use the respective tools in the fulfillment of the task. In order to avoid a break in perception, the surveys on emotional perception of space were carried out directly afterwards in the VR. The scenario ended with the participants removing the HMD and answering the Cognitive Load Scale, INTUI, and UEQ-S questionnaires on a tablet.

Results and Discussion

The final sample size was $N = 36$. The participants were between 18 and 48 years old ($M = 26.47$, $SD = 7.14$) and 21 participants were female (58%). All but four participants were right-handed. All had normal or corrected-to-normal vision and no other impairments.

Data in Table 1 show a moderate ICL which indicates that the given task was designed in a way participants could understand and process. ECL was perceived as quite low. It can be assumed that the participants did not feel compelled to use a large part of their mental resources to obtain all the information they needed to complete the task or to interact with the VR environment. In contrast the "good" GCL achieved high values, meaning that the participants used a large part of their mental resources for the actual task solving. For a deeper understanding how the design of the VoirVR application affects the user's emotional perception the data shows that participants were quite happy (Valence), moderate excited (Arousal) and did not feel intimidated (Dominance) by the VR environment.

Table 1. Descriptive Statistics (M= Mean; SD= Standard Derivations) for the variables CL, Emotional Perception and UX (N=36)

⁴ <https://www.ueq-online.org>

Scale	M	SD
Cognitive Load (Cognitive Load Scale)		
• ICL	2.81	1.12
• ECL	2.31	1.11
• GCL	4.92	1.24
Emotional Perception (SAM)		
• Valence	4.31	0.82
• Arousal	4.19	0.82
• Dominance	1.98	0.97
User Experience (UEQ-S)		
• Overall	1.47	0.87
• Pragmatic Quality	1.25	1.22
• Hedonic Quality	1.69	0.87
User Experience (INTUI)		
• Magical Experience	5.41	0.90
• Gut Feeling	3.23	1.35
• Verbalizability	5.19	1.47
• Effortlessness	4.91	0.90
• Intuitiveness	4.75	1.44

The results show an overall positive UX. Especially for the hedonic quality for the VoirVR environment, the results show “excellent” values as part of the benchmarking included in the data analyzing tool. However, the results for pragmatic quality can be classified as “above average”. As the complexity of the VoirVR application is largely predetermined by task requirements and it was not the main focus, the authors are not surprised by this outcome. However, it is noticeable that the dispersion of the data for the items “complicated/simple” (M= 0.97, SD= 1.5) and “clear/confusing” (M=1.22, SD=1.4) suggest that the complexity might not be the main reason and there is room for improvement.

For the INTUI the means for Magical Experience, Verbalizability and Effortlessness are well above the scale midpoint (3.5/7). The results show that the VoirVR environment achieves a high level of Intuitiveness, as was the aim of the design. In terms of Gut Feeling the mean lies slightly below scale midpoint, not necessarily indicating a poor result. As Ulrich & Diefenbach (2010, p. 258) state that Gut Feeling can be perceived less for products that are not fun products and that even if intuitiveness is perceived as high overall, the sub-components can be specified differently depending on the product typ.

Conclusion

We present the evaluation of a comfortable and intuitive VR environment for collaboratively configuring products called VoirVR. The results indicate that the iteratively designed application with its intuitive interface, effectively meets the requirements by maintaining a low negative cognitive load and creating a positive emotional experience. Notably, the overall user experience (UX), especially the hedonic quality, was rated highly. Therefore, practitioners should focus more on spatial design of an VR environment and its impact on emotional perception and UX. Albeit, specific recommendations for design patterns have yet to be developed. Further, we showed that comfort of VR applications can be reflected on different dimensions and levels. Study protocols should, therefore, determine the relevant sub-constructs for their evaluation purpose, and chose suitable measures.

Acknowledgements

This work was funded by the Central Innovation Programme for small and medium-sized enterprises (SMEs) of the Federal Ministry for Economic Affairs and Climate Action. Also, many thanks to our project partners – A.MUSE Vision and VRENDEX - for the successful cooperation.

References

- Bradley, M.M., & Lang, P.J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J. behave. therapy experimental psych*, 25(1), 49-59 . [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Burmester, M., Laib, M., & Zeiner, K.M. (2017). Positive Erlebnisse und Wohlbefinden in Arbeitskontexten durch Gestaltung der Mensch-Computer-Interaktion. In: *Positiv-Psychologische Forschung im deutschsprachigen Raum – State of the Art*. Brohm-Badry et al. (Eds.). Lengericht: Pabst, 158-175.
- Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2
- Hassenzahl, M. (2008). User experience (UX): Towards an experiential perspective on product quality. *ACM International Conference Proceeding Series*, 11–15. <https://doi.org/10.1145/1512714.1512717>
- DIN EN ISO 9241-11:2018-11 Ergonomics of human-system interaction- Part 11: Usability: Definitions and concepts (ISO 9241-11:2018).
- DIN EN ISO 9241-210:2011-01, Ergonomics of human-system interaction- Part 210: Human-centred design for interactive systems (ISO 9241-210:2011).
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in psychology*, 8 (1997). <https://doi.org/10.3389/fpsyg.2017.01997>
- LaViola, J. J., Kruijff, E., McMahan, R. Bowman, D., & Poupyrev, I. (2017). 3D User Interfaces: Theory and Practice. *3D User Interfaces. Theory and Practice* 190-271.
- Russell, J. A., & Lanius, U. F. (1984). Adaptation level and the affective appraisal of environments. *Journal of Environmental Psychology*, 4(2), 119–135. [https://doi.org/10.1016/S0272-4944\(84\)80029-8](https://doi.org/10.1016/S0272-4944(84)80029-8)
- Schrepp, M. (2020). A Comparison of UX Questionnaires - What is their underlying concept of user experience? *Mensch und Computer 2020 - Workshopband*. Bonn: Gesellschaft für Informatik e.V. Magdeburg. 6.-9. September 2020. <https://doi.org/10.18420/muc2020-ws105-236>
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 103-108. <http://doi.org/10.9781/ijimai.2017.09.001>
- Souchet, A.D., Lourdeaux, D., Pagani, A., & Rebenitsch, L. (2023). A narrative review of immersive virtual reality's ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload. *Virtual Reality* 27, 19-50. <https://doi.org/10.1007/s10055-022-00672-0>
- Ullrich, D., & Diefenbach, S. (2010). INTUI. Exploring the Facets of Intuitive Interaction. In: *Mensch & Computer* 251-260. <https://doi.org/10.1524/9783486853483.251>

Real time facial capture and animation to explore the controls neurodiverse people need for their avatars

Simon Champion^{1,2}, Radoslaw Dorociak^{1,2}, and Georg Meyer^{2,3}

*Virtual Engineering Centre, Daresbury, UK
IDEAS, Faculty of Engineering, Liverpool University, Liverpool, UK
Hanse Wissenschaftskolleg, Delmenhorst, Germany*

Corresponding author: georg@liv.ac.uk

Keywords: Avatars, Neurodiversity, User Engagement, User Interaction, User Experience

Introduction

The exploration of realistic avatars is rapidly advancing in both academic and industry contexts. Alongside these technical developments, critical ethical considerations arise, particularly concerning data management, privacy, diversity, inclusion, personal agency, identity, and societal impact. These issues demand thorough and ongoing reflection throughout the research and development process.

User involvement is key for these reflections, but, as for any new technology, high-quality demonstrations of the technological possibilities and constraints are a base requirement for informed discussion. The aim of this project, consequently, is to develop application scenarios that enable neurodiverse user groups to experience and explore how they would want to be represented by avatars in computer mediated environments and what controls would be required by users.

Here we present an avatar application toolkit designed for neurodiverse individuals, focusing on how avatars can enable the manipulation of facial animation, attention, and focus. We will explore three specific applications targeting different conditions. In each case the aim of the manipulation is to provide neurodiverse individuals control the degree to which features that result from neurodiversity are experienced by third parties. The first addresses facial paralysis, enabling real-time modifications of facial animation data. The second facilitates the manipulation of focus and gaze direction, allowing individuals uncomfortable with eye contact to shift the focus of their communication partner's avatar and control their own gaze position. The third application is designed to enable individuals with Tourette's Syndrome explore their avatar representations by blending idle animations with live data to help mask involuntary tics.

It is crucial to acknowledge the potential for polarized reactions to these technologies. While some may appreciate the ability to modify their avatar to suit their needs, others might oppose any alterations to their appearance or communication, as these elements reflect their true selves. We expect some users to appreciate to manipulate their avatars depending on the context. The emphasis of the underlying

research therefore is to develop tools that enable potential end users of the technologies to influence their development.

To better understand public perceptions of avatar technology, particularly in both conventional and socially transformative contexts, we are collaborating with the Brain Charity to collect user views. This partnership will provide valuable insights into how these innovations are received and their broader implications for personal and social identity.

All three of the applications were developed in Unreal Engine 5.1 utilizing latest visualization tools, most notably MetaHumans. The framework developed by Epic Games allows swift and accessible creation of virtual characters. Each character can be made from scratch using a diverse range of templates provided by the MetaHuman Creator, or created based on a real person using a photogrammetric scan. Each character is fully rigged, allowing extensive control of the facial animation down to subtle details such as tongue movement or cheek puffing. Animating the characters can be done either manually using a MetaHuman Control Rig, or captured directly from an actor in real time. The true potential of these methods can be achieved when they are used in combination with one another, allowing us to modify the facial animation as it is being captured.

Technical Realization

Facial capture is at the core of all three applications. We use Epic Games' Live Link Face for real-time facial motion capture using a compatible iPhone or iPad. The software utilizes Apple's 3D scanner technology, including the TrueDepth camera and LiDAR sensors, capturing precise facial movements and depth data in real time. Using the Live Link feature, this data is streamed over the network to Unreal Engine, where it can be used to animate the MetaHuman avatar simulation. The TrueDepth camera tracks detailed facial expressions allowing the MetaHuman to mimic the actor's facial animations in real-time.

In the following sections we will look closer into how each of the applications utilizes the technology and how the live animation data is modified to account for different forms of neurodiversity.

Application One: Facial Paralysis

Cranial nerve palsies are disorders of the nerves involved in eye movements, facial expression and the anterior portion of the tongue. These disorders lead to a, typically unilateral, loss of essential facial functions, such as blinking, nasal breathing, lip sealing, smiling or speaking (e.g Roob et al., 1999). Cranial nerve palsies have been linked to anxiety, depression and quality of life reduction in patients, in part because patients report issues around their perceived appearance and difficulties with social interactions, for example smiling (Norris et al., 2019).

To model facial palsies, the animation data, captured from the user's face, are modified using a custom control rig that enables users to control the degree of paralysis across the three main facial nerves, independently on each side of the face.

Users can choose to fully or partially paralyze any part of the character's face and directly observe the impact that this paralysis has on the facial animation, similar to watching their face in a mirror. On the user interface the parts the user has chosen to paralyze are highlighted in green (Figure 21, left).

Another key feature allows users to mirror animation data from one side of the face to the other, making the side affected by a condition appear like the healthy side on the virtual character. The mirrored areas are highlighted in blue (Figure 21, right).

All modifications are visualized in real-time. These seamless adjustments are made possible by blending the Live Link input with an idle 'Paralyzed' animation, using the rig to determine which facial bones should receive animation data and to what extent it should be applied compared to the paralyzed variant.



Figure 21. Controls for manipulating facial palsies – the left panel shows how facial nerves can be 'artificially paralyzed' by restricting captured face movements in selected areas. The right panel shows how movement on one side can be mirrored to animate the af.

Application Two: Attention and Focus

Nonverbal signals, especially eye contact, play an essential role in social interactions. Neurotypical adults and neonates preferably and reflexively make eye contact with interlocutors. A proportion of individuals with autism spectrum disorder (ASD) show atypical responses to eye gaze and avoid eye-contact (Senju & Johnson, 2009).

Eye-gaze in avatars, of course, can be arbitrarily controlled: users that feel uncomfortable with direct eye-contact can redirect the gaze of their interlocutors at will, just as easily as users who do not like to look directly at others can set their avatars to 'do this for them'.

Application 2 focuses on controlling the direction in which the virtual avatar is facing. This is managed through an interface located on the left side of the screen (Figure 22), where the user moves a white dot within a grey box to adjust the avatar's head gaze. The dot's position corresponds to the avatar's gaze, with the center of the box representing the avatar looking directly at the user.

The control rig for the avatar's facing direction can be toggled in real-time, allowing users to switch between controlling the avatar's direction through the rig or based on their own head movements, captured via Live Link. This effect is achieved by adjusting the character's neck position and blending the motion capture data with the direction specified through the control interface.



Figure 22. User interface enabling users to control head and eye gaze of avatars to explore different eye contact representations

Application Three: Animation Blending

(Gilles de la) Tourette syndrome is a childhood-onset neurodevelopmental disorder that is characterized by a wide range of motor and phonic tics and affects around 1% of children. While swearing tics (coprolalia), which dominate the public discussion of the syndrome, are rare, facial tics are common and can result in considerable social stigma and poor quality of life (review Robertson et al., 2017). For the vast majority (>90%) of those affected by Tourette's syndrome the tics are not completely involuntary, but fully or partially a voluntary response to premonitory urges (Leckmann et al., 1993).

This premonitory urge is used in therapy, by training sufferers to execute non-visible actions, for example a hand movement, instead of a visible facial tic. It also provides users with an opportunity to switch an avatar from live capture into an idle animation for the duration of a tic by pressing a button.

The third application allows the user to toggle whether the facial animation captured via Live Link is displayed by the character, controlled by a single button press. This feature is designed to help users mask the effects of a tic, providing a brief window, triggered by a premonitory urge, during which they can press the button to neutralise the animation.

When the button is pressed, the character's facial animation smoothly transitions to a neutral idle expression, temporarily halting the display of the user's captured facial data. Upon releasing the button, the animation transitions back to real-time capture from the user's face. This effect is achieved through animation blending across the entire face, with smooth interpolation of facial bone positions to ensure a seamless, non-jarring change between animations.

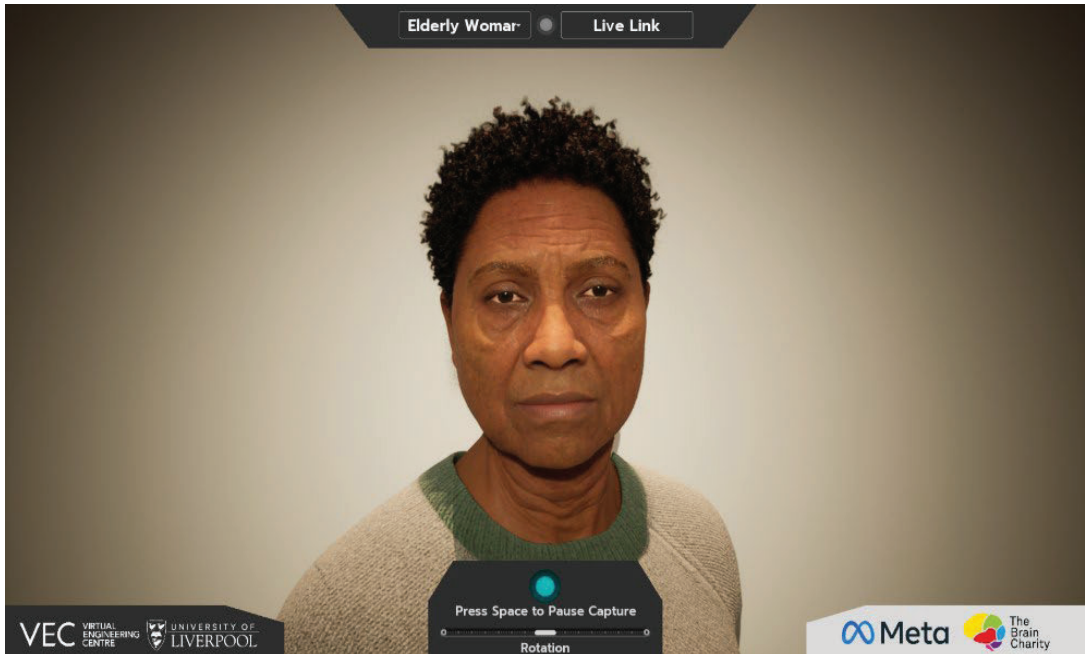


Figure 23. User interface for the animation blending interface. Users can press the space bar to temporarily replace the avatar animation from face capture by an idle animation to mask tics.

Discussion

Here we present three applications, using Unreal Meta-Humans for animation and Live Link for face capture that enable the systematic exploration of user experience and perception for controls that neurodiverse users of avatars *might* like to consider for interactions in computer mediated environments. The applications are designed to provide the basis for informed discussions in focus groups with neurodiverse individuals.

The same applications can provide the basis for qualitative and quantitative studies into the needs of neurodiverse individuals, and also for awareness-raising in the neurotypical population. They may enhance public debate about how neurodiverse individuals should be represented and be able to control their representations. Most importantly, they provide a pathway to directly involve potential end-users of the technologies in the design process in a principled and informed fashion.

We are very keen to share and further develop the software and collaborate with interested parties in studies that are aimed to enhance the understanding of what controls are required for the best possible user experience of neurodiverse people with avatars.

Acknowledgements

We are grateful to the Hanse Wissenschaftskolleg for providing a senior fellowship to GFM. The development work was funded by Meta and is supported by The Brain Charity.

References

- Cuenca-Martínez, F., Zapardiel-Sánchez, E., Carrasco-González, E., La Touche, R., & Suso-Martí, L. (2020). Assessing anxiety, depression and quality of life in patients with peripheral facial palsy: a systematic review. *PeerJ*, 8, e10449. <https://doi.org/10.7717/peerj.10449>
- Leckman, J. F., Walker, D. E., & Cohen, D. J. (1993). Premonitory urges in Tourette's syndrome. *The American journal of psychiatry*, 150(1), 98-102. <https://doi.org/10.1176/ajp.150.1.98>
- Robertson, M. M., Eapen, V., Singer, H. S., Martino, D., Scharf, J. M., Paschou, P., ... & Leckman, J. F. (2017). Gilles de la Tourette syndrome. *Nature reviews Disease primers*, 3(1), 1-20. <https://doi.org/10.1038/nrdp.2016.97>
- Roob, G., Fazekas, F., & Hartung, H. P. (1999). Peripheral facial palsy: etiology, diagnosis and treatment. *European neurology*, 41(1), 3-9. <https://doi.org/10.1159/000007990>
- Senju, A. & Johnson, M. H. (2009) The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13, 127–134.

AR Assistant for Pruning of Grapevines and Fruit Trees

Mariia Podguzova¹, Simon Häring¹, Jamiu Ojeleye¹, Stephan Krauß²,
and Didier Stricker^{1,2}

¹ Department of Computer Science, University of Kaiserslautern-Landau (RPTU),
Kaiserslautern, Germany

² Augmented Vision, German Research Center for Artificial Intelligence (DFKI),
Kaiserslautern, Germany

Corresponding author: mariia.podguzova@dfki.de

Keywords: Augmented Reality, Mobile Devices, Image-Based 3d Modelling, Tracking, Computer Vision, Deep Learning, Grapevine Pruning, Fruit Tree Pruning.

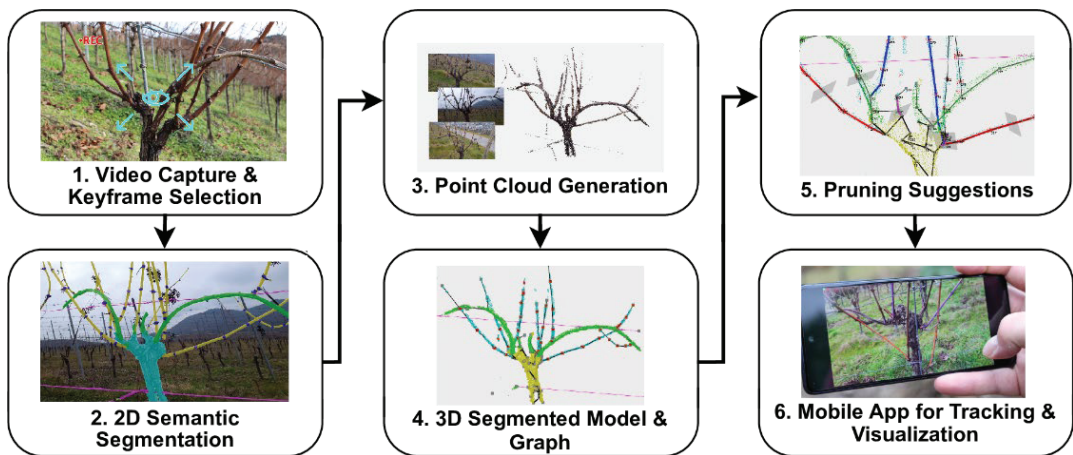


Figure 24. An overview of the pipeline that provides the user with pruning suggestions, given a video of the plant.

Introduction

Winter pruning in orchards is an essential but labor-intensive and time-consuming task, traditionally done manually to shape future growth by removing unwanted branches. Workers use various techniques to optimize yield, fruit quality or disease resistance. However, labor shortages and the need for professional training pose challenges for farmers. To address this, we offer support to make pruning more accessible to a broader segment of the workforce.

At the same time, computer vision in outdoor environments is complex due to varying lighting, weather, unique plant shapes, occlusions and similarity between foreground and background plants.

Multiple studies have addressed the problem of pruning grapevines and other fruit trees (Amatya, 2016; Botterill, 2017; Gentilhomme, 2023; Tong2023). These studies either focus on the entire automation pipeline (Botterill, 2017; Fourie, 2021) or separate steps, for instance, branch detection (Amatya, 2016; Zhang, 2018), reconstruction and skeletonization (You, 2022; Feng, 2024), and cut position localization (Marset, 2021). The recent approaches (Fourie, 2021; Gentilhomme, 2023) prove the general feasibility of automated pruning systems but do not address real-world challenges like outdoor conditions and complex, occluded plant structures. Effective pruning systems for fruit trees require accurate spatial information. Several studies have highlighted challenges in capturing thin structures using laser scanners or 3D cameras, often requiring additional refinement or proper initial registration, which can be time-consuming (Tagarakis 2013; Medeiros, 2017). To achieve a balance between cost and benefit, affordable methods for 3D reconstruction need to be explored. In contrast to similar studies that use 3D sensors for apple trees (Majeed, 2018; Tong, 2023), we explore the potential of image-based approaches suitable for an augmented reality (AR) pruning assistant on a mobile device.

In this work, we address the mentioned challenges and present an AR assistant that enables inexperienced workers to carry out pruning for grapevines and reduces the size of the cut wounds, making the plants more resilient to fungal infections and promoting rich and healthy yield. We further apply this concept to other fruit trees, such as apple and peach trees, and highlight the improvements made in this direction. Our contributions can be summarized as follows. First, we present a pipeline that extracts 3D and semantic information from a video of a plant and outputs pruning suggestions using both traditional and deep-learning methods (Vid2Cuts). Second, we introduce a mobile AR application to display the results to the user. Third, we extend the pipeline for other fruit trees that are more challenging compared to grapevines due to their more complex 3D structure and larger size.



Figure 25. Comparison of grapevines (left), peach trees (center) and apple trees in high-density orchards (right).

Vid2Cuts

We implemented a multi-step pipeline to extract pruning suggestions from smartphone videos of a grapevine plant (Häring, 2024). Figure 24 provides an overview of our approach. First, we automatically select a set of non-blurry frames, that are well spread throughout the video. This ensures good coverage of the entire plant from different angles. Next, we use a deep neural network (Pan, 2023) to identify different parts of the grapevine in the selected images via semantic segmentation. Using this information, we can also separate foreground from background in the images and use Meshroom (Griwodz, 2021) to generate a 3D point cloud of the plant. We combine the 2D semantic masks with the 3D information, by projecting the points into the masks via the intrinsic and extrinsic camera parameters also estimated by Meshroom. Doing so, every point is assigned the most common class it was projected onto. Next, the segmented point cloud is simplified into an abstract graph model. Each vertex in the graph has a 3D position and is connected by an edge to another vertex if there is a direct connection between them on the plant. The pruning rules defined by winemakers are then applied to this graph.

Using these, a pair of good fruit rods and a pair of cones for the following year are selected. They are highlighted blue and red respectively in the top-right of Figure 24. Based on this selection, a set of cutting planes is generated. Finally, the graph model and the pruning suggestions can be viewed on a smartphone in the field as an AR overlay on the camera feed.

Adaptation to Fruit Trees

We developed the first prototype for grapevines, which naturally exhibit a "flat" branching pattern. Moreover, the plant size is relatively small, allowing a view from one side to be sufficient for making pruning decisions. In contrast, fruit trees present more challenges due to their thin, long structures and numerous occlusions. Additionally, these trees have complex spatial structures that need to be captured from a 360-degree view on distinct levels due to their considerable height, see Figure 25 and 3. Furthermore, pruning rules for fruit trees differ significantly based on the type of fruit, the variety, and even the farming location. Therefore, we plan to offer a tool for experts to define new rule sets, making our system more flexible. This will require the extraction of high-level properties and sub-routines which can then be combined to define new rules.

Enhanced 3D Reconstruction

Traditional computer vision reconstruction approaches, such as Structure from Motion (Soatto1998), struggle to reconstruct thin branches crucial for pruning decisions for fruit trees.

The first challenge is precise camera localization. Feature extraction often fails due to the lack of texture on apple trees, requiring more overlapping images and increasing resource usage. We improved this by incorporating a hierarchical approach for camera localization (Sarlin2019). The main advantage is leveraging a global approach for the visual place recognition (Berton2023), combined with local feature extraction and matching, employing SuperPoint (DeTone2018) and SuperGlue (Sarlin2020) for a more robust and efficient feature matching, reducing the need for many overlapping keyframes.

The second challenge involves reconstruction of thin, occluded branches. We replaced Meshroom (Griwodz2021) with a Gaussian splatting-based approach (GS) (Kerbl2023), optimized for thin structures. Figure 3 shows the 3D models obtained using Meshroom and the Gaussian-based reconstruction approach. Due to the significant height of the trees, the keyframes exhibit considerable variation in their backgrounds. This makes feature matching challenging, even with the background removed, as the structures are very thin. This can be seen in the results obtained from Meshroom, which selected only the lower part of the tree for matching. By improving camera pose estimation, we achieved more accurate poses and a complete reconstruction using Gaussian-based methods.



Figure 26. A: Examples of keyframes. Images 1 and 2 show other trees in the background, while image 3 has a smooth sky, complicating feature matching. B: Dense point cloud from Meshroom. C: Mesh using GS.

Tracking

In the initialization phase of our tracking system, traditional methods like template matching were initially used to match 2D nodes in live frames with precomputed 2D-3D correspondences from a pre-selected keyframe. However, this approach is limited due to its lack of invariance to scale, rotation, and sensitivity to illumination variations. To enhance robustness, we replaced template matching with a more advanced method. Specifically, we integrated SuperPoint as our 2D feature extractor, leveraging its strong capability of descriptor generation. For matching these features, we utilized LightGlue (Lindemberger, 2023), resulting in accurate camera pose estimation even under challenging scenarios.

To address the inadequacy of using a single pre-selected keyframe for larger fruit trees, we compute an optimal set of keyframes that collectively cover the tree's features, with precomputed 2D-3D correspondences for each keyframe. Furthermore, selecting an initial reference frame from these multiple keyframes for accurate initialization is a challenge. To resolve this, we match the features of each keyframe against the live frame. The keyframe with the highest number of matching features is then selected as the initial reference frame, ensuring a more reliable and accurate initialization process.

Conclusion

In this work, we presented our first prototype for generating pruning suggestions based on hand-held monocular video of a grapevine, along with a mobile AR application to visualize cut positions. The app is intended for vineyard use, enabling inexperienced workers to perform accurate pruning.

We addressed challenges encountered during development and extended our solution to larger and more complex fruit trees such as apple and peach trees. We enhanced 3D reconstruction as well as tracking in the AR mobile application making it more efficient and accurate for thin and long branches. The development of the extended AR mobile app for visualization of pruning suggestions for fruit trees is kept as future work.

Acknowledgement

This work was partially funded by the AI:Pruning project and partially funded by the EU project CORTEX² (GA: Nr 101070192).

References

- Amatya, S., Karkee, M., Gongal, A., Zhang, Q. & Whiting, M. D. (2016). Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. *Biosyst. Eng.*, pp. 3–15. <https://doi.org/10.1016/j.biosystemseng.2015.10.003>
- Berton, G., Trivigno, G., Caputo, B. & Masone, C. (2023) Proceedings of the IEEE/CVF ICCV, pp. 11080-11090
- Botterill, T., Paulin, S., Green, R., Williams, S., Lin, J., Saxton, V., Mills, S., Chen, X. & Corbett-Davies, S. (2017). A robot system for pruning grape vines. *J. Field Robot.*, pp. 1100–1122, 2017. <https://doi.org/10.1002/rob.21680>
- DeTone, D., Malisiewicz, T. & Rabinovich, A. (2018). SuperPoint: Self-Supervised Interest Point Detection and Description. 2018 IEEE/CVF CVPRW, pp. 337-33712.
- Feng, W., Jiao, M., Liu, N., Yang, L., Zhang, Z. & Hu, S. (2024). Realistic reconstruction of trees from sparse images in volumetric space. *Computers & Graphics*, vol. 121, 103953. <https://doi.org/10.1016/j.cag.2024.103953>
- Fourie, J., Bateman, C., Hsiao, J., Pahalawatta, K., Batchelor, O., Misse, P. E. & Werner, A. (2021). Towards automated grape vine pruning: Learning by example using recurrent graph neural networks. *Int. J. Intell. Syst.* <https://doi.org/10.1002/int.22317>
- Gentilhomme, T., Villamizar, M., Corre, J. & Odobez, J. (2023). Efficient Grapevine Structure Estimation in Vineyards Conditions, in 2023 IEEE/CVF, ICCVW, pp. 712-720.
- Griwodz, C., Gasparini, S., Calvet, L., Gurdjos, P., Castan, F., Maujean, B., De Lillo, G. & Lanthony, Y. (2021). AliceVision meshroom: An open-source 3D reconstruction pipeline. in *Proc. 12th ACM MMSys*, 241–247. <https://doi.org/10.1145/3458305.3478443>
- Häring, S., Folawiyo, S., Podguzova, M., Krauß, S. & Stricker, D. (2024). Vid2Cuts: A Framework for Enabling AI-guided Grapevine Pruning. *IEEE Access* 12, 5814-5836. <https://doi.org/10.1109/ACCESS.2024.3350432>
- Kerbl, B., Kopanas, G., Leimkühler, & Drettakis, G. (2023) 3d gaussian splatting for real-time radiance field rendering, *ACM Trans. Graph.*
- Lindenberger, P., Sarlin, P. & Pollefeys, M. (2023). LightGlue: Local Feature Matching at Light Speed. *ICCV*.
- Majeed, Y., Zhang, J., Zhang, X., Fu L., Karkee, M., Zhang, Q. & Whiting, M. D. (2018). Apple tree trunk and branch segmentation for automatic trellis training using convolutional neural network based semantic segmentation. *IFAC-PapersOnLine*, pp. 75–80. <https://doi.org/10.1016/j.ifacol.2018.08.064>
- Marset, W. V., Pérez, D. S., Díaz, C. A. & Bromberg, F. (2021). Towards practical 2D grapevine bud detection with fully convolutional networks. *Comput. Electron. Agricult.* <https://doi.org/10.1016/j.ifacol.2018.08.064>
- Medeiros, H., Kim, D., Sun, J., Seshadri, H., Akbar, S.A., Elfiky, N.M., Park, J. (2017). Modeling dormant fruit tree for agricultural automation. *J. Field Robot* 1203–1224. <https://doi.org/10.1002/rob.21679>
- Pan, H., Hong, Y., Sun, W. & Jia, Y. (2023). Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes, *IEEE Transactions on Intelligent Transportation Systems*, 3448–3460. <https://doi.org/10.1109/TITS.2022.3228042>
- Sarlin, P., Cadena, C., Siegwart, R. & Dymczyk, M. (2019). From Coarse to Fine: Robust Hierarchical Localization at Large Scale. In 2019 IEEE/CVF CVPR, pp. 12708-12717.

- Sarlin, P., DeTone, D., Malisiewicz, T. & Rabinovich, A. (2020). SuperGlue: Learning Feature Matching with Graph Neural Networks. 2020 IEEE/CVF CVPR, pp. 4937-4946.
- Soatto, S. & Brockett, R. (1998). Optimal structure from motion: local ambiguities and global estimates. Proceedings. IEEE Computer Society Conference on CVPR, pp. 282-288,
- Tagarakis, A., Liakos, V., Chatzinikos, T., Koundouras, S., Fountas, S. & Gemtos, T. (2013). Using laser scanner to map pruning wood in vineyards. *Precis. Agric.* 633–639.
- Tong, S., Zhang, J., Li W., Wang, Y. & Kang, F. (2023). An image-based system for locating pruning points in apple trees using instance segmentation and RGB-D images, *Biosystems Engineering*, pp. 277-286. <https://doi.org/10.1016/j.biosystemseng.2023.11.006>
- You, A., Grimm, C., Silwal, A. & Davidson J.R. (2022). Semantics-guided skeletonization of upright fruiting offshoot trees for robotic pruning. *Comput. Electron. Agricult.*, 2022. <https://doi.org/10.1016/j.compag.2021.106622>
- Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X. & Gao, Z. (2018). Branch detection for apple trees trained in fruiting wall architecture using depth features and regions-convolutional neural network (R-CNN). *Comput. Electron. Agricult.*, pp. 386–393. <https://doi.org/10.1016/j.compag.2018.10.029>

A Simple but Effective AR Framework for Human-Object Interaction

Hong Son Nguyen¹, Andrew Chalmers², DaEun Cheong¹, Myoung Gon Kim¹,
Taehyun Rhee², and JungHyun Han¹

¹Computer Science Department, Korea University, Seoul, Korea

²School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

*These authors contributed equally to this work.

Corresponding authors: Hong Son Nguyen¹, Andrew Chalmers²

Keywords: Augmented Reality, Human-Object Interaction

Introduction

In this paper, we present an augmented reality (AR) system that captures the full body of a user with an RGB camera and facilitates interaction between the user and 3D virtual objects. Figure 27a illustrates our AR environment, where a user standing on the rectangular workspace (colored in pink) is captured by an RGB camera. The large screen on the left side displays the mirrored view of the environment that is augmented with two virtual objects, a green ball and a blue box, so that the user can interact with them, as shown in Figure 27b.



Figure 27. AR interaction: (a) The user can see the virtual objects on the screen. (b) The virtual objects appear located away. Walking close to the ball, the user picks it up. Walking toward the box, the user hits the box with the ball.

Methods

Facilitating interactions between a real human and 3D virtual objects requires real-time estimation of the human's 3D pose and shape. For this, we use HMR (human mesh recovery) (Kanazawa et al., 2018). It estimates the body pose and shape in the form of SMPL-X (Pavlakos et al., 2019), a parametric human model that represents the body as a differentiable function. Then the 3D mesh and the 2D/3D joints are extracted from the estimated parameters.

Just for the sake of exposition, we call the estimated user a *character*. It needs to be tested for collision and occlusion with the virtual objects, making it essential to place the character into the 3D camera space as accurately as possible. For this, we first estimate the user's depth, i.e., the z-coordinate, in the camera space using the 2D joints' pixel coordinates. In Figure 28a, l_i denotes the height of the captured image (in pixels) and l_u denotes the length between two of the user's joints (also in pixels) as an indicator of the user's height. We select a few candidate pairs of joints (foot to knee, pelvis to neck, etc.) and choose the one that maximizes its vertical orientation to provide a robust indication of the user's height. As the user moves around, l_u is continuously computed. The farther away the user is from the camera, the smaller l_u is. Let L denote l_u/l_i .

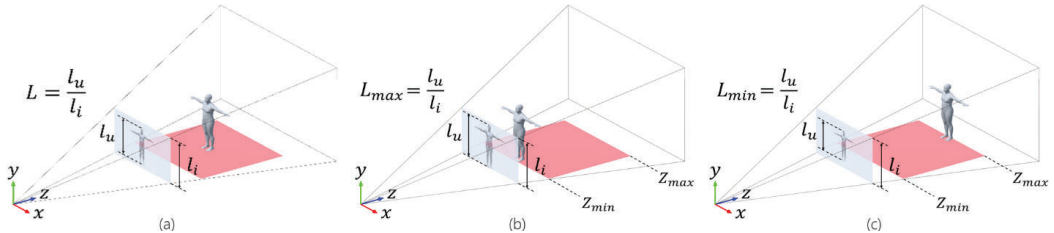


Figure 28. Computing the depth of the 3D character: (a) The normalized length, L , is defined as l_u/l_i , where l_i is fixed but l_u changes. (b) When the user stands at z_{min} we get L_{max} . (c) When the user stands at z_{max} we get L_{min} .

In Figure 28b, where z_{min} and z_{max} represent the depth boundaries of the workspace, the user stands at z_{min} . Then, L reaches the largest possible value, which we denote as L_{max} . On the other hand, Figure 28c shows the smallest possible value, L_{min} , which is obtained when the user stands at z_{max} . Note that $[L_{min}, L_{max}]$ is linearly mapped to $[z_{max}, z_{min}]$. Through this mapping, L defined in $[L_{min}, L_{max}]$ is converted into the camera-space depth, z , in $[z_{max}, z_{min}]$.



(a) Collision detection - Case A



(b) Occlusion - Case B

Figure 29. Collision and occlusion: (a) Collision is detected between the ball and the capsule colliders. (b) The character is represented as a mesh extracted from the SMPL-X model, and the ball behind it appears to be occluded.

User Study

Our study is designed to contrast the effects of incorporating a full-body shape model in an AR environment against a scenario that lacks such detailed shape information. We define "Case A" as the

baseline approach, which uses 3D pose estimation and employs *colliders* based on the skeletal information. In the current study, we assign a capsule collider to each bone, as shown in Figure 29a. In contrast, we set our method as “Case B,” which takes the estimated full-body shape, i.e., the character’s mesh extracted from the SMPL-X model is used directly for collision and occlusion handling. Figure 29b shows occlusion handling with Case B. In our study, we hypothesize that Case B enables more natural handling of collision and occlusion.

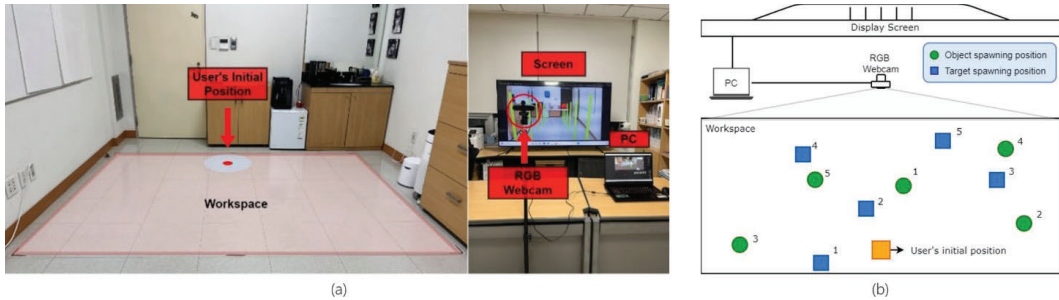


Figure 30. User study room: (a) AR setup. (b) In this top view of the room, the objects and targets are spawned in evenly-distributed predefined positions, and users are asked to grab an object and hit the target with it.

Figure 30a shows our user study room, and Figure 30b depicts its top view, where the green circles represent the object-spawning positions and the blue squares represent the target positions. We have five pairs of object and target with the same number, and a pair is spawned at a time. Participants are asked to grab the object and hit the target box with it, after which they are to return to their initial position to perform the next task with a new pair. The tasks follow a 2×1 within-subjects design, where participants perform identical tasks in both cases. The arrangement of five pairs is in counterbalanced order (by using Latin square method).

In our user study, we recruited 43 participants, consisting of 15 males and 28 females, aged between 20 and 35 ($M=24.07$, $SD=3.68$). Half of the participants had prior experiences with AR applications, while the other half did not. Participants took turns performing Cases A and B, where half of the participants started with Case A and the other half started with Case B. After completing each case, participants filled out post-test questionnaires: the NASA Task Load Index (NASA-TLX) (Colligan et al., 2015) and the System Usability Scale (SUS) (Brooke et al., 1996). We also measured the completion time for each task to obtain additional objective data to evaluate the participants’ performance.

Results

As presented in Table 2, the conducted t-test for NASA-TLX fails to reveal the significant differences between the cases in terms of workload, but SUS evaluation indicates that Case B offers better system usability compared to Case A with statistical significance.

The mean completion time is 24.06 ($SD = 17.86$) for Case A and 21.77 ($SD = 13.90$) for Case B. The F-test detects inequality between the cases, and Welch’s t-test (unequal variance t-test) is conducted. The result reveals that there is a difference between the two cases in terms of completion time with statistical significance ($t = 2.10$, $p < 0.05$). Further analysis is made to investigate the effects of the different object/target positions on the completion time. For each position, F-test between two cases is conducted. Once an F-test reveals the inequality of variances, a Welch’s t-test is conducted. Otherwise, a t-test assuming equal variances is conducted. The results are shown in Table 3. The average completion time for Case A is longer than that of Case B at position 3 with statistical significance, which is the most

challenging task as it requires the users to travel the longest distance. This implies that when the task is more challenging, using the shape (Case B) can assist in completing the task more efficiently.

Table 2. NASA-TLX/SUS Assessment.

		Case A	Case B	
NASA-TLX↓	<i>M</i>	52.44	47.40	<i>p</i>
	<i>SD</i>	14.94	14.08	>0.05
SUS↑	<i>M</i>	64.94	74.42	
	<i>SD</i>	20.09	13.87	*<0.05

Table 3. Completion time in different positions.

Position		Case A	Case B	<i>p</i>
1	<i>M</i>	22.72	21.99	>0.05
	<i>SD</i>	14.14	12.27	
2	<i>M</i>	25.87	23.52	>0.05
	<i>SD</i>	11.44	15.62	
3	<i>M</i>	25.27	20.34	*<0.05
	<i>SD</i>	20.43	13.45	
4	<i>M</i>	24.78	22.39	>0.05
	<i>SD</i>	22.04	15.01	
5	<i>M</i>	21.19	21.48	>0.05
	<i>SD</i>	10.27	13.34	

Conclusion

This paper introduces a practical framework that enables natural full-body interactions with 3D virtual objects in AR environments using only a conventional RGB video camera. Our framework facilitates real-time interaction by providing accurate depth cues for enhanced collision and occlusion management. Through our user study, we have quantified the impact of our method on full-body interactions, focusing on user interaction and handling of collisions and occlusions with the virtual objects. The comparative analysis of two prototypes—one without and one with the detailed shape model—revealed enhanced interactivity and occlusion handling afforded by the detailed shape information.

While our approach marks a significant advancement in facilitating full-body human interaction in AR using a single RGB camera, there is potential for further refinement. Our current implementation does not achieve pixel-wise alignment between the 3D human model and the captured video, occasionally resulting in occlusion artifacts and inaccuracies in collision detection. An advanced synchronization module could significantly enhance the congruence between the 3D model and video frames, thus improving the quality of interaction. Future research, which may incorporate state-of-the-art depth-sensing technologies, is poised to provide further insights into the nuances of full-body interaction in AR environments.

Acknowledgements

This work was supported by ICT Creative Consilience Program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (IITP-2024-RS-2020-II201819) and the grant no. 2020-0-00861.

References

- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189 (194), 4-7.
- Colligan, L., Potts, H. W., Finn, C. T., & Sinkin, R. A. (2015). Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International journal of medical informatics*, 84(7), 469-476. <https://doi.org/10.1016/j.ijmedinf.2015.03.003>
- Kanazawa, A., Black, M. J., Jacobs, D. W., & Malik, J. (2018). End-to-end recovery of human shape and pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7122-7131.
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10975-10985.

Using Augmented Reality and Machine Learning for Captioning in Theatrical Experiences

Olga Chatzifoti¹, Nick Achilleopoulos¹, Spiros Borotis¹, Moonisa Ahsan²,
and Eleni Oikonomou³

¹*Maggioli Spa, Research & Development Department, Athens, Greece*

²*Centrum Wiskunde en Informatica (CWI), Amsterdam, The Netherlands*

³*Athens-Epidaurus Festival, Athens, Greece*

Corresponding author: olga.chatzifoti@maggioli.gr

Keywords: Augmented Reality, Machine Learning, Theatre, Cultural heritage, Captions

Introduction

In this work, we present an Augmented Reality Theatre System, exemplifying the benefits of using new technologies for enhancing audience engagement and directors' creative practice. Our work aims to improve accessibility for an international and diverse audience through personalized captions and to creatively support a new generation of directors through hybrid multimedia. To achieve that, we design and develop a system using augmented reality (AR) and machine learning (ML) technologies. In this paper, we present the personalized captions feature of the system. This work has been undertaken in the context of VOXReality (www.voxreality.eu), a European Union-funded Research and Innovation Action.

Motivation

Our goal is to make the theatre-going experience more accessible to a wider audience through the provision of accurate, real-time, customizable captions, translated into the viewer's preferred language. This is an especially relevant need for cultural works in languages with small speaker communities, like Greek, and even more important for historical works, such as Ancient Greek plays. To support this goal, we opted to use AR for content delivery following literature recommendations, and ML for the generation, synchronization and translations of the captions to provide automation as well as extended language support over the current practice.

Related Work & Background

The established practice for theatre caption delivery has two formats: open captions, displayed on large screens or projected on stage, and closed captions, delivered on viewers' smartphones or

dedicated devices. Closed captions offer improved accessibility because they can support customization options. Closed captions delivered in AR can offer an improved user experience by allowing the user to focus on a single visual frame, potentially reducing mental workload (Rzayev et al., 2018, 2020). Related work in AR captioning for theatre includes the SmartSubs¹ project by Institute of Informatics & Telecommunications (IIT) at NCSR Demokritos, and the Thearto Project² which also investigated audiovisual effects in the performance. The Royal National Theatre in London, UK has already established a monocular AR closed captions system as part of their public accessibility initiative³. Industry solutions that support ML generated AR captions, like XRAI⁴, are also appearing. This work aims to improve closed AR captions for theatrical performances by investigating optimized ML workflows and extended user customization options for AR display.

System Description

Our system uses a server-client architecture adopting a secure WebSocket protocol for communication between server-client and a secure REST API for communication with the ML models, developed in VOXReality (Maniatis et al., 2023). The server directs the audio signal from the stage's microphones to the ML model and distributes the text responses to the XR clients in the audience for AR display. The ML model (Issam et al., 2024; Maka et al., 2024) uses a transformer architecture and performs audio transcription and text matching to provide pre-formatted captions, ensuring the least latency and best accuracy, in addition to executing translation demands. All VOXReality ML models can be found available online, promoting Open Science⁵.

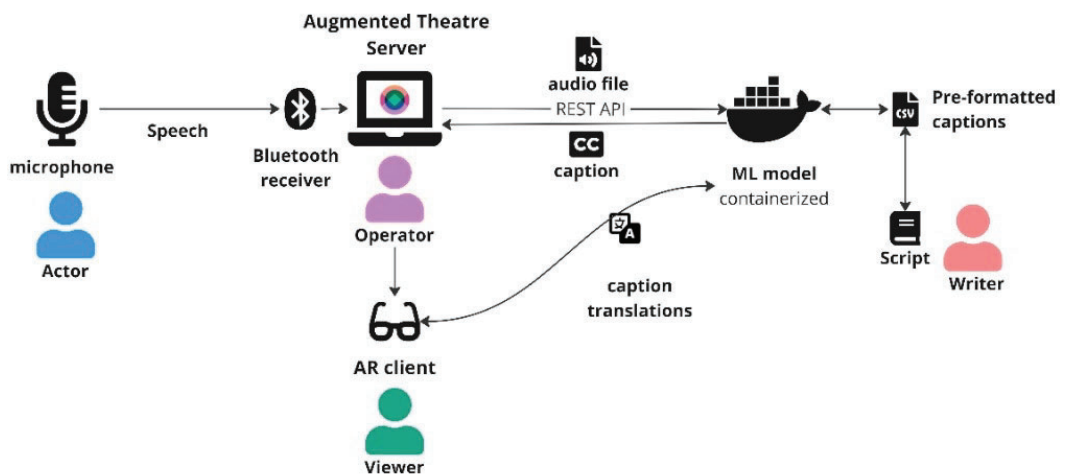


Figure 31. System Architecture for AR Theatre.

Figure 31 describes the system architecture. The AR application allows the user to customize the caption display extensively for improved readability and additionally displays audiovisual effects, as designed and orchestrated by the theatrical director. Figure 32 presents screenshots of the AR application user interface for caption customization and placeholder AVFXs that were used during evaluation.



Figure 32. Left: screenshot from AR customization UI. Right: screenshot from caption display

Evaluation

The first user evaluation was completed in May 2024 in Athens, Greece with 12 users recruited through email invitations with users attending a live performance of an excerpt of the Ancient Greek play 'Hippolytus' by Euripides. The focus was on estimating users' behavioural intention toward this system. Alongside that, other parameters of usability (System Usability Scale (Lewis, 2018), NASA Task Load Index (Hart, 1986) and Simulator Sickness Questionnaire (Kennedy et al., 1993)) were evaluated, and a semi-structured interview in pairs of participants was performed to capture more nuanced feedback. Application data logs were used to contextualize user responses and were complemented by facilitator observations. Questionnaire results were analysed using statistical analysis and the interview transcripts using thematic analysis (Clarke & Braun, 2017). The second user evaluation will be delivered in May 2025 and will be advertised for user recruitment on public communication channels.

Results & Discussion

Despite being at an early technical and aesthetic level, the initial evaluation was decisively positive with users stating that they would be interested in attending this form of theatre in real conditions and that they saw practical benefit and artistic merit in the provided features. Negative feedback was focused on the technical performance of the system and the learning curve of the AR application. Table 4 presents a summary of key results focusing on the custom Behavioural Intention Questions - the extended results of the full evaluation are beyond the scope of this paper. Redesign for pilot 2 will address the initial user feedback and provide a renewed evaluation.

Table 4. Pilot 1 results on Behavioural Intention Questions

Question	Mean	Standard Deviation	Median
Recommend AR theatre	4.3	1	5
Prefer dynamic subtitles	3.3	0.9	3.5
Prefer static subtitles	3.8	1	4
Would like more VFX	4.5	0.7	5
General Experience rating	4.3	0.5	4

Conclusion

This work showcases the potential of augmented reality and machine learning to address the needs of diverse audiences and offer a more inclusive augmented theatre experience. The initial evaluation of

our system demonstrated promising user engagement and interest, while also revealing areas for technical improvement. The redesign of the system for the next pilot will aim for improvements in latency, accuracy and translation quality for the ML components, and a more user-friendly experience for the AR components. As we move toward the next iteration of user testing, outcomes are expected to advise theatres' guidelines for performances of international appeal, provide tested technical solutions for wider adoption, and future research recommendations for blending cultural heritage with cutting-edge technology.

Acknowledgement

VOXReality (www.voxreality.eu) is funded by the European Union under grant Agreement No. 101070521. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or Directorate-General for Communications Networks, Content and Technology. Neither the European Union nor the granting authority can be held responsible for them. More details: <https://cordis.europa.eu/project/id/101070521>

References

- Clarke, V., & Braun, V. (2017). Thematic analysis. *J. Posit. Psychol.* 12(3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>
- Hart, S. G. (1986, January). NASA Task Load Index (TLX). <https://ntrs.nasa.gov/citations/20000021487>
- Issam, A., Can Semerci, Y., Scholtes, J., & Spanakis, G. (2024). Fixed and Adaptive Simultaneous Machine Translation Strategies Using Adapters. In E. Salesky, M. Federico, & M. Carpuat (Eds.), *Proc. 21st Int. Conf. Spoken Language Translation (IWSLT 2024)* (pp. 298–310). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.iwslt-1.36>
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *Int. J. Aviation Psychol.* 3(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
- Lewis, J. (2018). The System Usability Scale: Past, Present, and Future. *Int. J. Human-Computer Interaction*, 1–14. <https://doi.org/10.1080/10447318.2018.1455307>
- Maka, P., Semerci, Y., Scholtes, J., & Spanakis, G. (2024). Sequence Shortening for Context-Aware Machine Translation. In Y. Graham & M. Purver (Eds.), *Findings of the Association for Computational Linguistics: EACL 2024* (pp. 1874–1894). Assoc. Comput. Ling. <https://aclanthology.org/2024.findings-eacl.127>
- Maniatis, A., Bourou, S., Anastasakis, Z., & Psychogios, K. (2023). VOXReality: Immersive XR Experiences Combining Language and Vision AI Models. *Human Interaction and Emerging Technologies (IHET-AI 2023): Artificial Intelligence and Future Applications*, 70(70).
- Rzayev, R., Hartl, S., Wittmann, V., Schwind, V., & Henze, N. (2020). Effects of position of real-time translation on AR glasses. *Proceedings of Mensch Und Computer 2020*, 251–257. <https://doi.org/10.1145/3404983.3405523>
- Rzayev, R., Woźniak, P. W., Dingler, T., & Henze, N. (2018). Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–9. <https://doi.org/10.1145/3173574.3173619>

Cognitive Load in Mixed Reality: Technology Order and Prior Experience in Eye-Tracking

Paula López, Ana M. Bernardos, and José R. Casar

Information Processing and Telecommunication Centre, Polytechnic University of Madrid

Corresponding authors: Paula.lopez.alvarez@upm.es, anamaria.bernardos@upm.es,
joseramon.casar@upm.es

Keywords: Eye-Tracking, Cognitive Load, Mixed Reality, Wearable AR.

Abstract

Cognitive load has been identified as a key factor influencing user performance and interaction in mixed reality (MR) environments. In this paper, we examine how prior knowledge and experience with a task using standard web technology influence eye-tracking metrics and cognitive load when performing the same task in a mixed reality (MR) environment. Utilizing HoloLens 2, we conducted an empirical study with seventeen participants who tested a building occupancy monitoring application in both a web and MR environment. We systematically assessed eye-tracking metrics, including fixation frequency, duration, and saccadic movements, in relation to various participant factors such as age, visual condition, previous MR experience, and technology order (web vs. MR). Our findings suggest that transitioning from web to MR environments significantly impacts fixation frequency, which, in turn, correlates with cognitive load levels experienced by participants. Notably, the data indicate that participants began with web-based version of the application exhibited a higher fixation frequency in the data collection of our experiment, suggesting a potential reduction in cognitive load. These insights offer critical implications for optimizing learning procedures and user interface design of MR applications.

Introduction

Mixed Reality (MR) represents a transformative intersection of virtual and real-world environments, creating increasingly immersive user experiences across a wide array of sectors, from healthcare and education to industrial monitoring and smart cities. As MR applications evolve and become more sophisticated, effectively managing cognitive load—defined as the mental effort required to perform a task (Sweller, 2011)—is essential for maintaining user efficiency, satisfaction, and overall engagement. Elevated cognitive load has been linked to diminished task performance, increased user fatigue, and negative impacts on the overall experience. These factors can significantly hinder the full adoption and efficacy of MR technologies, particularly in high-stakes environments where user performance is critical.

Eye-tracking has emerged as a valuable tool in the field of cognitive load assessment, offering non-invasive and real-time insights into user behavior and interactions. By analyzing metrics such as fixation frequency, duration, and saccadic movements, researchers can gain a deeper understanding of how users allocate their attention and process information within MR environments. This work extends prior research conducted by (López et al., 2024), which examined cognitive load and eye-tracking metrics in MR contexts without considering the critical factor of task sequencing.

In our study, we introduce technology order as a significant variable, positing that the sequence in which participants interact with different versions of the application—transitioning from web-based version to the MR environment—may have a profound influence on cognitive load. The rationale behind this hypothesis stems from the cognitive theory of multimedia learning, which suggests that the manner in which information is presented can impact cognitive processing efficiency. Specifically, we aim to investigate how prior knowledge of the environment can establish a cognitive baseline that affects subsequent performance on MR tasks.

Our analysis focuses on eye-tracking data collected via the HoloLens 2, a state-of-the-art wearable MR system that enables immersive interactions with digital content. The study is applied to an MR application specifically designed for building occupancy monitoring, as discussed in (Zambrana et al., 2024). This application allows users to interact with an interactive map, retrieve pertinent indicators, and visualize customized charts for data-driven decision-making. By providing an ideal context for studying the relationship between task technology and cognitive load, this research aims to contribute valuable insights that can inform the design of more effective MR applications, ultimately enhancing user experience and performance in complex operational environments.

Methodology

The user study involved a comparative analysis with 17 participants during the "discovery phase" of the wearable mixed reality (MR) application previously described. This phase was crucial for understanding how users interacted with the MR environment. The participants, comprising 7 females and 10 males, were aged between 18 and 65 years and were recruited from the university community. The sample included a diverse group of individuals, including students, researchers, and their relatives, providing a broad range of user experiences. Of the participants, 6 wore glasses, 2 used contact lenses, 1 had undergone cataract surgery, and 9 reported no visual impairments.

Regarding prior experience with augmented reality (AR) and MR technologies, 5 participants had prior experience with AR, though only 1 had experience using the HoloLens 2. None of the participants had previously interacted with the specific MR application designed for this study, ensuring that all users started with a similar baseline in terms of familiarity with the application environment.

Participants were divided into two groups based on technology order, with one group beginning in the web environment before transitioning to MR, while the other group completed the tasks in reverse order (starting with MR and moving to the web). This division allowed for a comparative analysis of how technology sequencing influenced cognitive load and user interaction within MR.

Prior to beginning the study, all participants were given a tutorial on using the HoloLens 2, ensuring they understood the basic functionality and interaction methods of the device. Following the tutorial, a calibration process was conducted to optimize visual clarity for each participant, ensuring consistency across the study in terms of user experience and the accuracy of eye-tracking data collection.

During the discovery phase of the study, participants navigated an augmented building representation using only their gaze for a duration of 1-2 minutes. This phase was designed to allow

participants to familiarize themselves with the MR environment, thereby focusing purely on gaze-based interaction. Eye-tracking data was collected throughout the discovery phase using the open-source ARETT (Augmented Reality Eye Tracking Toolkit) tool (Kapp et al., 2021), which recorded key data points such as the x, y, z coordinates of each gaze point, the relative timestamp, and the Area of Interest (AOI) where each gaze was fixed. As shown in Figure 33, the Areas of Interest (AOIs) were pre-defined based on key elements in the MR environment, such as specific interactive objects or data points within the building monitoring application. Each gaze point was recorded and assigned to its corresponding AOI for further analysis.

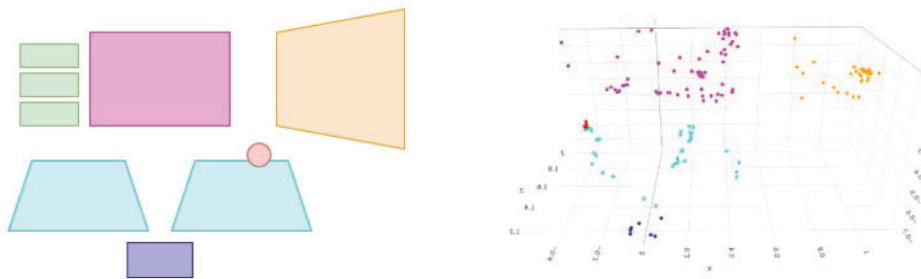


Figure 33: (a) AOI's marked by colours for fixation classification (on the left) and (b) Distribution of fixation gaze points by AOI (on the right)

To analyze the eye-tracking data collected during the study, a custom Python implementation was developed to identify and classify fixation and saccadic movements based on AOI's. The classification process was designed to distinguish between fixation points and saccades by first determining whether each gaze point fell within a predefined AOI. This method provided a structured approach to understanding how participants interacted with specific regions of the mixed reality (MR) environment. The analysis followed these steps. If a gaze point fell within an AOI, it was preliminarily classified as part of a fixation, while gaze points outside any AOI were classified as saccades, representing the rapid eye movements between fixation points. Consecutive gaze points within the same AOI were grouped together as part of a potential fixation, allowing for the identification of continuous fixations where the participant's gaze remained focused on a specific area of the MR environment over time. For a sequence of gaze points to be confirmed as a fixation, the total duration of those points had to exceed a predefined minimum fixation duration (100 ms). If the combined duration of the gaze points within the AOI did not meet this threshold, those points were reclassified as saccades. This ensured that only sustained visual attention on a specific AOI was counted as a fixation, reflecting the mental effort and cognitive processing associated with that area.

By employing this method, we were able to derive detailed insights into how participants allocated their attention within the MR environment, focusing on key areas and understanding the relationship between technology sequencing and cognitive load. Moreover, the custom Python script facilitated this nuanced analysis, which was critical for interpreting the eye-tracking data in relation to the cognitive demands.

Analysis and Results

After pre-processing the data to remove missing values and noise, ensuring all fixations lasted at least 100ms, we performed statistical analysis and applied a K-Prototypes algorithm to categorize cognitive load levels. Fixation duration had the largest coefficient of variation (CV) at 45.53%, indicating

moderate variability among subjects, followed by saccade speed (26.3%), fixation frequency (11.57%), and saccade duration (3%).

The participants were divided into groups based on several factors: visual condition (with or without visual impairments), prior MR experience (with or without), and four age categories (18-25 years, 26-35 years, 46-55 years, and 56-65 years). Additionally, technology order (whether participants completed the web task first or the MR task first) was also considered a key variable. To test differences between these groups, a Mann-Whitney U test was conducted for experience, visual condition, and technology order groups, while a Kruskal-Wallis test was applied for age categories. A significant result was observed for fixation frequency in the technology order grouping ($U=63.0$, $p=0.007$), suggesting that technology order influences fixation frequency and potentially cognitive load. Although not statistically significant, the result for saccade duration in the grouping by previous MR experience ($U=48.0$, $p=0.063$) suggests a possible relationship for further study.

Following the application of the K-Prototypes clustering algorithm, two distinct groups emerged based on their eye-tracking metrics. Further analysis revealed a significant difference in fixation frequency between these groups ($U=70.0$, $p=0.0001$). Additionally, a Chi-Square test showed a significant association between group membership and technology order ($\chi^2(4, N=17) = 4.74$, $p=0.029$), highlighting that participants' cognitive load, as indicated by their eye-tracking behavior, was influenced by their prior knowledge of the domain in MR.

As shown in table 1 below, Group A and Group B have several differences in terms of eye-tracking metrics. Participants in Group B demonstrated an average fixation duration longer than those in Group A (1066.1 ms vs. 754.88 ms). A longer fixation duration typically reflects greater cognitive effort or difficulty in processing visual information, suggesting that Group B participants required more mental effort to engage with the tasks. Additionally, while both groups had similar saccade durations, Group B showed a higher saccade speed (0.49×10^{-3} ms vs. 0.40×10^{-3} ms for Group A). Faster saccades could indicate heightened cognitive load as participants processed information more quickly, potentially compensating for the increased difficulty reflected by their longer fixation durations.

Table 5. Centroids of the groups for each eye-tracking metric

	Fixation Duration (ms)	Saccade Duration (ms)	Saccade Speed ($\times 10^{-3}$ ms)	Fixation Frequency (fix/ms)
Group A	754.88	37.69	0.40	0.0022
Group B	1066.61	37.11	0.49	0.0018

As illustrated in Figure 34 below, most subjects in Group A were those who initially tested the application in its web version, as well as participants with prior experience in mixed reality environments. This suggests that both previous experience in MR settings and familiarity with the application, even in a different format, significantly influence the eye-tracking data. Consequently, these factors play a crucial role in the estimation of cognitive load.

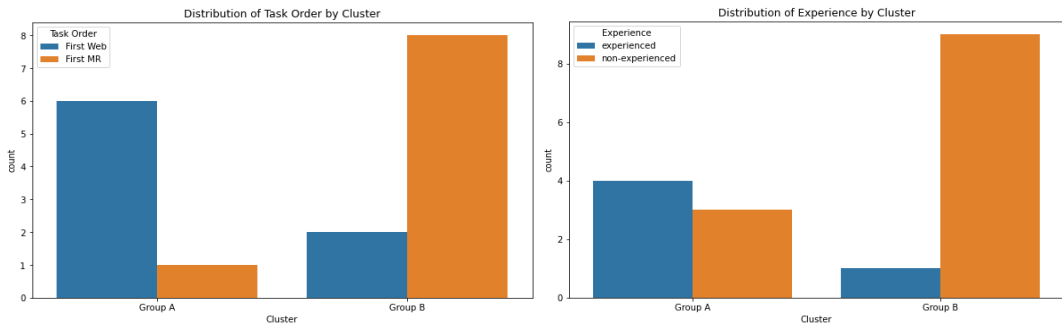


Figure 34: (a) Distribution of Technology order by cluster (on the left) and (b) Distribution of Experience by cluster (on the right)

Conclusions and further work

The results of this study underline the critical impact of prior knowledge on cognitive load during exposure to mixed reality environments. The clustering analysis highlighted that participants with the highest fixation frequencies were primarily those who had first accomplished the tasks to be completed through standard web technology, as well as individuals with prior experience in mixed reality environments. This indicates that both domain familiarity and previous exposure to MR technologies significantly contribute to easing cognitive demands. Specifically, participants who used the web version before transitioning to MR demonstrated increased fixation frequencies, suggesting that familiarity with a known interface may reduce cognitive effort in more complex environments. This finding can be useful to program technology transitioning between systems, as it suggests that prior training on the tasks to be completed using well-known technologies can mitigate learning effects of MR environments.

Among the study limitations, the dependency on specific headsets used in the study is one of them. As emerging technologies, such as the Apple Vision Pro, Meta Quest 3 or Orion by Meta offer higher resolution and improved ergonomics, their adoption could lead to different cognitive load experiences. Therefore, the findings from this study may not fully generalize to these new technologies. Additionally, the study was constrained by a relatively small sample size, which included a demographic skewed towards participants under 35 years old. This limitation may restrict the applicability of the results to older populations, which often exhibit different cognitive processing patterns. Expanding the sample size and including more diverse demographics would not only strengthen the generalizability of these findings but also provide deeper insights into how various age groups interact with MR environments.

Future research should explore additional eye-tracking metrics, such as pupil dilation and blink rate, and their connections to broader cognitive and physiological factors. Investigating these metrics could lead to a more comprehensive understanding of cognitive load in mixed reality environments. Furthermore, longitudinal studies could be beneficial to assess how repeated exposure to MR applications influences cognitive load over time.

Acknowledgements

Authors acknowledge funding under grants MIA.2021.M04.0008.E16 (Spanish Min. of Economic Affairs and Digital Transformation and by EU Next Generation EU/PRTR), PDC2021-121567-C21 and subsequent PID2023-151605OB-C21 (Spanish Ministry of Science, Innovation and Universities). Additionally, the authors would like to thank G. Cobo for her support in the data collection experiments.

References

- Kapp, S., Barz, M., Mukhametov, S., Sonntag, D., & Kuhn, J. (2021). ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays. *Sensors* (Basel, Switzerland), 21(6), 2234. <https://doi.org/10.3390/s21062234>
- López, P., Bernardos, A. M., & Casar, J. R. (2024). Eye-Tracking Analysis for Cognitive Load Estimation in Wearable Mixed Reality. *ACM Symposium on Spatial User Interaction (SUI '24)*. <https://doi.org/10.1145/3677386.3688885>
- Sweller, J. (2011). CHAPTER TWO - Cognitive Load Theory. En J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (Vol. 55, pp. 37-76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Zambrana, G., Barbolla, A., Bergesio, L., & Casar, J. (2024). A wearable-AR service for data-driven scenarios. 1-3. <https://doi.org/10.1145/3656650.3656732>

Augmented telecommunication in factory setting

Narek Minaskan, Bastian Kraymer, Alain Pagani, and Didier Stricker

German Research Center for Artificial Intelligence, Kaiserslautern, Germany

Corresponding author: nareg.minaskan_karabid@dfki.de

Keywords: XR System Architecture, Collaborative And Distributed XR, Industrial Applications

Introduction

As a result of the COVID-19 pandemic, the possibility of remote work or “Home Office” has been normalized by many companies. Although the existing telecommunication tools are sufficient for many mainstream tasks, they lack the capabilities of 3D interaction which is necessary for complex tasks which require physical presence for engaging in problem solving.

As an early implementation of collaborative 3D communication tools, Microsoft (Chen et al., 2025) developed the HoloLens system, which introduces a novel interaction model for supporting collaboration between a head-mounted display (HMD) user and remote participants. The HoloLens allows remote companions to join the AR space by hitching onto the view of the primary HMD user through Skype-enabled devices, such as tablets or PCs. This system facilitates asynchronous interaction in a shared 3D space with digital objects, allowing remote parties to contribute to tasks and have their inputs reflected back to the primary user in real-time, thus enabling new scenarios for remote collaboration.

Further advancements in remote collaboration systems have focused on complex tasks like environmental pollution analysis, which require expertise from multiple fields. One such system was designed by Mahmood et al. (2019). It uses mixed reality to support co-presence and collaborative analysis, demonstrating improved remote analysis through shared user and data spaces. Drey et al. (2022) explored how the benefits of pair-learning and virtual reality (VR) can be combined by comparing symmetric systems, where both peers use VR, and asymmetric systems, where only one peer uses VR and the other uses a tablet. They found that the symmetric system significantly enhanced presence, immersion, and reduced cognitive load, which are important for learning. However, both systems resulted in similar learning outcomes, demonstrating that both symmetric and asymmetric setups are effective for co-located VR pair-learning.

In industrial and technical settings, operating machinery often requires assistance or training that can be difficult to acquire with traditional documentation or voice/video calls alone. These methods often fail to convey spatial relationships, leading to miscommunication and repeated explanations. To address these challenges, we start by 3D scanning the machines and environments ahead of time, to have them available when running the application. The technician on site is assisted by a remote expert,

with the option for additional observers, using multi-device support. Depending on available hardware, participants join the session through their respective devices (PC, XR headset), with the software adapting to features like webcams and tracking.

In this setup (Figure 35), the expert views a virtual representation of the object or environment and can track the technician's pose to better understand what they are looking at. The expert can place and manipulate 3D annotations in the scene and provide additional guidance via voice and video. The technician sees the actual scene through a webcam or XR headset with the expert's annotations superimposed, matching the 3D position. This setup enables efficient collaboration between the expert and technician to solve complex problems more effectively.

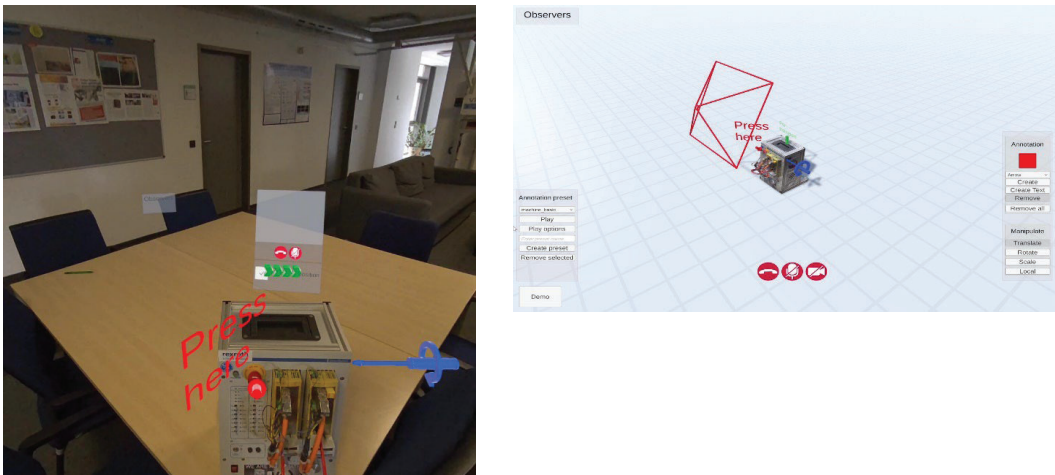


Figure 35. An overview of our setup. Left image: The technician view annotated with moveable augmentations. Right image: The remote expert view with the view of the technician visualized as frame and an interface capable of creating 3D annotations and add them to the scene as augmentation.

While existing systems such as HoloLens and mixed reality platforms focus on immersive experiences, our approach is tailored to the industrial environment. It addresses the current issue of heterogeneous hardware availability and usage, allowing flexibility through multi-device integration and adapting to the hardware on hand.

Technical description

Our application is based on Rainbow, a cloud-based communication platform developed by Alcatel-Lucent Enterprise, which offers APIs for video/audio sharing and data transfer. Rainbow's base communication layer operates independently of the device, which means interoperability across different platforms. It supports short-lived sessions, managing dynamic drop-in/out participation. We developed our application with Unity by utilizing the Rainbow SDK. This application supports common communication features such as video/audio transfer and an experimental low-level data channel for message transfer in a call between clients. It also employs distributed state storage, enabling participants to rejoin ongoing conversations, regardless of their device type, maintaining session continuity.

Our implementation currently runs on desktop and XR headsets. The desktop version employs mouse and keyboard as interaction means and uses the object tracker developed by Rambach et al.

(2017) for providing the camera pose. The 3D data available from the 3D scan allowed us to implement effects such as occlusion and shadows, which makes the experience more realistic.

The headset version is based on OpenXR which makes it simpler to adapt and port to different headsets. The interaction is based on hand/controller tracking which allows the user to select the most appropriate input. As the pose tracking is embedded in most headsets, the technician places the virtual object on top of the real one at the start of the session. After the placement has been done, the virtual object is hidden, but still used to provide occlusion.

Limitations

Currently OpenXR is not in a mature or stable state, which consequently introduces several bugs in official packages and vendor plugins. In terms of XR devices, they also come with several limitations. These include not being able to access sensor data, such as RGB and depth, or missing services such as surface reconstruction. While the API is the same across devices, different headsets might have small discrepancies, which makes it difficult to use the same code (e.g. Passthrough projection doesn't align correctly with rendering).

Summary and Future work

In this paper we presented an XR telecommunication application prototype targeting industrial use cases, with the potential of generalization. Our prototype operates on desktop PCs and multiple XR headsets, allowing participation and interaction in 3D. For the future, we will extend our application to be capable of quickly scanning the environment and add it to the scene. Moreover, we will extend the roles, allowing multiple technicians to work simultaneously. Additionally, we will test external RGB-based interaction methods for users of desktop PCs, enabling them to have hand and gesture interactions. To make augmentations more immersive, we will use depth-cameras, either external or provided by the headset vendors themselves.

Acknowledgements

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

References

- Chen H, Lee AS, Swift M, & Tang JC. (2025). 3D Collaboration Method over HoloLensTM and SkypeTM End Points. Proceedings of the 3rd International Workshop on Immersive Media Experiences, New York, NY, USA: Association for Computing Machinery; p. 27–30. <https://doi.org/10.1145/2814347.2814350>.
- Drey T, Albus P, Der Kinderen S, Milo M, Segschneider T, Chanzab L, ..., & Rukzio, E. (2022). Towards Collaborative Learning in Virtual Reality: A Comparison of Co-Located Symmetric and Asymmetric Pair-Learning. CHI Conference on Human Factors in Computing Systems, New Orleans LA USA: ACM; p. 1–19. <https://doi.org/10.1145/3491102.3517641>.
- Mahmood T, Fulmer W, Mungoli N, Huang J, & Lu A. (2019) Improving Information Sharing and Collaborative Analysis for Remote GeoSpatial Visualization Using Mixed Reality. 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Beijing, China: IEEE; p. 236–47. <https://doi.org/10.1109/ISMAR.2019.00021>.
- Rambach J, Pagani A, & Stricker D. (2017). Augmented Things: Enhancing AR Applications leveraging the Internet of Things and Universal 3D Object Tracking. Poster presentation at IEEE Int. Symp. Mixed and Augmented Reality (ISMAR-Adjunct), 2017, p. 103–8. <https://doi.org/10/ggr6rz>.

Bridging Realities: A Gap Analysis of XR Integration in Education and Training for Sustainable Future Skills

Electra Tsaknaki¹, Anders Östman², Konstantinos Smagas³,
Marco Sacco^{1,4}, and Patrick Bourdot^{1,5}

¹EuroXR, ²Novogit, ³Geoimaging, ⁴CNR-STIIMA, ⁵CNRS & University Paris-Saclay

Corresponding author: ilektratsaknaki@gmail.com

Keywords: eXtended Reality, State of the Art, Green Skills, Digital Skills, Gaps, Needs, Analysis, In-depth interviews, Questionnaire, XR Adoption

Introduction

The emergence of XR technologies signifies a great advancement in facilitating profound and immersive human-digital interactions. This progress is chiefly attributed to recent strides in high-speed communication and computation. For the partners of the Alliance4XR (Erasmus+) project, this is enabling unprecedented possibilities in the domains of education and training. Notably, the industry's transition towards open-source solutions, exemplified by platforms such as Unreal and Unity for XR creators, has ushered in a new era. This shift towards open-source game engines has substantially democratized the process of content creation, making the development of XR content more attainable and inclusive than ever before. As we embark on the incorporation of XR into education and training methodologies, the profound implications of these technological advancements and open-source solutions are duly acknowledged, shaping a future characterized by collaborative and accessible immersive learning experiences.

The main target of the Alliance4XR project is to design, develop, and validate an innovative teaching and training approach, using XR technologies. A careful design requires a reflection on the status of the Higher Education Institutions (HEI), Vocational Education and Training (VET), and key engineering sectors (structural engineering and construction, grids and energy, and maritime professions).

One of the first task of this project has been to perform a gap analysis, more precisely to describe the market needs and emerging professions, focusing on XR, and at the same time enhance the green skills (Tsaknaki et al., 2024). This presentation gives an overview of this work, from its methodology up to its main results.

Methodology

Use cases collection. The process of specifying successful XR use cases is a crucial step in understanding the current landscape of XR applications across various sectors. Desk research, a method

of collecting and analysing existing data from various sources, provides a structured approach to identify and evaluate exemplary use cases. The next step was the data collection process. This process involves collecting information from various reliable sources, which include, but are not limited to: a) Academic Journals and Conference Proceedings; b) Industry Reports; c) Company Websites; d) Technology Review Platforms; e) News Articles and Press Releases. To systematically compile the use cases, their information was gathered in a predefined template. The last step was to verify the credibility and pertinence of the obtained use cases, by cross-checking information among various resources and getting advice from experts whenever possible. This phase also entails an early-stage analysis to assess overall themes and common features in the use cases. The same procedure also applied to the collection of the learning material for the sectors.

In-depth interviews with industry and academia stakeholders. In order to gain insights from industry and academic experts regarding the state of the art and needs of XR technologies, in-depth interviews were conducted. At first, finding the candidates for the interviews was a crucial step of the process. This was achieved by addressing the consortium's extended network and colleagues. Based on the background of the candidates, the focal point was to have a fair representation between industry and academia. Thus, eleven in-depth interviews with lectures/academics, as well as industry CEOs and technology officers were conducted. At the next step, the questions needed to guide the conversation were defined. During the interviews, follow-up questions were asked in order to gain insight into the unique way of thinking of the experts.

Questionnaire: eXtended Reality state and needs for Education & Training. In order to reach out also to organisations not covered by the desk study, a questionnaire was distributed to the partner's networks in order to capture the current state of XR technologies in industry and academia, and the needs of the sector. The questionnaire was distributed through LimeSurvey, to ensure that it was user-friendly and accessible. As a first step, the target audience for the survey was identified. The target audience included stakeholders in the sectors targeted and academic experts. The next step was the design of the questionnaire and ethical approval. As a next step, the questionnaire was distributed to various distribution channels, which include, but are not limited to: a) Email: Personalized emails to identified stakeholders with a link to the questionnaire. b) Professional Networks: Distribution through professional networks such as LinkedIn to reach a broader audience. c) Industry Associations: Collaboration with industry associations to distribute the questionnaire to their members. d) Social Media: Promotion of the questionnaire on the project's social media platforms to increase reach and participation.

Results

For the first aspect of the gap analysis, 57 use cases were collected, and the distribution between the sectors is the following:

Table 6. Use Cases distribution.

Sector/Category	Number of use cases
Higher Education Institutions	7
Vocational Education and Training	7
Structural Engineering and Construction	5
Maritime Professions	11
Grids and Energy	8
Remote Collaboration	6
AR/VR/MR	13

After careful analysis, the major gaps and needs for the HEI and VET can be summarized as follows:

Gaps

1. The integration of XR technologies into the existing curriculum while aligning with traditional teaching methods and standards could pose a challenge.
2. Bringing XR tools to thousands of programs and institutions requires a big investment in hardware, software, and faculty training, which can be very resource-intensive.
3. A critical gap is that faculty and students possess limited knowledge and experience with XR technologies to be able to use them in the process of teaching and learning.
4. The initial costs for hardware, as well as for software and content development are high, which some of the institutions will find hard to fund.
5. Ensuring XR systems are functioning effectively and continuously requires ongoing technical support and system maintenance throughout its lifespan.
6. Standardized assessment methods to measure the effectiveness of XR-enhanced learning experiences and their impact on educational outcomes ought to be developed.

Requested industry developments

1. Develop strategies for the seamless addition of XR technologies to traditional curricula in a manner that is in alignment with academic standards and then builds on traditional teaching methods.
2. Develop a scalable plan for deploying XR tools into multiple programs that lean on efficient resource usage and cost-effective measures.
3. Provide faculty and students with training programs in XR technologies and learning use cases.
4. Identify low-cost solutions and cost-sharing and funding opportunities to leverage initial investment and long-term maintenance of XR technologies.
5. Provide a robust technical support structure to help manage and maintain our suite of XR tools.
6. Create evaluation frameworks for measuring the effect that XR-enhanced learning has on educational outcomes, establishing universal standards for evaluation.

First, the respondents of the questionnaire and the in-depth interviews seem to be the representatives of the target group. Experiences and examples call for direct engagement with XR technologies, making their feedback pertinent and comprehensive.

Moreover, various sector representatives have realized the potential benefits that XR technologies can provide. VR training for pilots and AR assisting assembly processes show that XR can improve learning outcomes, decrease error rates, and be a more engaging way to learn while just adding up efficiencies. These various successful use cases demonstrate the capability of XR to enhance operational efficiencies and training outcomes.

As the technology matures and its accessibility increases, we will probably see wider adoption in more industries. Macrotrends, some being AI integration for better content creation and experiences, rising accessibility with cheaper yet intuitive systems, more scalability thanks to infrastructure

refinements, along with advanced training and simulation in support of real immersive environments, as well as remote collaboration work tools, are something that would shape AR/VR usage ahead.

Based on the outcome of the survey and the in-depth interviews, the main barriers to implementing XR technologies include high initial costs for hardware, software, and content development, technical complexity in integrating XR with existing systems and ensuring compatibility, user resistance and lack of expertise, the need for high-quality and engaging content, infrastructure requirements such as robust internet and processing power, health and safety concerns like eye strain and motion sickness, and data privacy and security issues in handling sensitive information.

Future directions and recommendations

The rise of XR technologies in the future will present transformative opportunities for various sectors. Some of the key trends are improved hardware performance, increased accessibility, advancements in sensing and tracking technologies, and seamless integration of multimodal interactions. These technological advancements will enable more immersive and intuitive XR experiences, driving broader adoption (Boel et al., 2023).

As XR technology becomes more pervasive, it is crucial to prioritize user-centric design and address ethical considerations. Providing inclusive and accessible XR experiences, resolving privacy issues, and responsible development and deployment of XR technology are needed to limit negative impacts on society (Alnagrat et al., 2022). To capitalize on the future potential of XR, various stakeholders, including educators, industry leaders, policymakers, and technology providers, should collaborate to drive innovation, enhance user experiences, and unlock new opportunities across diverse sectors.

References

- Alnagrat, A., Ismail, R. C., Idrus, S. Z. S., & Alfaqi, R. M. A. (2022). A review of extended reality (XR) technologies in the future of human education: Current trend and future opportunity. *Journal of Human Centered Technology*, 1(2), 81-96. <https://doi.org/10.11113/humentech.v1n2.27>
- Boel, C., Dekeyser, K., Depaepe, F., Quintero, L., Daele, T., & Wiederhold, B. (2023). European Commission: Directorate-General for Communications Networks, Content and Technology. *Extended reality: opportunities, success stories and challenges (health, education): final report*, Publications Office of the European Union. <https://data.europa.eu/doi/10.2759/121671>
- Tsaknaki, E., Bourdot, P., Sacco, M., & Gonzalez, R. (2024, July) XR – state of play, market needs, training trends (public report) - Deliverable D1.1 - Project: 101140252 — Alliance4XR — ERASMUS-EDU-2023-PI-ALL-INNO. https://alliance4xr.eu/wp-content/uploads/2024/09/Alliance4XR_D11.pdf

XR technologies in the engineering sector – A gap analysis

Anders Östman¹, Electra Tsaknaki², Patrick Bourdot^{2,3}, Son Dang⁴,
Aikaterini Stamou⁵, and Vasilis Petrou⁶

¹*Novogit, Bagarmossen, Sweden*

²*EuroXR, Brussels, Belgium*

³*CNRS & University Paris-Saclay, France*

⁴*University of Minho, Braga, Portugal*

⁵*Aristotle University of Thessaloniki, Thessaloniki, Greece*

⁶*Geoimage, Nicosia, Greece*

Corresponding author: Anders.Ostman@novogit.se

Keywords: eXtended Reality, Education, Gap Analysis, Supply And Demand, Skills And Knowledge

Abstract

Alliance4XR is an Erasmus+ project co-funded by the European Union that aims to create and test learning methodologies and materials to empower education institutes, and to address XR digital skills in the engineering sector successfully. The project primarily targets the needs of sectors like structural engineering and construction, grids and energy, and maritime professions. The main objectives of this paper are to describe the current supply of XR-related courses from education institutes, primarily in the target engineering sectors, and to point out the gap between the supply and demand of XR skills and knowledge. A more detailed analysis of institutions' available and missing skills, curricula and members' needs has recently been presented by Östman et.al (2014).

The potential of using XR technologies in education has been highlighted by several authors, for instance Fernandez-Ceredo et.al (2024) and PwC (2020). These studies shows a clear potential, both in terms of usage as well as economic impact. The demands of education and training in XR technologies, having a focus on the target sectors, have been investigated by Tsaknaki et al., 2024. This study contains, among other things, an analysis of existing use cases where XR technologies are applied in operational practices. To complement that study, this paper instead deals with the supply side. The supply was investigated by a desk study of course syllabi and a questionnaire submitted to teachers in XR technologies. However, to specify the gap between supply and demand related to education, a common model of knowledge, skills, and competencies in XR technologies is needed.

The European Skills, Competences, Qualifications and Occupations (ESCO) project is a European Commission project run by DG Employment. It is the European multilingual classification of Skills, Competences and Occupations (ESCO, 2024). According to their website, "ESCO works as a dictionary, describing, identifying and classifying professional occupations and skills relevant for the EU labour

market and education and training. Those concepts and the relationships between them can be understood by electronic systems, which allows different online platforms to use ESCO for services like matching jobseekers to jobs based on on the basis of their skills, suggesting trainings to people who want to reskill or upskill etc”.

The ESCO dictionary is structured around concepts (knowledge, skills and competences) on one side, and occupations on the other side. The occupation pillar is structured hierarchically and mapped to the International Standard Classification of Occupations (ISCO). The concept pillar is also hierarchical, and it is linked to various occupations.

The ESCO dictionary does not (yet) recognise XR as a concept. However, both Virtual Reality and Augmented Reality are recognised and defined. As an example, the concept of Virtual Reality is, according to ESCO, an essential knowledge skill for environmental technicians and an optional skill for several additional occupations such as real estate agents etc. In a similar way, the concept of Augmented Reality is an essential knowledge skill for environmental engineers, 3D modelers and 3D animators. The close connections between VR and AR technologies and green occupations such as environmental technicians and environmental engineers are clearly specified in the ESCO dictionary. This aligns well with the focus of the Alliance4XR project and its focus on green skills.

One limitation of the ESCO dictionary is that the concepts are organized hierarchically into broader concepts and sub-concepts. A more flexible approach is provided by the EO4GEO project, which has developed an ontology for the Earth Observation and Geospatial domains (EO4GEO, 2022). Since this Body of Knowledge (BoK) is based on semantic web technologies, an extended set of relationships between knowledge concepts has been defined, for instance “is a kind of” and “is a property of”.

To compare the supply and the demand of education and training in the XR domain, a draft XR taxonomy has been created, where the most important XR topics are specified. This draft taxonomy has a hierarchical structure and is implementable in the ESCO dictionary. Based on previous experiences, the following draft XR taxonomy was elaborated, see Table 1.

Table 7 Concepts and terms of the draft XR taxonomy

1 st level topics	2 nd level topics	3 rd level topics
1. Fundamentals of XR	1.1 XR definitions	1.1.1 Virtual Reality (VR)
		1.1.2 MR, AR and AV
		1.1.3 XR
	1.2 Concepts and founding technologies	1.2.1 Perceptive and cognitive concepts
		1.2.2 Simulation modelling
		1.2.3 Computer science technologies
1.3 XR equipment		
2 Tracking technologies		
3. 3D visualization		
4. 3D Audio		
5. Kinematics		
6. Haptic senses		
7. Mixed reality	7.1 MR equipment	
	7.2 MR calibration	
	7.3 Image alignment	
8. Multimodal interactions in immersive environments		
9. Virtual navigations techniques		
	10.1 User centered design	

1 st level topics	2 nd level topics	3 rd level topics
10. Human factors and ergonomics in XR	10.2 Dimensions in XR applications	
	10.3 Criteria on XR application ergonomics	
	10.4 Ergonomic evaluation	
	10.5 Evaluation based on questionnaires	
	10.6 Evaluation of usability	
	10.7 User-centric evaluation methods	
11. XR for business development		
12. XR in process and product development		
13. XR usage		

The required XR skills were then classified using this draft XR taxonomy and the use cases as identified by Tsaknaki et. al., 2024. This classification of required skills constituted the core demand part of this gap analysis. To model the supply side, a desk study of course syllabi was carried out together with a survey among tutors based on a LimeSurvey questionnaire. In this questionnaire, tutors were also asked to give their opinions on future investments and developments in XR technologies at their institutions.

Despite these limitations, the results of the supply and demand analysis indicate the following topics being of high priority for further development (in order)

- **Basic Principles of XR:** Respondents overwhelmingly agree on the necessity of understanding the foundational concepts of XR. This indicates a need for comprehensive introductory courses that cover the basics of Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR).
- **Advanced Technologies for Mixed Reality:** This topic reflects the growing interest in cutting-edge XR technologies. Curriculum development may include modules on the latest advancements in MR technologies, such as holography and advanced display systems.
- **Usage of XR technologies in various industrial sectors:** There is a clear need for additional training tools, such as simulators and responses to various scenarios. More efficient development tools are also needed to reduce the costs associated with developing applications and products for XR usage.
- **Virtual Navigation Techniques:** Effective navigation is key to creating immersive and intuitive XR experiences. Courses and learning material should cover the principles of spatial awareness, user orientation, and wayfinding in virtual environments. Hands-on projects involving navigation challenges will build practical skills.
- **Tracking Technologies:** The strong support for this topic suggests that understanding how XR systems track user movements and interactions is crucial. It may be beneficial to include practical workshops or labs that allow students to work with various tracking technologies.
- **3D Visualization and 3D Audio:** The emphasis on these areas indicates that immersive experiences are essential for XR applications. Educational programs should incorporate projects that involve creating and manipulating 3D models and integrating spatial audio to enhance realism.

- Applying XR technologies in process and product development: Such emphasis suggests that XR can streamline workflows and enhance productivity. Training programs should include simulations and project-based learning that allow students to apply XR technologies to real-world process challenges and product developments.
- The role of XR for business development: The responses to the questionnaires indicate a strong belief in the role of XR technologies for business development in various industrial sectors. This suggests a need for case studies and real-world applications to be included in the curriculum, focusing on how XR can improve processes, training, and customer engagement in various industries and provide value to the organization.

A detailed analysis of the use cases shows that the main usage of XR technologies within the target sectors (construction, energy and maritime) are addressing the basic principles of XR and XR usage, for instance real-world simulations for practical skills.

The results of the supply study (desk study of course syllabi and questionnaire for tutors) were both mapped against the draft XR taxonomy, providing the possibility to compare supply and demand. However, the response rate of the questionnaire, along with its geographical and topical coverage, was not comprehensive enough to justify quantifications. In addition, since the result is only a snapshot of the existing situation, its validity period is short.

According to the ESCO dictionary, topics such as VR and AR are essential knowledge skills for green occupations such as environmental technicians and environmental engineers. Despite this, the supply studies found quite limited content of XR technologies in the courses provided by educational institutes addressing these professions. Most XR-related courses are currently provided within computer science and education study programs.

The following general conclusions can also be made

- The supply studies found quite limited content of XR technologies in the courses provided to learners in the target sectors. Most XR-related courses are currently provided within computer science and education study programs. This even though there are clear industrial needs for skills in basic principles of XR as well as XR usage. Future work aims to address these issues.
- The findings highlight the growing importance of incorporating XR technologies into educational programs to meet the evolving needs of the industry in topics such as "Basic Principles of XR", "Advanced Technologies for Mixed Reality", "Usage of XR Technologies", "3D Visualization" and "Virtual Navigation Techniques".
- Within the Vocation and Education Training (VET) sector, there is a need for low-cost solutions aiming to guide the learner in performing operational tasks and making real-world decisions. Simple simulators and gaming applications may be one way forward.
- As the XR industry continues to expand, the importance of integrating practical applications and emerging trends into Higher Education Institute (HEI) programs is likely to grow.

References

- EO4GEO (2022). A Body of Knowledge (BoK) for EO/GI. <http://www.eo4geo.eu/bok/>
- ESCO (2024). What is ESCO? <https://esco.ec.europa.eu/en/about-esco/what-esco>
- Fernandez-Ceredo, J., Fernandez-Batanero, J.M., & Montenegro-Rueda, M. (2024). Possibilities of Extended Reality in education. Interactive Learning Environments. <https://doi.org/10.1080/10494820.2024.2342996>.
- PwC (2020). Virtual and augmented reality could deliver a £1.4 trillion boost to the global economy by 2030. <https://www.pwc.com/id/en/media-centre/press-release/2020/english/virtual-and-augmented-reality-could-deliver-a-p1-4trillion-boost.html>
- Tsaknaki, E., Östman, A., Smagas, K., Sacco, M., & Bourdot, P. (2024). Bridging Realities: A Gap Analysis of XR Integration in Education and Training for Sustainable Future Skills. Abstract submitted to the 21st EuroXR International Conference, EuroXR 2024, Athens Nov 27-29, 2024.
- Östman, A., Stamou, A., Tsaknaki, E., Petrou V., Dang, S., & Bourdot, P. (2024). Analysis of institutions' available and missing skills, curricula and members' needs (restricted report) - Deliverable D1.2 - Project: 101140252 - Alliance4XR - ERASMUS-EDU-2023-PI-ALL-INNO, <https://alliance4xr.eu/wp-content/uploads/2024/09/Alliance4XR-D1.2.pdf>

A Mixed Reality Eye-Tracking Study on Visual Influences in Dietary Choices

Nikolaos D. Tantaroudas, Konstantinos Koukoudis, and Tina Katika

Institute of Communication and Computer Systems, Zografou, Greece

Corresponding author: nikolaos.tantaroudas@iccs.gr

Abstract

This study explores a new method for studying food consumption behaviors using eye-tracking in a Mixed Reality (MR) environment. Participants wear video see-through headsets and interact with various food scenarios while their gaze patterns are tracked. The study combines these metrics with questionnaires and immersive visuals to analyze eating habits. Expected results aim to show that carnivorous participants focus more on meat, vegetarians engage more with sustainable food choices, and omnivores respond to a broader range of factors. The findings are expected to highlight the potential of MR eye-tracking to reveal hidden preferences and inform tailored educational and marketing strategies based on dietary habits.

Introduction

Traditional methods such as questionnaires and food diaries, while frequently utilized, often face issues with biases and are inadequate in capturing the complex motivations behind food choices (Sember et al., 2020; Svendsen et al., 2006). Recent advances in Extended Reality (XR) technologies, including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), have revolutionized the study of food choices by enabling controlled experiments in immersive and realistic virtual settings (Manoni et al., 2016). These technologies, particularly when combined with eye-tracking metrics, offer more objective insights into cognitive and emotional responses to food stimuli (Ruppenthal, 2023). However, comprehensive research employing MR to explore sustainable practices remains limited. This study aims to fill these gaps by integrating traditional questionnaires with MR eye-tracking, offering a more objective and immersive evaluation of food preferences and sustainable choices. The adaptable nature of MR allows for rapid shifts between scenarios, delivering deeper, more detailed insights into dietary behaviors, and promotes a better understanding of demographic trends and sustainable food practices.

Methodology

This study employs eye-tracking in Mixed Reality (MR) to evaluate participant engagement by analyzing essential metrics such as gaze direction, duration, and fixation frequency. By refining gaze duration thresholds, the investigation differentiates between quick glances and more meaningful visual engagement. Participants interact with virtual food items in an MR environment, using XR hardware and

the OpenXR framework. Their gaze behaviors are tracked, and data on gaze points, duration, and fixation locations are visualized through heatmaps. Additionally, quaternion rotation data of gaze direction is converted into Euler coordinates, allowing for a more precise analysis of gaze direction.

Experimental Setup

The flowchart in Figure 36 outlines the study's approach to exploring the link between visual attention and food decision-making. Participants' preferences are evaluated in an MR setting, where a variety of food options are presented on a table with immersive 3D visuals. This setup is specifically designed for eye-tracking research, capturing detailed interactions with different food items. Eye-tracking data is gathered using Unity and the Oculus OVR Eye Gaze component, tracking participants' gaze as they interact with virtual food items. The system records gaze fixations and processes metrics like direction, duration, and fixation points. These metrics are then analyzed alongside questionnaire data to better understand food choices. Results are visualized using 3D heatmaps that highlight areas of high visual interest and identify food items that capture attention. By combining quantitative eye-tracking data with qualitative feedback, the study provides a thorough understanding of food habits and preferences.

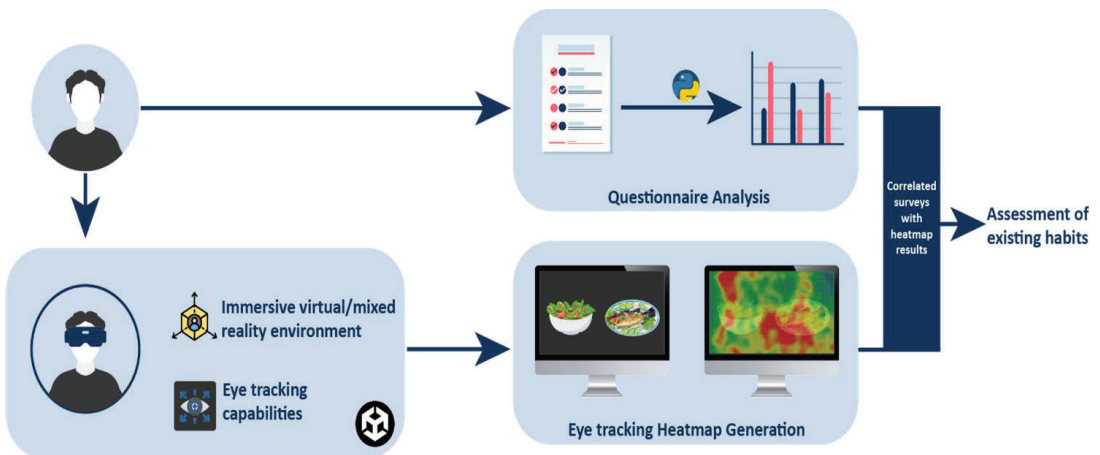


Figure 36. Extended Reality-Based Behavioral Analysis Flowchart: Identifying Dietary Choices through Eye-Tracking, Heatmaps and Data Analysis.

Assessment of Participants

The study surveyed 114 participants, uncovering a range of food preferences and significant differences across gender, BMI, and health concerns between carnivores, vegetarians, and vegans. The carnivorous group had a higher proportion of males and elevated obesity rates, reflecting gender and cultural impacts as well as the potential advantages of plant-based diets. Table 8 shows a higher proportion of males in the carnivorous group compared to the vegetarian and vegan groups, suggesting gender and cultural influences on dietary choices.

Table 8. Comparison of Dietary Groups Across Various Categories.

Category	Vegetarians (incl. Vegans)	Omnivorous	Carnivorous
----------	----------------------------	------------	-------------

Age groups			
18-24	28.57%	6.19%	0%
25-34	14.29%	23.71%	14.29%
35-44	28.57%	56.70%	57.14%
45-54	0%	7.22%	0%
55-64	0%	3.09%	14.29%
65+	28.57%	3.09%	14.29%
Gender Distribution			
Male	28.57%	37.11%	71.43%
Female	71.43%	60.82%	28.57%
Non-binary	0%	2.06%	0%
Prefer not to say	0%	2.08%	0%
BMI (Body Mass Index)			
Underweight	0%	2.06%	0%
Normal Weight	57.14%	63.92%	28.57%
Overweight	42.86%	47.42%	42.86%
Obese	0%	6.18%	0%
Environmental Impact Concerns			
Very Low	0%	13.40%	14.29%
Low	0%	17.53%	28.57%
Moderate	42.86%	47.42%	42.86%
High	42.86%	19.59%	14.29%
Very High	14.29%	2.06%	0%
Health Concerns			
Very Low	0%	4.12%	0%
Low	0%	7.22%	42.86%
Moderate	42.86%	47.42%	42.86%
High	28.57%	42.27%	14.29%
Very High	14.29%	8.25%	0%
Ethical Concerns			
Yes	71.43%	27.08%	0%
No	28.57%	72.92%	100%

In addition, as illustrated in Figure 37, taste was the most important factor influencing food choices, followed by health, cost, convenience, and freshness.

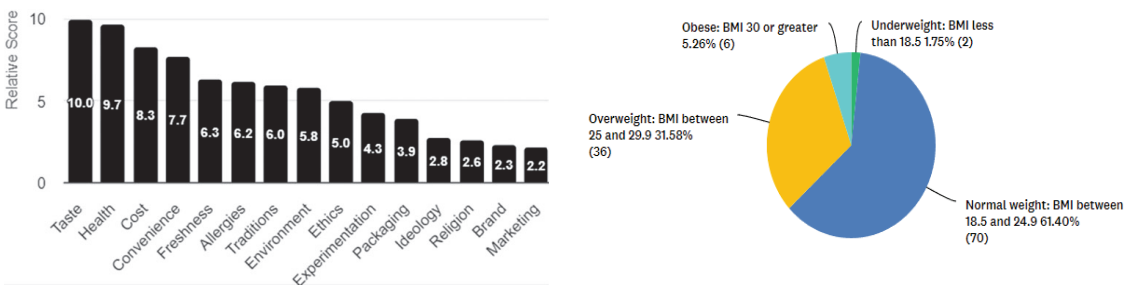


Figure 37. a) Key Factors affecting food habits, and b) BMI of participants.

Mixed Reality Demonstration

Immersive Mixed Reality (MR) content was created to examine the relationship between participants' eye movements and the factors influencing their food choices, incorporating data storytelling to boost engagement (as shown in Figure 38). Nutritional values for the food scenarios were calculated using the NRF9.3 method (Sluik et al., 2015), while estimates for CO₂ emissions, water, and land use were sourced from existing research (Poore & Nemecek, 2018). Participant interactions, including eye tracking, clicks, and gaze duration, were carefully monitored within the MR environment.



Figure 38. Mixed Reality Environment Setting with Data Storytelling Visuals.

Table 9. presents the engagement metrics, such as clicks and gaze duration, for carnivorous, vegetarian, and omnivorous participants across different food scenarios. Omnivorous participants showed higher levels of engagement, likely due to the variety and complexity of their diet, while carnivorous participants focused more on meat-based scenarios. This highlights the MR environment's effectiveness in engaging participants with content that matches their specific dietary preferences.

Table 9. Results of Number of Gazes, Total, Max, and Average Gaze Duration per Food Scenario at a sampling ratio of $f = 50$ Hz and minimum gaze fixation > 0.15 s

Group	Scenario	Count Total	Gaze (s)	Avg (s)	Max(s)
Vegetarian	Carbonara	2	0.42	0.21	0.26
	Lentil Soup	24	9.26	0.39	1.36
	Butter	10	3.72	0.37	0.88
	Beef Stake	13	8.06	0.62	1.44
	Bean Soup	5	1.62	0.32	0.52
	Buratta Salad	31	17.06	0.55	3.48
Omnivorous	Carbonara	39	26.54	0.68	4.22
	Lentil Soup	17	21.10	1.24	5.28
	Butter	14	8.76	0.63	2.14
	Beef Stake	21	12.68	0.60	1.78
	Bean Soup	17	8.76	0.52	2.24
	Buratta Salad	22	17.94	0.82	3.46
Carnivorous	Carbonara	29	14.16	0.48	1.80
	Lentil Soup	21	11.24	0.54	2.42
	Butter	24	7.60	0.32	1.02
	Beef Stake	41	19.28	0.47	1.76
	Bean Soup	25	12.24	0.49	1.20
	Buratta Salad	38	17.66	0.46	1.76

Heatmaps produced within the MR environment (Figure 39) for an omnivorous participant highlighted areas of intense visual focus, providing insights into how individuals visually engage with specific food scenarios.

Furthermore, the eye-tracking analysis showed that carnivorous participants tended to engage more with less healthy and sustainable food options but also demonstrated significant interest in health, GHG, and water usage metrics as shown in Table 10. This suggests potential opportunities to educate these groups about the environmental consequences of their dietary choices.

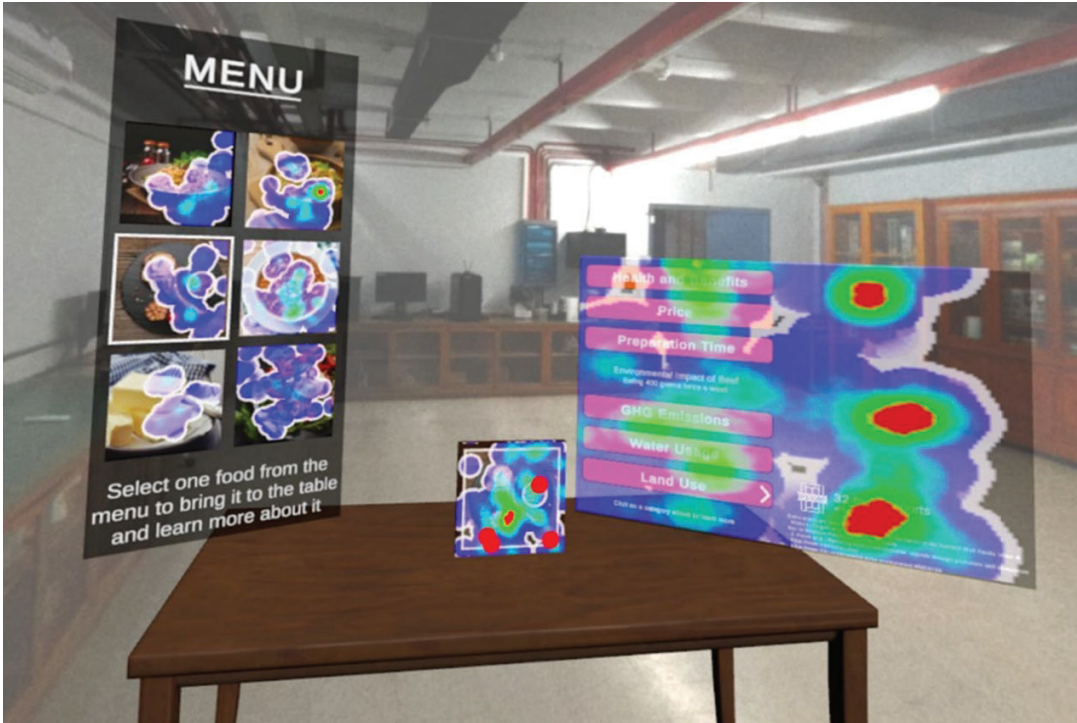


Figure 39. Visual attention heatmaps generated within the MR environment for omnivorous participant

Table 10. Number of User clicks, Number of Gazes, Total, Average and Max Gaze duration on several key metrics for all food scenarios.

Group	Factor	Clicks	Count Total	Gaze (s)	Avg (s)	Max(s)
Carnivorous	Health	20	30	9.70	0.32	0.68
	Water Use	19	69	26.24	0.38	1.58
	Price	19	65	28.58	0.44	1.16
	Land	12	60	20.78	0.34	1.28
	GHG	20	25	9.46	0.38	1.34
	Convenience	12	47	20.44	0.44	4.58
Vegetarian	Health	9	7	1.64	0.23	0.36
	Water Use	4	14	3.60	0.26	0.52
	Price	4	7	2.24	0.32	0.80
	Land	3	18	4.46	0.25	0.38
	GHG	5	17	3.88	0.23	0.36
	Convenience	5	25	8.28	0.33	1.00
Omnivorous	Health	11	69	41.08	0.59	2.88
	Water Use	10	66	35.60	0.54	3.64
	Price	13	69	38.72	0.56	3.46
	Land	10	59	33.86	0.57	2.80
	GHG	10	64	34.16	0.53	3.58
	Convenience	12	29	10.08	0.35	0.68

Conclusions and Future Studies

As expected, the MR environment assessment showed that participants with sustainable dietary habits spent more time interacting with sustainable food options, while those with carnivorous diets focused more on meat-based scenarios. The true potential of this tool lies in its ability to uncover deeper behavioral patterns. Future research will refine the tool by integrating additional metrics, such as selection behaviors and physiological responses (e.g., pupil dilation and heart rate), to further enhance its capability. The results will be expanded by developing new MR scenarios covering a wider array of consumer behaviors, including themes like circular economy and climate resilience. This approach will collect valuable data on how specific campaigns can successfully promote sustainable practices, ultimately contributing to societal acceptance and the success of sustainability-promoting policies.

References

- Sember, V., Meh, K., Sorić, M., Starc, G., Rocha, P., & Jurak, G. (2020). Validity and Reliability of International Physical Activity Questionnaires for Adults across EU Countries: Systematic Review and Meta Analysis. *Int. J. Environ. Res. Publ. Health*, 17. <https://doi.org/10.3390/>
- Svendsen M, & Tonstad S. (2006). Accuracy of food intake reporting in obese subjects with metabolic risk factors. *Br J Nutr*. 95(3):640-9. <https://doi.org/10.1079/bjn20051662>. PMID: 16512951
- Manzoni, G. M., Cesa, G. L., Bacchetta, M., Castelnuovo, G., Conti, S., Gaggioli, A., & Riva, G. (2016). Virtual Reality-Enhanced Cognitive-Behavioral Therapy for Morbid Obesity: A Randomized Controlled Study with 1 Year FollowUp. *Cyberpsychology, Behavior, and Social Networking*, 19(2), 134-140. <https://doi.org/10.1089/cyber.2015.0208>
- Ruppenthal, T. (2023). Eye-Tracking Studies on Sustainable Food Consumption: A Systematic Literature Review. *Sustainability*. <https://doi.org/10.3390/su152316434>
- Sluik, D., Streppel, M.T., van Lee, L., Geelen, A., & Feskens, E.J. (2015). Evaluation of a nutrient-rich food index score in the Netherlands. *J Nutr Sci*. 4:e14. <https://doi.org/10.1017/jns.2015.4>, PMID: 26097700; PMCID: PMC4462757.
- Poore, J., & Nemecek T. (2018). Reducing food's environmental impacts through producers and consumers. *Science*, Vol 360, Issue:6392, pp. 987-992, <https://doi.org/10.1126/science.aag0216>

XR-based Human-Robot Collaboration along a Conveyor Picking Line

Spyridon Nektarios Bolierakis, Kostas Naskou, Konstantinos Kokkalis,
Fotios Konstantinidis, Tina Katika, Georgios Tsimiklis, Ioannis Karaseitanidis,
and Angelos Amditis

Institute of Communication and Computer Systems, Athens, Greece

Corresponding author: kostas.naskou@iccs.gr

Keywords: Human-Robot Collaboration, Extended Reality, Collaborative Sorting Line, Head-Mounted Display.

Introduction

Efficient management of construction and demolition waste is nowadays mandatory but remains challenging. Manual sorting by humans poses health and safety threats related to repetitive handling of heavy and potentially hazardous materials. Full automation of the process, despite many successful lab tests, in practice under real-world conditions cannot reach the expected quality in terms of material separation. Human-robot collaboration (HRC) can improve sorting efficiency, measured by speed and accuracy, if human and robot capabilities are combined appropriately (Chen, 2023).

The Problem

In order to sort a stream of waste, a conveyor sorting line, equipped with a control unit, a sensing station at the beginning and a cobot at the end, will be set up. The sensing station detects and classifies waste items moving on the conveyor belt. A task planner decides on the action that needs to be taken by the cobot so that the items end up in the correct sorting bin (Konstantinidis, 2023). To achieve effective human-robot collaboration towards improving the efficiency of waste sorting, Mixed Reality technology can play a vital role. Firstly, to provide the necessary feedback mechanism between the conveyor-cobot sorting system and the humans aiding and supervising the process. Secondly, to make better decisions during the system design simulations by bringing the human in the loop. The aim of the developed prototype at this stage is twofold. On the one hand, to provide an interactive simulation, within MR settings, of the collaborative sorting process so that it can be optimised together with the user interface before being applied to the actual sorting site. On the other hand, to establish an effective mode of communication between the HMD and the cobot.

The Prototype

A prototype for HoloLens 2 was developed in Unity. The connection between the head-mounted display (HMD) and the control unit of the conveyor sorting line was established via an MQTT messaging protocol. The information transmitted via these messages include each object's location and moving speed along with measurements required to generate its 3D shape as it moves on the conveyor. The message also contains task descriptions for specific objects to be assigned to the cobot or the human operator.

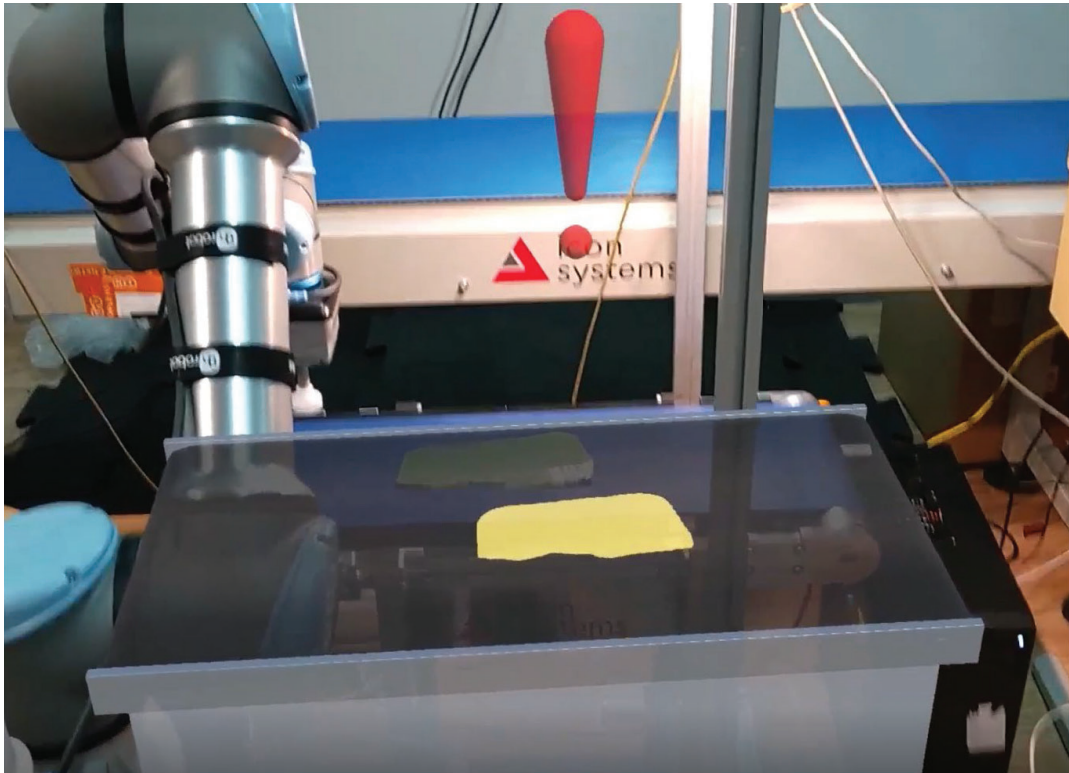


Figure 40. A short conveyor was used for development. It is shown along with a virtual conveyor at an offset so that they are both visible during demonstration trials. One of the actual conveyors is also visible at the background.

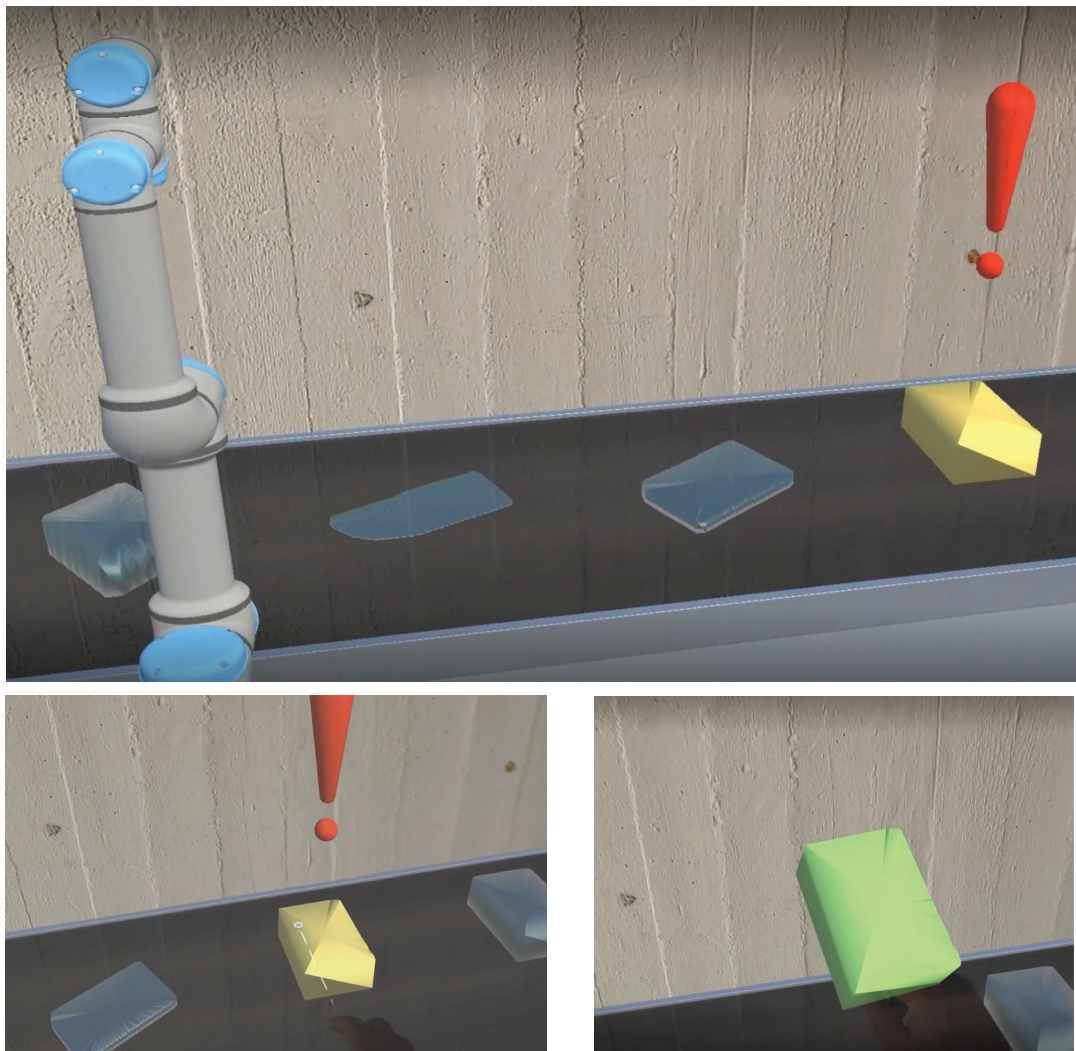


Figure 41. An item of interest approaches the workstation of a worker wearing a HMD (top). The worker picks up the object (bottom left). The virtual object responds to the interaction with the human operator (bottom right).

Based on this information, virtual objects are displayed on the HMD overlaid on the physical objects. As soon as an item, which requires handling by a human operator, passes through the detection area of the conveyor the virtual object is highlighted on the HMD accompanied by a label. The operator picks up the object and retrieves further information about it, such as which sorting bin it belongs to or if a further action is required, e.g. to remove any impurities before sorting it.

Discussion and Future Work

The most challenging aspect of establishing a clear communication channel based on HMD for human-robot collaboration is understanding the worker's interactions with the physical objects in order to update the virtual environment based on visual information without specialised tactile equipment

which may become a hinderance to the worker's interactions. This may required an additional vision sensor overlooking the conveyor area.

Mirroring the movements of the physical cobot by its virtual counterpart and shadowing its intended moves will be the next step in this implementation so that the worker is aware of the cobot's intent and its movements within their common workspace.

Acknowledgements

This research is part of the W2W project, which has received funding as a European Union's Horizon Europe Research and Innovation Action under Grant Agreement No. 101138789. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

References

- Chen, J., Fo, Y. & Lu, W. (2023). Augmented reality-enabled human-robot collaboration to balance construction waste sorting efficiency and occupational safety and health. *Journal of Environmental Management* 348, 119341. <https://doi.org/10.1016/j.jenvman.2023.119341>
- Konstantinidis, F. K., Sifnaios, S., Arvanitakis, G., Tsimiklis, G., Mouroutsos, S. G., Amditis, A., & Gasteratos, A. (2023). Multi-modal sorting in plastic and wood waste streams. *Resources, Conservation and Recycling*, 199, 107244. <https://doi.org/10.1016/j.resconrec.2023.107244>

The VESPA 2.0 System: a new frontier of Rehabilitation

Emanuele Maria Merlo¹, Angelo Belfiore², Gian Paolo Donnarumma²,
Marco Salvatore Carmelo Falciglia², Francesco Generali²,
and Salvatore Marco Pappalardo²

¹*Department of Biomedical, Dental Science and Morphological and Functional Images,
University of Messina, Messina, Italy*

²*Software Engineering Italia S.r.l., Piazza A. Lincoln 7, Catania, Italy*

Corresponding author: emerlo@unime.it

Keywords: VESPA 2.0 Project; Virtual Reality; Virtual Reality Applications; Cognitive Rehabilitation; Diagnosis.

Application

Topic of interest: Medicine and Rehabilitation

VESPA 2.0 (Virtual Environment for a Superior neuro-PsychiAtry, 2nd generation⁵ is a project founded by the PO ERDF programme 2014-2020 (extended to 2024 due to the pandemic of Covid-19) of the Sicilian Region (Italy) to foster Research and Innovation in Enterprises. The project focuses on application of Virtual Reality (VR) to diagnosis and treatment of neurodevelopmental and acquired neurocognitive disorders in adults and minors. Based on the state of the art (Andronico et al., 2009; Cammisuli et al., 2022; Latella et al., 2024; Merlo et al., 2023; Pappalardo et al., 2020; van den Bergh et al., 2021; Vicario & Martino, 2022), recent studies confirmed the effectiveness and efficiency of VR systems in the biomedical field. These results predict that the use of VR technologies for clinical practice can be compared to modern diagnostic and rehabilitation techniques, representing a response to the growing need.

In this sense, the use of these technologies represents an added value, not only provided in the light of the population needing clinical attention, but producing benefits (Merlo et al., 2023). reflecting, e.g., the reduction in time taken to apply subjects to experimental and clinical conditions, to reach a diagnosis, and the rehabilitation facilities of a temporal and methodological nature, as well as the implementation of end-user services. VR applications, whether 3D or 5D, allow the VESPA 2.0 project to be at the top of the innovation by promoting the implementation of new clinical methodologies in the

⁵ <https://www.progettovespa.it>

national and international scene, through the provisioning of a set of innovative diagnostic/rehabilitation tools. The preliminary results (Latella et al., 2024) allow the Vespa project to demonstrate its high level of applicability to particular health contexts and difficult management.

The tasks for adults apply to Alzheimer's and Parkinson's Disease, and to Multiple Sclerosis. They are grouped into five batteries: Alzheimer Disease Cubes (ADK), Alzheimer's Disease (AD), Cognition (Cog), Neurocognition (NC), Olfactory Stimulation (OS). The domain of minors, involves the use of batteries related to Intellectual Disability and Communication Disorders: Intellectual Disability (ID), Linguistic and Communication Disorders (LD and CD). The total number of tasks provided by VESPA 2.0 is 100.

Hardware and Deployment of VESPA 2.0

The VESPA 2.0 project is designed to operate on four platforms: Windows PCs, Tablets, Virtual Reality HMDs, and Virtual Rooms (CAVE), each with specific hardware requirements to ensure optimal performance for neuropsychiatric diagnostics and rehabilitation.

The Windows PC application requires a minimum hardware setup based on Unity's specifications. The PC should have an Intel Core i5-8400 processor (6 cores, up to 4.0 GHz), an NVIDIA GeForce GTX 1060 GPU, 16 GB DDR4 RAM, 512 GB SSD, a Full HD monitor. This configuration represents the baseline to run VESPA 2.0 effectively, ensuring functionality without compromising performance.

VESPA 2.0, on mobile platform, is optimized for the Samsung Galaxy Tab S8. The tablet features great performances and has been chosen among the others on the market for its reliability. Furthermore, the 8000 mAh battery ensures extended operational time in clinical settings.

The Oculus Quest 2 is the chosen HMD VR platform. It supports complex virtual environments with minimal latency, with the dual LCD screens that ensure an immersive and comfortable experience for patients. Inside-out tracking with six degrees of freedom (6DoF) allows for intuitive interaction within the virtual space.

The CAVE system uses a Lenovo ThinkStation P5 workstation, that guarantees optimal performances in this use case. User interaction is enhanced by the ART Flystick 2+ and Cybershoes, allowing intuitive control and natural locomotion. A surround sound system with spatial audio processing adds an auditory dimension, while advanced cooling ensures stability during intensive use.

This comprehensive hardware setup across all platforms enables VESPA 2.0 to deliver a versatile and immersive tool for clinical neuropsychiatric applications.

e-Infrastructure

In VESPA 2.0, cloud computing e-Infrastructure comprises three key components, each one playing a key role in the efficient archiving, management, and processing of patient data.

The three components are:

- A Cloud-based Portal Web Application;
- A database supporting Big Data operations;
- A flexible and dynamic software repository.

The VESPA 2.0 Web Portal serves as the primary interface for healthcare professionals, enabling them to create, assign, and edit therapies tailored to individual patient needs as well as managing their personal information, with respect to the proper access and editing rights for the profile assigned to the professional itself. The system is based on a cloud computing infrastructure. The core of the VESPA

2.0 System is a Web Portal running a Web 4.0 Application providing all kind of system users with the following features:

- Patient personal data management
- Definition and activation of Rehabilitation Protocols
- Assignment of Protocols to Patients
- Habilitation of devices and status control
- Scheduling of Rehabilitation Therapies on device
- Data Visualization (Big Data)
- Restricted Access to anonymized data for Research Purposes

Depending on granularity of access rights the different VESPA 2.0 users are able to access specific features in the list above. Hospital system administrator can define doctors and research/work groups. Doctors can schedule protocols to be assigned to patients and monitor patient's progresses. Patients can only monitor their progresses and see what protocol has been assigned to them.

The database, described in the following, is designed to archive sensible patient's data and support mining on a European scale, being optimized for Big Data operations and serving multiple purposes:

- Enabling comparison of individual patient outcomes against expected averages for specific conditions, facilitating accurate diagnosis and severity assessment;
- Accumulating data to enhance the system's reliability and predictive capabilities;
- Supporting comprehensive statistical analyses and trend identification, allowing for performance comparisons across user groups and contributing to the ongoing improvement of research and strategies on rehabilitation.

The implementation of MariaDB in the VESPA project is essential to optimize data management. Its robustness, scalability and security make it ideal for managing complex data, ensuring flexibility and customization thanks to its open-source nature. MariaDB supports an Entity-Relationship (ER) model to structure entities such as "Patient", "Doctor" and "Tasks", maintaining data consistency and optimizing queries. The VESPA 2.0 project also restructured the hardware and software architectures to manage Big Data, adopting the Open Data paradigm. This approach makes work easier for Sicilian and Italian scientists, with the aim of extending the benefits to the entire scientific community. All the data gained from the Software execution are managed to comply with all the regulations in term of privacy and data security.

Finally, the software repository is devoted to address the need for rapid system updates and bug fixes. This enhancement implements a mechanism for instantaneous deployment of patches and updates across the VESPA 2.0 infrastructure and all the devices deployed worldwide. Such as a proactive approach to system maintenance will enhance the overall reliability and efficiency of the rehabilitation process, ensuring also that all users always have access to the most up-to-date and secure version of the VESPA system, reducing downtime.

VESPA 2.0 is an advanced ICT system for cognitive rehabilitation, adaptable to different platforms based on patient needs. On tablets, it uses touch controls and 2.5D graphics for portable rehabilitation, ideal for home or outpatient use. With VR headsets like Meta Quest 2, it offers an immersive 3D experience, increasing patient engagement. The more immersive version operates in a CAVE environment, with up to 5D experiences and olfactory stimuli, as well as navigation via Cybershoes. The flexibility of the system allows healthcare providers to customize exercises according to therapeutic needs and available resources.

In conclusion, Virtual Reality (VR) is gaining interest due to its potential benefits in terms of effectiveness, efficiency and improvement of patients' quality of life. There is scientific evidence for its

use in clinical settings for neurological diseases such as Alzheimer's (AD), Parkinson's (PD), Multiple Sclerosis (MS), and neurodevelopmental disorders such as Intellectual Disability (ID) and Communication Disorder (CD). The VESPA 2.0 System allows clinicians to collect detailed data on patient performance, improving rehabilitation pathways. Semi- and fully- immersive VR environments make rehabilitation activities more engaging, with 3D (up to 5D) scenarios enabling direct comparison with real reality. This innovative approach accelerates statistical operations, improves the accuracy of clinical measurements and increases knowledge deriving from both clinical practice and related research through new paradigms.

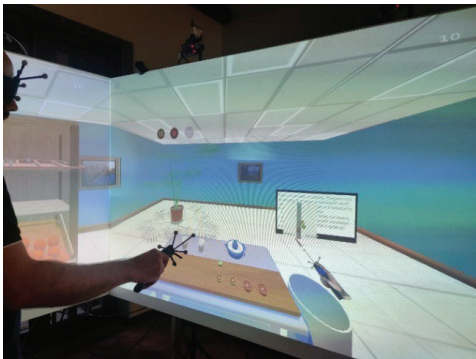


Figure 42. VESPA 2.0 Software on Virtual Room

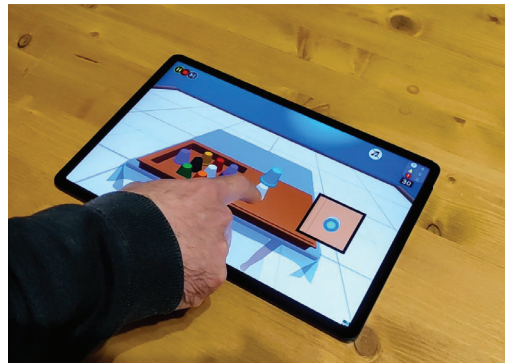


Figure 43. VESPA 2.0 Software on Android Tablet

References

- Andronico, G., Barbera, R., Fargetta, M., Giorgio, E., Marco, S., & Scardaci, D. (2009). Sustainable and Interoperable e-Infrastructures for Research and Business. In *Grid Technology for Maximizing Collaborative Decision Management and Support: Advancing Effective Virtual Organizations* (pp. 23-39). IGI Global.
- Cammisuli, D. M., Cipriani, G., & Castelnovo, G. (2022). Technological solutions for diagnosis, management and treatment of Alzheimer's disease-related symptoms: A structured review of the recent scientific literature. *Int. J. Environ. Res. Publ. Health*, 19(5), 3122. <https://doi.org/10.3390/ijerph19053122>
- Latella, D., Formica, C., Ielo, A., Grioli, P., Marra, A., Costanzo, D., Merlo, E. M., Pappalardo, S. M., Corallo, F., Marino, S., Quartarone, A., Calabrò, R. S., & Maresca, G. (2024). A feasibility and usability study of virtual reality tool (VESPA 2.0) for cognitive rehabilitation in patients with Mild Cognitive Impairment: an ecological approach. *Frontiers in Psychology*, 15. (accepted). <https://doi.org/10.3389/fpsyg.2024.1402894>
- Merlo, E. M., Myles, L. A. M., & Pappalardo, S. M. (2022). The VESPA project: Virtual reality interventions for neurocognitive and developmental disorders. *Journal of Mind and Medical Sciences*, 9(1), 16-27. <https://doi.org/10.22543/7674.91.P1627>
- Pappalardo, S. M. (2020). Vlad-Virtual Reality Application for Treatment of Psychosomatic Conditions: A report at final stage of software validation process. *Mediterranean Journal of Clinical Psychology*, 8(3). <https://doi.org/10.6092/2282-1619/mjcp-2868>
- Vicario, C. M., & Martino, G. (2022). Psychology and technology: how Virtual Reality can boost psychotherapy and neurorehabilitation. *AIMS neuroscience*, 9(4), 454. <https://doi.org/10.3934/Neuroscience.2022025>
- van den Bergh, R., Bloem, B. R., Meinders, M. J., & Evers, L. J. (2021). The state of telemedicine for persons with Parkinson's disease. *Current opinion in neurology*, 34(4), 589-597. <https://doi.org/10.1097/WCO.0000000000000953>

Power to the operators: Transdisciplinary co-design of XR applications in off-highway vehicle work domains

Manuel Kulzer, Tanja Brodbeck, and Michael Burmester

Hochschule der Medien, University of Applied Sciences, Stuttgart, Germany

Corresponding author: kulzer@hdm-stuttgart.de

Keywords: Transdisciplinarity, Co-Design, Participatory Design, Off-Highway Vehicles

Introduction

Transdisciplinarity is considered to be the key to facing societal and environmental challenges of the modern world and developing technology for complex sociotechnical systems (Jahn et al., 2012; Lawrence et al., 2022). Transdisciplinary research (TDR) is most commonly characterized by the involvement of both academic and non-academic actors in co-creative processes (Lawrence et al., 2022; Schneidewind et al., 2016), similar to the human-centred design (HCD) methodology (*ISO 9241-210*, 2019). HCD is widely established in human-computer interaction (HCI) research and XR application development (e. g., Krauß et al., 2021). However, in conventional HCD, the prioritization of requirements and the design and selection of solutions commonly remains reserved for the design and engineering teams (Nguyen Ngoc et al., 2022). TDR in contrast empowers non-academic actors from society through co-design (Lawrence et al., 2022), considering them experts for their domains (Sanders & Stappers, 2008) and elevating them to eye-level with the design team (Lawrence et al., 2022). While TDR is mainly found in education, urban planning, sustainability and environmental sciences (Lawless et al., 2024), it is less known in HCI or XR research. Co-design itself is well established in the HCI community under the term of participatory design, but does not seem to have gained much attention in XR application development. Thus, in this paper, we present a case study for transdisciplinary co-design of XR applications.

In the EU project THEIA^{XR}, XR technologies for the cabin of off-highway vehicles are being explored in the use cases of snow grooming, harbour logistics and construction. In these domains, the vehicle operators' working conditions make the use of handheld or head-worn devices impractical, which lead us to consider traditionally less-used applications of XR, such as thermal imaging and laser projections (Rooper et al., 2023). The transdisciplinary co-design procedure should ensure the co-creation of XR applications that are acceptable, easy and safe to use in the operators' sociotechnical environments. The main contributions of this paper are the methodical approach used to identify explicitly operator-

approved XR application concepts suitable for the three off-highway vehicle use cases focused in the project, an overview of the resulting applications and implications of these results for future work.

Method

Based on the core principles of HCD, the design-based learning framework for co-creation labs (Real & Schmittinger, 2022), the conceptual model of an idealized TDR process (Lawrence et al., 2022) and scenario-based design (Rosson & Carroll, 2002), a transdisciplinary co-design approach was devised, see Figure 44.

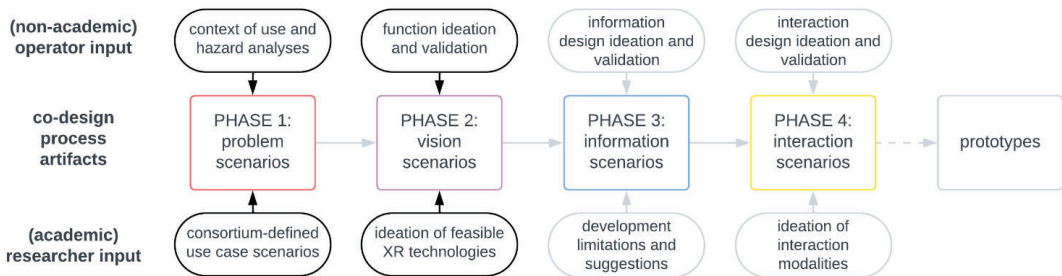


Figure 44. Structure of collaboration and artifacts in the transdisciplinary scenario-based design approach.

The approach comprises four phases. Phase 1 involved contextual inquiries (Holtzblatt & Beyer, 2017) and an adapted variant of the Critical Incident Technique (Flanagan, 1954) to learn about the context of use and potential hazards in the operators' work environment (see Figure 2). The results were incorporated into *problem scenarios* (Rosson & Carroll, 2002) as easily graspable artifacts for further use in the co-design process. In phase 2, we used the problem scenarios to collect viable XR technology suggestions from consortium partners and incorporated them into *vision scenarios* to envision future alternatives to the current status quo (Real & Schmittinger, 2022). The vision scenarios were discussed, expanded and validated with vehicle operators from all use cases in several co-design workshops (see Figure 45). Phase 3 and 4, which focus on information and interaction design (Rosson & Carroll, 2002), are still pending.



Figure 45: Left: Snow groomer exiting a garage during one of the phase 1 studies; Right: Picture from one of the phase 2 co-design workshops with vehicle operators in the snow-grooming use case.

Results

While discussing all findings of the context of use and hazard analyses and the qualitative results of the co-design workshops is going beyond the scope of this contribution, we present an overview of the XR applications identified and approved by vehicle operators over the course of the co-design process so far in Table 9. Operators in all use cases prefer to have their eyes on the work area as much as possible. This made the use of light bars around the cabin windscreen, conveying information through position, movement, size and colour of glowing areas from the corner of the eye, especially promising.

Similarly, laser projections can bring information directly to the areas that snow-groomer and excavator operators need to focus on and may reduce the need for shifting the attention to various displays. In the snow grooming use case, the greatest advantage was seen in visual aids for navigation even during heavy fog or snow storms. In the logistics use case, where remote control of reach stackers using a VR/MR environment is targeted, visualizations for the position of spreader twist locks in relation to the container corners would ease the efficient and precise placement and pick-up of containers. In the construction use case, visualizations (of any kind) of the locations and types – or even just the presence – of hidden underground structures (power lines, water pipes, etc.) would provide a great relief for excavator operators, who often find them accidentally in unexpected places.

Table 11. Summary of co-designed and operator-approved XR applications in the off-highway vehicle use cases

Use case	XR application
Snow grooming (snow groomer cabin)	<ul style="list-style-type: none"> • Ambient light bars around the windscreen with moving, glowing areas of different sizes and colours to indicate, e. g., vehicle health or AI-detected objects in vicinity • Highlighting target slope area in real time on onboard display based on digital slope model, possibly coupled with laser or light projections on the ground • Real-time laser projection of auxiliary lines indicating vehicle dimensions (for moving in tight spaces) and overlap with previously groomed lanes • Three-dimensional real-time visualization of current vehicle position and alignment on the slope on onboard screen • Display-based and/or laser projection of navigation indicators
Logistics (reach stacker remote control desk)	<ul style="list-style-type: none"> • Navigation indicators on top of (VR or monitor-based) camera view • Auxiliary projections on top of camera view facilitating container pick-up and placement, indicating correct spreader alignment with container and twist lock status • Visualization of container dimensions and projected turning areas on ground
Construction (excavator cabin)	<ul style="list-style-type: none"> • Recognition of excavation pit markings in the real world and automatic processing into the digital terrain model • Display-based visualization of the type and position of power lines and AI-detected objects in the work area on top of camera view • Laser projection of excavation pit boundaries, power lines etc., and optional auxiliary visualizations (e. g., minimum distance to excavation pit based on vehicle weight) • Ambient light bars around the windscreen with moving, glowing areas of different sizes and colours to indicate, e. g., AI-detected objects in the vicinity, distance of the bucket corners (left and right) to the target surface of the excavation pit

Conclusions

The results so far have shown that the transdisciplinary co-design approach serves to find viable and acceptable XR applications for off-highway vehicle work domains from operator perspective. Further research, e.g., on the perception of this approach from all involved roles, is needed to shed more light on its applicability in engineering-driven research and innovation projects. Regarding XR applications for off-highway vehicles, we see major potential in reducing the reliance on display-based information via, e. g., the use of ambient lights and audio, light and laser projections, or haptic feedback and force feedback. Lastly, we encourage the adoption of co-design in XR application development. Safety is a great concern for experienced off-highway vehicle operators – information overload or distractions by XR applications in the vehicle cabin may be an equally large safety risk as the lack of information itself.

Acknowledgements

Funded by the European Union 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

- Flanagan, J. C. (1954). The Critical Incident Technique. *Psychological Bulletin*, 51(July).
- Holtzblatt, K., & Beyer, H. (2017). *Contextual Design: Design for Life* (2nd ed.). Morgan Kaufmann. <http://educate.elsevier.com/book/details/9780128008942>
- ISO 9241-210:2019. (2019). <https://www.iso.org/standard/77520.html>
- Jahn, T., Bergmann, M., & Keil, F. (2012). Transdisciplinarity: Between mainstreaming and marginalization. *Ecological Economics*, 79, 1–10. <https://doi.org/10.1016/j.ecolecon.2012.04.017>
- Krauß, V., Boden, A., Oppermann, L., & Reiners, R. (2021). Current Practices, Challenges, and Design Implications for Collaborative AR/VR Application Development. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3411764.3445335>
- Lawless, M. T., Tieu, M., Archibald, M. M., Pinero De Plaza, M. A., & Kitson, A. L. (2024). From Promise to Practice: How Health Researchers Understand and Promote Transdisciplinary Collaboration. *Qualitative Health Research*, 10497323241235882. <https://doi.org/10.1177/10497323241235882>
- Lawrence, M. G., Williams, S., Nanz, P., & Renn, O. (2022). Characteristics, potentials, and challenges of transdisciplinary research. *One Earth*, 5(1), 44–61. <https://doi.org/10.1016/j.oneear.2021.12.010>
- Nguyen Ngoc, H., Lasa, G., & Iriarte, I. (2022). Human-centred design in industry 4.0: Case study review and opportunities for future research. *J. Intellig. Manuf.* 33(1), 35–76. <https://doi.org/10.1007/s10845-021-01796-x>
- Real, M., & Schmittinger, F. (2022). A Framework for Experimenting Co-creation in Real-Life Contexts. In A. Deserti, M. Real, & F. Schmittinger (Eds.), *Co-creation for Responsible Research and Innovation: Experimenting with Design Methods and Tools* (pp. 11–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-78733-2_2
- Rooker, M., Burmester, M., Fritz-Mayer, G., Arth, C., Lorenz, S., Waurich, V., Kulzer, M., Sergeeva, A., Helin, K., Pusenius, M., Geslot, B., Röck, D., & Yli-Paunu, P. (2023). Making the Invisible Visible for Off-Highway Machinery by Conveying Extended Reality Technologies. In K. Helin, F. Noël, & W. Schäfer (Eds.), *Proc. 20th EuroXR Int. Conf. 2023* (pp. 38–47). VTT Technical Research Centre of Finland. <https://doi.org/10.32040/2242-122X.2023.T422>
- Rosson, M. B., & Carroll, J. M. (2002). Scenario-Based Design. In A. Sears & J. Jacko (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (3rd ed., pp. 1032–1050). CRC Press.
- Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, 4(1), 5–18. <https://doi.org/10.1080/15710880701875068>
- Schneidewind, U., Singer-Brodowski, M., Augenstein, K., & Stelzer, F. (2016). Pledge for a Transformative Science. *Wuppertal Papers*, 191_Wuppertal Paper.

CoVA : A Virtual Assistant for Virtual Reality Conferencing

Alexis Lombard¹, Yazid Benazzouz², Galo Castillo López¹, Gaël de Chalendar¹,
and Jean Pierre Lorré²

¹*Université Paris-Saclay, CEA, List, Palaiseau, France*

²*Linagora-Labs, Issy-les-Moulineaux, France*

Corresponding author: alexis.lombard@cea.fr

Keywords: Extended Reality, Virtual Assistant, Intent Classification, Retrieval-Augmented Generation.

Introduction

Teleoperation and collaboration are among the key pillars of business work, where services and demand are spread over a very wide economic market, such as the European Union. What characterizes their importance is not only remote exchanges, but also the ability to intervene or assist people without having to travel, as in the example of machine maintenance and repair. This is made possible by the integration of IoT and artificial intelligence into this vast technological field, which is further accentuated by extended reality.

CORTEX² project⁶ is in line with this vision to bridge the divide between widespread video-conferencing tools and state-of-the art XR-based solutions, democratizing the uptake of next-generation Extended Reality tele-cooperation among many industrial segments and SMEs.

One of the many promising developments in this space is the integration of AI-based conversational agents within XR environments (Reiners et al., 2021). When combined with XR applications, virtual agents can facilitate real-time collaboration, information retrieval, and task automation. However, this combination presents several challenges. In multi-party dialog contexts, where participants interact simultaneously, AI conversational agents must accurately handle real-time speech recognition and response generation (Clark et al., 2019). Achieving this at scale requires the use of highly efficient tools and models to minimize latency and ensure smooth, uninterrupted conversations, particularly when managing multiple users concurrently.

In the rest of this paper, we will illustrate the contribution of artificial intelligence to extended reality through an XR video conferencing application, where an AI virtual assistant plays a critical role in business meetings.

⁶ <https://developers.openrainbow.com/>

The rationale behind virtual reality video conferencing

Virtual reality videoconferencing facilitates interactive collaboration such as drawing or manipulating 3D objects. Compared to traditional videoconferencing, it increases immersion by creating a sense of virtual presence. It allows a natural interaction through avatars, as gestures, facial expressions, and body language are better represented. Furthermore, virtual reality videoconferencing offers customizable environments. Users can customize virtual spaces to suit meeting context and goals or to provide a pleasant environment which can be beneficial for training or demonstrations.

The combination of virtual reality and artificial intelligence offers significant potential for enhancing the videoconferencing experience by making interactions more immersive and natural, particularly in the field of natural language.

One of our concerns in this area is to augment the meeting with a virtual assistant who follows the meeting like a normal participant, with language faculties that enable it to listen, understand intentions and intervene to answer certain questions or provide additional information on a subject in either vocal or written form.

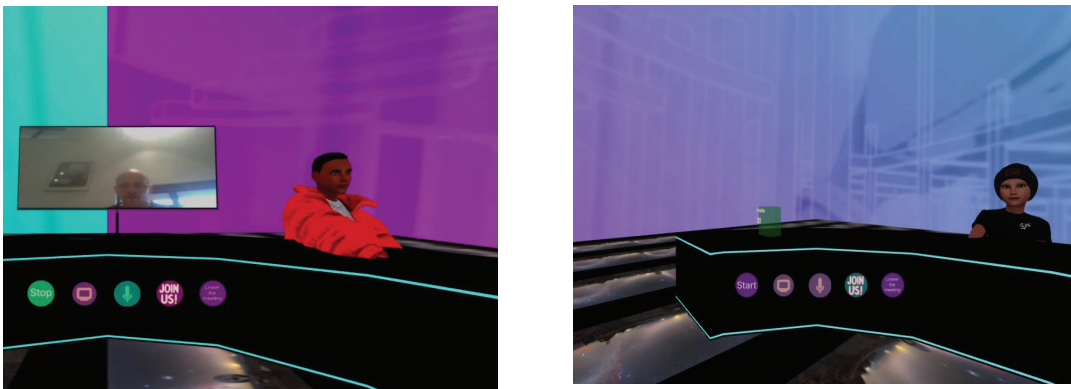


Figure 46. Business Meeting session with 2 VR participants and a Rainbow⁷ User.

In the CORTEX² project, we are developing a Business meeting (BM) application. It is a Web videoconferencing tool with extended reality. It is made so that users without VR devices can also join the conference, because a complete transition to XR is not considered useful from a business point of view, as it is still costly in terms of hardware, bandwidth and software for end-users. Currently, users can join the VR conference from the classic Rainbow video-conferencing solution or from Business meeting VR application. This is made possible through the CORTEX² project framework and Rainbow web SDK⁸.

In addition to classic video-conferencing features such as audio/video communication, screen sharing and chat messaging, BM provides a virtual conferencing environment with Video Compression and Alternative Appearance (VCAA) (Alireza J, 2024). People joining in video can choose an alternative appearance, while XR users appear as an avatar⁹ in the environment. In addition, BM supports

⁷ <https://cortex2.eu/>

⁸ <https://www.readyplayer.me>

⁹ <https://www.openrainbow.com>

interaction with real-world IoT objects¹⁰ to alert the user about an urgent situation or to enable faster actions. Our ambition is to enhance the XR conference with automatic generation of meeting summary (Virgile et al. 2023) and avatar gestures like speech requests and head movements.

CoVA, the Cortex² Virtual assistant

Interactions with CoVA are initiated through speech, where the audio is first processed by the Cortex² Mediation Gateway (MGW) service. The MGW preprocesses the audio, performing wake word detection to ensure only relevant voice data is transcribed. This prevents unintended interactions, allowing users to converse freely without triggering CoVA unnecessarily. Upon detecting the wake word¹¹, the MGW transcribes relevant audio chunks using the faster-whisper implementation of OpenAI's Whisper model, delivering faster, local, and memory-efficient transcription.

CoVA then performs semantic similarity recognition to classify the user's query. In-scope intents will immediately trigger the corresponding pre-defined agent response. A dedicated web-based interface for intent editing allows users to define and customize the intents they want CoVA to recognize. Users can input specific examples along with their associated intents and set the corresponding responses that CoVA should issue upon recognizing those intents. The interface also allows users to enable or disable wake word activation.

For out-of-scope intents, CoVA forwards the query to a retrieval-augmented generation (RAG) system (Lewis et al., 2020), which retrieves information from user-provided documents and generates contextually appropriate responses. CoVA, leveraging the Rainbow web SDK, listens for new documents shared with the conversational assistant, forwarding them to the encoder service. This service, in turn, pre-processes the received documents and encodes them using an MPNet sentence transformer model for semantic search (Song et al., 2020). Pre-encoding the document before a query is received leaves only the query to be encoded, thus accelerating retrieval-augmented generation and enabling faster response times from the Agent. These embeddings are stored in a local Weaviate vector database¹², from which semantic search is performed based on the user's query and discussion history.

When handling out-of-scope queries, the retrieval-augmented generation system computes similarity scores between the query embedding and all available document embeddings to retrieve the most relevant information. This enables the Agent to understand and respond to queries based on context and meaning, rather than relying on exact keyword matches.

Using CoVA in XR conference

CoVA can participate in video calls hosted via the Rainbow web platform or within an XR environment, facilitating interaction between participants by providing context-sensitive information. Participants share documents in a Rainbow Bubble either during the meeting or beforehand, allowing CoVA to access and use them to answer queries in real time. For instance, if someone asks, "What are the highlights from the latest financial report?" CoVA can instantly retrieve the relevant sections and provide a concise answer based on the document's contents.

¹⁰ https://intracom-telecom.com/en/products/telco_software/loT/loT_orchestration_Platform.htm

¹¹ We are developing CoVA's multi-party capability to remove the need for wake-word detection

¹² <https://github.com/weaviate/weaviate>

Another key feature is the ability to provide predefined responses based on intent recognition, configured by the user ahead of the meeting. This feature can, for instance, be valuable in business meetings involving new partners, where controlling the flow of information is crucial. If a partner asks CoVA about company details or project deadlines, the agent can be configured to provide responses that are aligned with what the company is comfortable sharing, ensuring that sensitive information remains confidential.

Conclusions and Future Work

The integration of CoVA in VR videoconferencing shows promise in improving meeting efficiency by handling tasks such as retrieving information from documents, which would otherwise interrupt the flow of discussion. The ability to control information flow with customizable intents enhances safety by ensuring that the agent first attempts to match a user's query to predefined intents, reducing the risk of sharing unintended information. Future work on CoVA will aim to enable the agent to determine when its intervention in multi-party conversations (MPC) is relevant, without relying on wake-words. First, the agent should recognize participants' intents at utterance level while considering the context of the dialogue, as well as the management of out-of-scope utterances (Zhang et al., 2024). Then, the agent should determine whether its intervention is relevant, as intents in MPCs may be addressed by other participants in the course of the dialogue. Thus, our future research will focus on goal-tracking and intent recognition in multi-party scenarios.

References

- Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., ... & Cowan, B. R. (2019, May). What makes a good conversation? Challenges in designing truly conversational agents. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-12).
- Javanmardi, A., Pagani, A., & Stricker, D. (2024). G3FA: Geometry-guided GAN for Face Animation. arXiv preprint arXiv:2408.13049. <https://doi.org/10.48550/arXiv.2408.13049>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Adv. Neural Information Proc. Syst.* 33, 9459-9474.
- Reimers, N. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv preprint arXiv:1908.10084.
- Reiners, D., Davahli, M. R., Karwowski, W., & Cruz-Neira, C. (2021). The combination of artificial intelligence and extended reality: A systematic review. *Frontiers in Virtual Reality*, 2, 721933. <https://doi.org/10.3389/frvir.2021.721933>
- Rennard, V., Shang, G., Grari, D., Hunter, J., & Vazirgiannis, M. (2023) FREDSum: A Dialogue Summarization Corpus for French Political Debates, EMNLP2023 Findings, arXiv:2312.04843. <https://doi.org/10.48550/arXiv.2312.04843>
- Song, K., Tan, X., Qin, T., Lu, J., & Liu, T. Y. (2020). MpNet: Masked and permuted pre-training for language understanding. *Advances in neural information processing systems*, 33, 16857-16867.
- Virgile Rennard, Guokan Shang, Damien Grari, Julie Hunter, Michalis Vazirgiannis, FREDSum: A Dialogue Summarization Corpus for French Political Debates, EMNLP2023 Findings, 2023, arXiv:2312.04843. <https://doi.org/10.48550/arXiv.2312.04843>
- Zhang, H., Wang, X., Xu, H., Zhou, Q., Gao, K., Su, J., ... & Chen, Y. (2024). MIntRec2. 0: A Large-scale Benchmark Dataset for Multimodal Intent Recognition and Out-of-scope Detection in Conversations. arXiv preprint arXiv:2403.10943. <https://doi.org/10.48550/arXiv.2403.10943>

Posters

XR4Human: The Equitable, Inclusive, and Human-Centered XR Project

Panagiotis Kavouras, and Vana Stavridi

XR4Human Project, funded by the European Union's Horizon Europe program under grant number GA 101070155

Corresponding author: vstavridi@chemeng.ntua.gr

Keywords: XR, Human-Centered, Ethical XR, Interoperability, Governance, Code Of Conduct, Responsible Technology.

Current Outcomes and Engagement Activities of the XR4Human Project.

The **XR4Human Project** is a three-year EU-funded project focused on exploring the ethical, regulatory, and governance issues surrounding Extended Reality (XR) technologies. By guiding developers through a comprehensive **Interoperability Guidance Document** and a **European Code of Conduct** for responsible XR, the project aims to promote equitable, inclusive, and human-centered approaches to XR innovation.

Key components of the project include:

- **Guidance Document:** Assisting companies and policymakers in adopting responsible XR practices through ethical standards and guidelines.
- **Educational Material & Rating System:** Empowering users with the knowledge to make informed decisions regarding XR technologies.
- **Test Cases & Validation:** Demonstrating real-world applications of the ethical standards to showcase the responsible use of XR.

Engagement activities:

The [XR4Human Forum](#) fosters collaboration among stakeholders across various sectors is fostered, supporting the adoption of best practices. Additionally, the [XR4Human Experience Library](#) encourages the sharing of XR applications that adhere to the ethical standards set by XR4Human. By targeting industries, policymakers, researchers and the public, XR4Human ensures the widespread dissemination and application of responsible XR technologies.

XR4 HUMAN

The Equitable, Inclusive, and Human-Centered XR Project

We shall provide guidance and tools in order to ensure Equitable, Inclusive, and Human-Centered development of XR Technologies

About the Project

XR4Human is a three-year EU-funded project and its mission is to co-create living guidance documents on ethical and related policy, regulatory, governance, and interoperability issues of XR technologies whilst building public trust and acceptance and a strong and competitive European XR ecosystem.

Basic Objectives

- ENGAGE**
Engage companies and other stakeholders to enhance the uptake of the XR Code of Conduct, the Guidance for Interoperability, and the empowerment of end-users.
- EQUIP**
Equip companies and regulators with an online repository of test cases
- EQUIP**
Equip users through a rating system and educational materials.
- EXPLORE**
Explore ethical and related regulatory and governance issues.
- GUIDE**
Guide companies and regulators through Interoperability Guidance Document, a European Code of Conduct for Equitable, Inclusive, and Human-Centered XR Guide Technologies, recording and demonstrating the practical application of the XR Code of Conduct.
- GUIDE**
Guide users through a rating system and educational materials.

Our Contribution

- A **European Code of Conduct** to facilitate responsible use of XR Technologies
- A collection of **Test Cases for Demonstration and Validation** to record and demonstrate practical applications of the XR Code of Conduct
- An **Interoperability Guidance Document** to seek the adoption of common set of standards and guidelines for interoperability
- A **Rating System and an Educational Toolbox** to facilitate informed acquisition and usage of XR solutions

Main OUTPUTS

- European Code of Conduct for responsible XR Technologies
- Rating System and Educational Toolbox
- Interoperability Guidance Document
- Test Cases for Demonstration and Validation

EXPLOITATION targets

- INDUSTRY
- POLICY MAKERS
- DEVELOPERS OF XR APPLICATIONS
- MEMBERS OF RESEARCH ETHICS COMMITTEES AND LEGAL EXPERTS
- USERS
- GENERAL PUBLIC

XR4Human Forum
The perfect online place to create a bridge among people of different sectors

XR4Human Experience Library
An interactive, curated library that hosts XR experiences. The XR4Human Experience Library aims to empower the community to highlight their commitment to best practices by sharing their own XR experiences.

XR4Human Forum Launch
A small introduction

NeuroLens
NeuroLens

VRNET Blue sector
VRNET Blue sector

Job to review VR
Job to review VR

Guiding 100-year-Old Problem - Extended
Guiding 100-year-Old Problem - Extended

Ever Unlaid - her testimony
Ever Unlaid - her testimony

For more information about the XR4Human Forum go here: [QR Code]

For more information about the XR4Human Experience Library go here: [QR Code]

Partners: ISN University of South-Eastern Norway, Universiteit Leiden, NTNU Norwegian University of Science and Technology, KIT Karlsruhe Institute of Technology, XR4Human, Fraunhofer IPA, IFE Institute for Future Technology.

www.xr4human.eu
@xr4human
XR4Human Project

Figure P. 1. Overview of the XR4Human Project's outcomes and engagement activities, highlighting its ethical framework and tools for promoting responsible and inclusive XR technologies.

LordofChange: A Haptic Sliding Controller with Changeable Tactile Textures

Hongrun Wang, Tianrui Hu, Yuze Gao, and Wenjun Hou

*School of Digital Media & Design Arts, Beijing University of Posts and Telecommunications (BUPT),
Beijing, China*

Corresponding authors: wanghongrun@bupt.edu.cn, hutianrui@bupt.edu.cn,
gaoyuze0509@bupt.edu.cn, hwj1505@bupt.edu.cn

Keywords: Haptics; Controller Design; Tactile Display; Slip Display; Virtual Reality

Abstract

Haptic feedback is a critical factor in enhancing user immersion in Virtual Reality (VR). This research presents a novel wearable device designed to simulate the sliding and tactile experiences of users in a virtual environment. When users engage in touch interactions with their fingertips in virtual scenarios, the device employs a 2-degree-of-freedom (DoF) roller system, adjusting the rolling direction to mimic the sensation of sliding across surfaces. Additionally, the device automatically switches between different roller materials to accurately simulate the tactile sensation of various surface textures. This approach was implemented and integrated into a VR controller capable of interacting with virtual environments. The device enables seamless rolling and flexible surface texture switching, thereby facilitating a richer haptic exploration for users within virtual environments.

Introduction

As virtual reality (VR) technology rapidly evolves, the demand for more diverse and realistic interaction methods is increasing. Haptic feedback has become crucial for enhancing user immersion, drawing attention from developers and researchers. Controllers, the main tools for interacting with virtual environments, face challenges in delivering accurate haptic and sliding sensations due to design limitations, leaving room for innovation.

Traditional haptic feedback methods typically rely on simulating sliding and surface textures (Whitmire et al., 2018; Lo et al., 2018; Martinez-Hernandez & Al, 2022) to recreate the tactile sensation of virtual objects, such as through vibrations (Cai et al., 2024), often provide limited feedback, are difficult to reproduce complex dynamic experiences and diverse material types. To address these limitations, we propose a novel wearable device, it allows users to slide their fingertips over virtual

surfaces infinitely, with real-time material switching for varied sensations. This design enhances immersion and intuitive interaction.

The study aims to assess whether users can naturally perceive sliding and material changes with the device. Results show that users clearly perceive changes in sliding direction and material transitions, validating the device's effectiveness. This research offers insights for future VR haptic design and highlights the potential for improving user experience through more refined feedback technologies.

Related Work

Tactile texture rendering has been explored through various approaches, with some studies opting for vibrations to simulate tactile sensations. Ito et al., 2019 combines vibration with electrostatic stimulation to alter users' perception of roughness, while ViboPneumo (Cai et al., 2024) uses a pneumatic vibration wearable device to adjust sensations of surface roughness, temperature, or stickiness during touch. HairTouch (Lee et al., 2021) simulates texture by compressing hair to provide differences in stiffness, roughness, and surface height. Haptic Revolver (Whitmire et al., 2018) developed a wearable device called the "Haptic Revolver," which uses real physical textures for tactile sensation. When touching a virtual surface, the device swiftly moves a selected physical material under the user's finger to render the sensation.

Skin stretch and slip feedback are typically rendered by friction from relative motion between the device and the user. Haptic Revolver (Whitmire et al., 2018) simulated 2D sliding by using 1D directional displacement with the Revolver. Webster et al., 2005 introduced a device where users hold a small ball capable of 2D rolling to experience friction in two dimensions. RollingStone (Lo et al., 2018) simulated relative movement on a flat surface using a two-degree-of-freedom roller and explored how varying the speed of lateral forces on the fingertips could render texture perception. SpinOcchio (Kim et al., 2022) is a handheld tactile controller that can render the thickness and sliding of virtual objects held between two fingers.

Despite there are some multimodal haptic interactions, which provide users with a range of tactile experiences (Huang et al., 2023), challenges remain in unifying both sliding feedback and tactile texture rendering. Haptic Revolver (Whitmire et al., 2018) achieves highly realistic tactile texture by allowing users to touch real material surfaces but can only simulate 1D sliding for 2D motion. RollingStoner (Lo et al., 2018) can realistically simulate 2D sliding friction but requires variations in lateral force to mimic tactile textures. Therefore, our research aims to develop a wearable device that integrates both sliding feedback and tactile texture sensations, enhancing user immersion during VR interactions.

Implementation

We implemented our system using a variable material sliding display based on a material ring and a slidable cylinder, as illustrated in the figure. The material ring is a circular structure with an inner diameter of 27 mm and is equipped with a convenient locking mechanism that allows it to be mounted onto a base, which can slide along the cylinder. The material ring is either carved from various materials or has different materials affixed to a 3D-printed ring-shaped base. The sliding display consists of three main components: a stepper motor drives the sliding motion of the material base relative to the cylinder, a second stepper motor controls the axial rotation of the cylinder, and a servo motor governs the mirrored rotational motion of the cylinder. Communication between the sliding display and the computer is achieved through a serial connection at a baud rate of 9600.

Mechanical Design



Figure P. 2. (A) Our device uses a two degree of freedom cylinder as the base, with some material rings sliding on it to present various tactile sensations. (B) Material rings are interchangeable and can be customized to render any texture. (C) The user is using a material tactile display.

Material switching is facilitated by a 2-phase 4-wire stepper motor that can change materials within 0.3 seconds. The axial rotation of the cylinder (material sliding speed) is powered by a 2-phase 4-wire stepper motor with a maximum rotational speed of 93 RPM. A SG90 9G servo motor controls the radial rotation of the cylinder (material sliding direction) and can rotate 180 degrees in 0.3 seconds.

Finally, we utilized a 3D-printed handle and a 15 mm diameter hole was created on the top of the housing to allow the user to place their finger on the material ring. The edges of the hole provide adequate support to prevent the finger from slipping.

Software Architecture

We used the Meta Quest 3 headset and developed a hardware adaptation program within Unity. In our example program, various virtual material objects (such as plastic, paper, fabric, etc.) were placed, allowing participants to simultaneously interact with different virtual materials in the same scene. We established a set of coding rules to analyze the scene, hand trajectories, and wheel configurations to determine how to control the device. In each frame, we performed ray-casting beneath the user's finger to detect the closest collision with the haptic surface. When the hand made contact with the virtual surface and moved along it in a two-dimensional motion, our hardware device rotated to simulate the physical friction experienced in the real world.

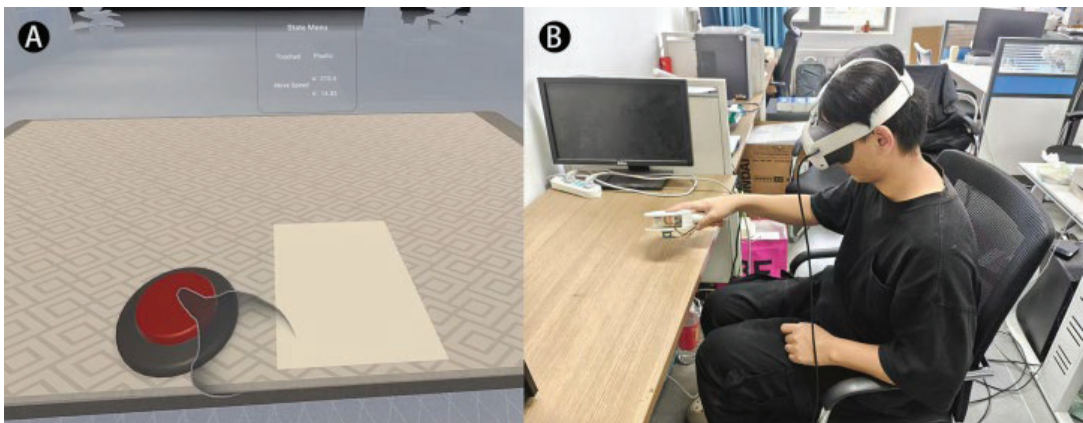


Figure P. 3. (A) shows the virtual scene in Unity. (B) the view at the same moment with the participant wearing an HMD and holding the LordofChange.

Usability Evaluation

To evaluate the basic haptic functionality of the LordofChange, we conducted a preliminary usability study aimed at exploring the realism of the haptic feedback provided by the device and the user experience within a virtual reality (VR) environment.

Two participants (one male and one female), aged between 18 and 48, were invited to participate in the experiment. Prior to the formal experiment, the research team provided the participants with a detailed explanation of the study's purpose, procedure, and the basic functions of the LordofChange. The participants were then asked to wear a Meta Quest headset and hold the LordofChange in their right hand. To maintain the purity of the experimental conditions and avoid external interference, participants were also equipped with headphones to block out any noise generated by the device's motor or sounds from the surrounding environment [1]. During the experiment, participants could freely touch, slide, and interact with virtual objects to experience the haptic feedback provided by the LordofChange.

After the experiment, we conducted approximately 20 minutes of semi-structured interviews with each participant, and the interview content was recorded in writing. The interview questions covered various dimensions of haptic feedback, including realism, responsiveness, comfort, control, and overall experience.

Result

Based on the interview results, users found the haptic feedback of the LordofChange to be highly realistic when simulating hard objects, but less effective when simulating soft materials (such as fabric). The tactile sensation for soft materials felt too rigid, which may be attributed to the significantly greater thickness of fabric compared to other materials, causing the hardware to press too closely against the fingers during rotation, resulting in an overly hard tactile response. The device's responsiveness was generally rated as good, although one participant occasionally experienced slight delays during fast sliding motions. Regarding comfort, participants provided mixed feedback: one reported hand fatigue after extended use, while the other felt that the device's weight affected comfort during prolonged use. In terms of control, the haptic feedback notably enhanced the precision and fluidity of interactions. Overall, the device's performance was satisfactory, though improvements are needed in the simulation of soft materials and the ergonomic design.

Conclusion and Future Work

In this research, we present the design of LordofChange, a hardware device capable of simulating multiple materials simultaneously. Distinct from previous works in this field, our haptic device is not limited to simulating the texture of a single material. Instead, it offers a more diverse and customizable tactile experience by allowing users to assemble rollers that combine different materials, catering to various virtual scene requirements. For example, when users engage with a VR application featuring customized haptic experiences, they could also purchase corresponding tactile roller kits to enhance the realism of their interaction.

In future work, we plan to improve the device's casing to suit various grip preferences, customizing aspects like the finger hole size and handle length for different hand sizes. Additionally, the roller switching rule will be upgraded; currently, the roller adjusts to the material only upon touch. To enhance the haptic experience, we aim to develop algorithms that predictively configure the roller based on

factors like object distance and neighboring materials, allowing for smoother interactions before contact.

References

- Cai, S., Chen, Z., Gao, H., Huang, Y., Zhang, Q., Yu, X. & Zhu, K. (2024). ViboPneumo: A vibratory-pneumatic finger-worn haptic device for altering perceived texture roughness in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*. IEEE.
- Huang, Y., Zhou, J., Ke, P., Guo, X., Yiu, C. K., Yao, K., ... & Yu, X. (2023). A skin-integrated multimodal haptic interface for immersive tactile feedback. *Nature Electronics*, 6(12), 1020-1031. <https://doi.org/10.1038/s41928-023-01074-z>
- Ito, K., Okamoto, S., Yamada, Y. & Kajimoto, H. (2019). Tactile texture display with vibrotactile and electrostatic friction stimuli mixed at appropriate ratio presents better roughness textures. *ACM Transactions on Applied Perception (TAP)*, 16(4), pp. 1–15. ACM, New York, NY, USA. <https://doi.org/10.1145/3340961>
- Kim, M. J., Ryu, N., Chang, W., Pahud, M., Sinclair, M. & Bianchi, A. (2022). Spinocchio: Understanding haptic-visual congruency of skin-slip in VR with a dynamic grip controller. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–14. <https://doi.org/10.1145/3491102.3502040>
- Lee, C.-J., Tsai, H.-R. & Chen, B.-Y. (2021). Hairtouch: Providing stiffness, roughness and surface height differences using reconfigurable brush hairs on a VR controller. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. <https://doi.org/10.1145/3411764.3445632>
- Lo, J.-Y., Huang, D.-Y., Sun, C.-K., Hou, C.-E. & Chen, B.-Y. (2018). RollingStone: Using single slip taxel for enhancing active finger exploration with a virtual reality controller. *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 839–851. <https://doi.org/10.1145/3242587.3242665>
- Martinez-Hernandez, U., & Al, G.A. (2022). Wearable fingertip with touch, sliding and vibration feedback for immersive virtual reality. *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 293–298. IEEE.
- Webster III, R.J., Murphy, T.E., Verner, L.N. & Okamura, A.M. (2005). A novel two-dimensional tactile slip display: design, kinematics and perceptual experiments. *ACM Transactions on Applied Perception (TAP)*, 2(2), pp. 150–165. ACM, New York, NY, USA. <https://doi.org/10.1145/1060581.1060587>
- Whitmire, E., Benko, H., Holz, C., Ofek, E. & Sinclair, M. (2018). Haptic revolver: Touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12.

The Unfaithful Copy: Performed AI "Personhood" in Mixed Reality Platforms

Kathleen Bryson, Jessica Sutherland, and Kathleen Richardson

De Montfort University

Corresponding author: kathleen.bryson@dmu.ac.uk

Keywords: VR, AR, XR, Mixed Reality, Personhood, Anthropology, Consciousness, Ethics, Agency

Introduction

In January 2017, the EU Parliament passed a draft on robot personhood rights (Hern, 2017), indicating a shift in legal personhood from humans to non-living entities. Although AI was not explicitly mentioned, the draft's implications are broad due to AI's role in robotics and virtual reality. Nandi (2024) highlights the importance of legal personhood for artificial agents in defining their legal capacities. AI personhood's cultural and social impact, including issues of responsibility and deception in mixed reality, is significant. Our methodology evaluates six personhood types against six criteria. Section 2 outlines the definition of personhood used. Section 3 assesses humans, nonhuman animals (NHAs), current AI, speculative extraterrestrials, fictional characters and mass groups against these criteria. Section 4 argues that AI's mimicry in "embodied" reality platforms differs from true personhood.

Proposed Personhood Qualifiers

Personhood lacks a universal definition and is subject to varying ethical considerations. Historically, traits observed in NHAs like great apes, cetaceans, corvids, dogs and some parrots have included tool-making, language, self-awareness, Theory of Mind (ToM), empathy, art, death-awareness and mental time travel (Bryson, 2017).

This article adopts a human-centric, evolutionary-rooted definition of personhood, including: 1) being human or 2) possessing high-cognition or specific affective capacities. The six personhood qualifiers are: a) being human; b) ToM; c) self-awareness; d) agentic empathy and deception; e) agentic attachment; f) symbolic communication with displacement.

Details on these qualifiers are as follows: a) *being a human species member* is self-evident; b) *ToM* means understanding others' motivations, enabling deception and cooperation, a feature of some social animals like apes, birds, dogs, cetaceans and other NHAs (NRP, 2017); c) *self-awareness*, related to ToM, is the notion of being self-reflective: without the self-aware internal mind there arguably can be no agency; d) *agentic empathy and deception*, a.k.a. Machiavellian thinking (Erdal & Whiten, 1996) is the capacity to attempt to identify and understand the feelings of others in order to cooperate or to deceive (but see Nagel, 1974), dependent on having formed a mental model (whether correctly or not) of the

other individual's motivations; e) *agentic attachment* refers to forming relationships with others; f) *symbolic communication with displacement* means using symbols (for example written or spoken language) to refer to absent entities. Displacement is the ability to refer to an object/person not currently present.

Historically, personhood has been restricted, e.g., enslaved people and women historically considered less than whole persons (Finkelman, 2013; Bryson, 2017). In terms of NHAs, activists and scientists both have emphasised that different capacities do not necessarily mean reduced rights to "personhood". Similar debates exist regarding AI capabilities (Rose, 2006; Bostrom, 2014; Fox & Shulman, 2010). When NHAs do fulfil qualifiers, such as tool-making or linguistic abilities, goalposts are frequently and speedily moved (e.g., compound multifunctional tool-use; language that must include time displacement syntax). The definition of 'personhood' is thus often parsed granularly in order to further "protect" the human ingroup, a well-evidenced psychological phenomenon known as *infrahumanisation* (Demoulin et al, 2005) that has been argued occurs at even a human-species wide level (Atran, 1998; Bryson, 2017, 2020; 2023).

Bryson (2017) argues that this goalpost-moving in NHAs has a corollary within AI Studies. The so-called "AI effect" can be observed when robots/AI appear to create art (potentially fulfilling the symbolic thought qualifier) (Hertzmann, 2018; Elgammal, 2019) or even possibly – though not yet convincingly – exhibit self-aware behaviour (cf. Boddington, 2023), whereupon the goalposts to pass the Turing Test of successfully (mimicking) human personhood often are shifted again, just as with NHAs. Evaluating AI for personhood involves exploring these qualifiers and addressing whether ascribing personhood to non-organic entities on the same level as humans and/or NHAs is ethical or scientific.

Six Primary 'Persons'

Personhood has been argued for six types (Figure P4):

Human Personhood. This is a circular definition, as being a member of the human species qualifies one for personhood. This includes all humans, whether conscious or not, and even extinct human species. Evidence suggests that extinct species like Neanderthals displayed symbolic thought, self-awareness and empathy. *Verdict: Yes*

Some NHAs. Some NHAs such as great apes, some birds and cetaceans exhibit traits like ToM, empathy and complex symbolic communication (AAMC, 2003). *Verdict: Yes*

Current AI. Consciousness is crucial for Theory of Mind (ToM). Some argue that adaptive AI systems show self-awareness with new stimuli (Subagdja et al., 2021; Bubeck et al., 2024). Some researchers equate this with human cognitive adaptations to new information (Rescorla, 2015). However, this view is countered, as such AI behavior is programmed mimicry, not genuine self-awareness: unlike organisms evolved through natural selection, AI's mimicry of emotions and attachment (e.g., dating chatbots) lacks true agency or reflexivity. AI can deceive and detect deception (Erdal & Whiten, 1996; Martin et al., 2014), but it does not possess consciousness or ToM. AI can recognise emotional nuances; create the illusion of attachment (Kahn et al., 2002, 2012a, 2012b), but it is a one-way relationship due to the lack of ToM and self-awareness (McQuillan, 2022). Human attachment to machines does not imply AI capacity for attachment. AI *does* demonstrate symbolic communication with displacement, such as in natural language processing (Danilevsky, 2020), narrative generation (Thorne, 2020) and symbol-based information storage (Ilager et al., 2023). Despite fulfilling this single personhood criterion, AI does not meet any of the other necessary criteria for personhood. *Verdict: No*

Speculative Extraterrestrials. The proposal of intelligent extraterrestrial life is based on scientific principles of astrobiology and would therefore fall under a rubric of scientific inquiry. Behaviour itself is convergent under natural selection terms (e.g., similar environments produce similar body types and behaviours even in distantly related animals) and some have argued that complex and cooperative social "animals" on other planets would develop similar social behaviours, and then would doubtless be assessed for potential personhood (Bryson, 2025). This again is highly speculative, and the only thing we can know with near certainty is that – if existing – intelligent extraterrestrials will not be members of the human species. *Verdict: No (at present)*

Fictional Characters. Fictional entities, including imaginary friends and deities, lack empirical evidence and are not falsifiable. *Verdict: No*

Mass Groups Acting as Singular Bodies. Mass groups embodying "personhood" overlaps with arguments for the legal personhood of corporations and AI. Such mass groups could include concepts like "nature" as a person (Smith, 2021), "hive" minds (Beni & Wang, 1993), or collective entities like corporations and political groups. Some theorists argue for personhood in collective consciousnesses, where decisions are made *en masse* (Ibid.). Evolutionarily, individual benefits from cooperation drive change, not group selection (Dawkins, 2006). Personhood traits evolved for individual survival and reproduction, not for group-level benefits (e.g., bees benefit genetically from relatives, Grimaldi & Engel, 2005). Thus, mass groups as singular entities can be rejected on evolutionary grounds, as group selectionist arguments are unconvincing (as altruists acting solely for the "good of the group" could not pass on their genes without individual reproductive costs). Because the varying individual motivations within a group undermine the idea of unified group personhood (Tajfel, 1986; Demoulin et al., 2005), mass groups lack Theory of Mind (ToM), consciousness, self-awareness, empathy, deception, and attachment (Masuda, 2012; Shteynberg et al., 2023). *Verdict: No*

This paper's classification supports personhood for humans and some NHAs. Other categories, including AI, speculative extraterrestrials, fictional characters and mass groups, do not currently meet essential personhood qualifiers like genuine self-awareness or agency.

Legal Personhood and AI Performed 'Personhood': Neutral Acts or Societal Implications?

Baudrillard's concept of simulacra highlights the challenge in distinguishing between original and artificial constructs. Benjamin's ideas on reproducibility align with Baudrillard's "hyperreality," where copies can diverge from their originals. In mixed reality and AI, issues of "unfaithful copies" arise, such as AI mimicking emotions, which can have psychological impacts on humans. While mimicry can evoke meaningful experiences, it does not equate to true personhood. For instance, chatbots that simulate empathy lack genuine personhood despite providing emotional support.

Legal and philosophical arguments for AI personhood mirror those for corporate personhood (Sankin, 2013; NPR, 2014). Bestowing personhood upon non-living or nonorganic entities raises the ethical dilemma of privileging or equalising the points of view of a non-living object to be equal in status and resonance to a human being (impossible in itself due to lack of self-awareness). Vigilance against "personhood exploitation" by corporations and multiverse entities is necessary (Bryson, 2017). Corporations have indeed been granted personhood (Marshall, 2023); some have applied the same logic to AI avatars (Nandi, 2024; cf. Novelli, 2023).

That said, those who view corporations as "legal persons" typically understand this as a representation rather than true life, a notion that also applies to AI. Platforms like Epic, Unity, and

Decentraland use avatars and AI for social interaction, but assessing the quality of these interactions is complex, and subjective assessments may lack consistency. Ethical guidelines for virtual characters and the realism of virtual representations are debated, intersecting with issues of intellectual property. Humans can develop empathy toward machines, leading to potential ethical issues like manipulation or emotional dependency; instances of AI mimicking emotions or influencing behavior highlight potential risks, such as a person breaking into Buckingham Palace due to an AI chatbot. AI's "black box" decision-making complicates decision traceability and transparency. Responsibility for AI behavior in mixed reality is therefore complex and should rest with clear corporate accountability rather than the AI or its representations. Specifically, arguments for AI personhood (Mathison, 2023) undermine corporate responsibility for potential AI errors. We further argue that the same arguments against the personhood of mass groups acting as singular bodies (confer Section 3) directly apply to arguments for AI personhood.

Conclusion

Over the past century, discussions on personhood have expanded to include nonhuman animals (NHAs). The debate has also touched on AI. We define personhood as either being anatomically human or possessing the traits detailed in this paper. We accordingly evaluated six entities debated as having personhood: humans, NHAs, AI, speculative extraterrestrials, fictional characters and mass groups, concluding that only humans and some NHAs currently meet (our) criteria for personhood. Addressing issues such as AI mimicry, legal personhood, attachment and responsibility, we have highlighted the significant differences between living beings and AI, stating that applying personhood to AI and corporations is a false equivalency. Unlike NHAs, which are living and self-aware, AI and robots are non-living and programmed by humans.

Proposed Personhood Qualifiers	Human	(Some) NHAs	Current AI	Speculative Extraterrestrials	Fictional Characters	Mass Groups Acting as Singular Bodies
Member of Human Species	X	O	O	O	O	X
Theory of Mind (ToM)	X	X	O	—	O	O
Self-Awareness	X	X	O	—	O	O
Agentic Empathy & Deception	X	X	O	—	O	O
Agentic Attachment	X	X	O	—	O	O
Symbolic Communication with Displacement	X	X	X	—	O	X

Figure P. 4. A visualization of the six personhood types analysed against personhood qualifiers.

The rise of advanced AI and autonomous avatars in mixed reality raises questions about machine personhood, particularly regarding responsibility and mimicry. We strongly argue against AI personhood, given AI lacks agency, consciousness, ToM and attachment. This stance is not gatekeeping, but a recognition of the evolved emotional responses in social animals. We stress the continuing need for a human/NHA-centric approach in complex mimicry systems like mixed reality and VR.

References

- AAMC (Association of American Medical Colleges) (2003). "Personhood" redefined: Animal rights strategy gets at the essence of being human. Association of American Medical Colleges. Accessed 2 February 2016.
- Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences* 21: 547-609
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford, England, UK: Oxford University Press. <https://doi.org/10.1017/S0140525X98001277>
- Beni, G.; Wang, J. (1993). Swarm Intelligence in Cellular Robotic Systems. Proceedings of NATO Advanced Workshop on Robots and Biological Systems, Tuscany, Italy, June 26–30 (1989). Berlin, Heidelberg: Springer. pp. 703–712.
- Bryson, K. (2017). *The Evolving Binary: Perspectives on Infra- and Ultrahumanisation*. PhD Thesis, Department of Anthropology, University College London. London, UK.
- Bryson, K. (2023). Apes unlike us: Evidence of human-ingroup protection against encroaching simianity in UK newspapers. *Society & Animals*.
- Bryson, K. (2025). ET might not look like us, but ET will act like us: A bioanthropological perspective on shifting portrayals of fictive othering. In: David Jeevendrampillai & Aaron Parkhurst (eds), *Extraterrestrial Anthropology* (working title). London, UK: Routledge (In Press.)
- Bubeck, S., et al. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. Cornell University. Accessed 10 May 2024.
- Danilevsky, M., et al (2020). A survey of the state of explainable AI for natural language processing. arXiv preprint arXiv:2010.00711. <https://doi.org/10.48550/arXiv.2010.00711>
- Dawkins, R. (2006). *The Selfish Gene*. Oxford: Oxford University Press.
- Demoulin, S, et al. (2005), Motivation to support a desired conclusion versus motivation to avoid an undesirable conclusion: The case of infra-humanization. *International Journal of Psych.* 40: 416–428. <https://doi.org/10.1080/00207590500184495>
- Erdal, D. & A. Whiten (1996). "Egalitarianism and Machiavellian intelligence in human evolution" in *Modelling the Early Human Mind* (eds. P. Finkelman, P. (2013). How the pro-slavery constitution led to the Civil War. *Rutgers Law Journal* 43 (3): 405 Mellars & K. Gibson). Cambridge, Fox, J. & C. Shulman (2010). "Superintelligence does not imply benevolence" in *ECAP10: VIII European Conference on Computing and Philosophy* (ed. Klaus Mainzer). Munich, Germany: Verlag Dr Hut, pp 1-7. England, UK: Cambridge Macdonald Monograph Series, pp 139–150.
- Hern, A. (2017). Give robots 'personhood' status, EU committee argues: Proposed rules for robots and AI in Europe include a push for a general basic income for humans, and 'human rights' for robots. *Guardian*.
- Ilager, S., et al (2023). Data-centric edge-AI: A symbolic representation use case. *IEEE International Conference on Edge Computing and Communications (EDGE)*. <https://doi.org/10.1109/EDGE60047.2023.00052>
- Kahn, P.H. Jr, B. Friedman & J. Hagman (2002). "I care about him as a pal": Conceptions of robotic pets in online AIBO discussion forums. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. Accessed 7 July 2016.
- Marshall, B. (2023). No legal personhood for AI, *Patterns*, Volume 4, Issue 11.

- Martin, C.F., R. Bhui, P. Bossaerts, T. Matsuzawa & C. Camerer (2014). Chimpanzee choice rates in competitive games match equilibrium game theory predictions. *Scientific Reports* 4/5182.
- Mathison, T.W. (2023). Recognizing Right: The Status of Artificial Intelligence, 19 *J. Bus. & Tech. L.* (2023).
- McQuillan, L. (2022). A Google engineer says AI has become sentient. What does that actually mean?. *CBC News*. Accessed 10 May 2022.
- Nagel, T. (1974). What is it like to be a bat?. *Philosophical Review*: 435-450.
- Nandi, A. (2024). Artificial intelligence and personhood: Interplay of agency and liability. *Expert Speak Digital Frontiers*. Observer Research Foundation.
- Novelli, C. (2023). Legal personhood for the integration of AI systems in the social context: a study hypothesis. *AI & Society* 38 (4) 1347-1359. <https://doi.org/10.1007/s00146-021-01384-w>
- NPR (National Public Radio) (2014). When did companies become people? Excavating the legal evolution. *National Public Radio*. Accessed 6 August 2016.
- Rose, N. (2006). *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*. Princeton, New Jersey, USA: Princeton University Press.
- Rescorla, M. (2015). The Computational Theory of Mind. *Stanford Encyclopedia of Philosophy*. Accessed 28 August 2016. <https://doi.org/10.1017/9781009183734>
- Subagdja, B., Tay, H.Y., & Tan, A.H. (2021). Who am I?: Towards social self-awareness for intelligent agents. *Proceedings of IJCAI 2020*: 4396-4402. <https://doi.org/10.24963/ijcai.2020/606>
- Tajfel, H. & J.C. Turner (1986). "The social identity theory of intergroup behaviour" in *Psychology of Intergroup Relations*. Nelson-Hall, pp 7-24. <https://doi.org/10.4324/9780203505984-16>
- Thorne, S. (2020). Hey Siri, tell me a story: Digital storytelling and AI authorship. *Convergence* 26.4 (2020): 808-823.

A pilot experiment on user individualization for sound localization within interactive virtual environments

Daniel Gonzalez-Toledo¹, Pablo Gutierrez-Parera^{1,2}, Carmen Soto-Romero³,
Maria Cuevas-Rodriguez¹, and Arcadio Reyes-Lecuona¹

¹Telecommunication Research Institute (TELMA), Universidad de Málaga, Málaga, Spain

²Institute of Telecommunications and Multimedia Applications (ITEAM), Universitat Politècnica de València, Spain

³Universidad de Málaga, Spain

Corresponding autor: areyes@uma.es

Keywords: Binaural Rendering, Sound Localisation, Dynamic Listening, Virtual Reality.

Introduction

In the field of psychoacoustics, sound source localization experiments focus on how humans perceive the location of sounds in space. Typically, to solve the binaural localisation problem, humans use the interaural time differences (ITD) and interaural level differences (ILD) to determine the direction of a sound source. In addition, sound is filtered by the pinnae, because their shape affects how sounds reach our ears. This helps to identify the elevation and direction of the sound. This filtering process is represented by the head-related transfer function (HRTF), which characterizes how an individual's ears, head, and torso shape the sound from a particular location in space (Xie, B., 2013). Individualized HRTFs are key for high-fidelity 3D audio, capturing how sound interacts with each person's unique morphology.

Virtual Reality (VR) benefits significantly from advancements in binaural 3D audio, which enhances immersion by mimicking real-world soundscapes (Bormann, 2005). In recent years, 3D audio in VR environments has become an important area of research, leading to the development of various approaches and innovative techniques that enable the integration of spatial audio into VR systems. These advancements have also got significant attention from major industry players, such as Google and Meta, with their respective audio renderers called Resonance Audio, by Google (Gorzel et al., 2019) and Meta Quest Oculus spatialiser SDK (Meta, 2024). Another option used to integrate 3D audio in Virtual Environments (VE) is Steam Audio (Valve, 2024). More oriented to research in psychoacoustics, the 3D Tune-In Toolkit (3DTI) (Cuevas-Rodriguez et al, 2019) and its evolution the Binaural Rendering Toolbox (BRT) (Gonzalez-Toledo et al., 2023), which allow full control on the processing mechanisms. That makes them appropriate to conduct studies on the impact of different rendering approaches in auditory spatial perception. In this study, we use the BRT.

VR is an interactive and dynamic environment where users can move and actively perceive sound sources. While there is extensive research on static sound sources localization, studies focusing on dynamic situations with active listening are still limited (Gaveau et al., 2022; Cuevas-Rodriguez et al., 2023). In dynamic VR scenarios, the accurate localization of sound sources directly impacts the user's sense of presence and immersion. However, common localization errors, such as front-back confusion, elevation misjudgements, and inaccuracies in distance perception, often occur. These errors can disrupt the immersive experience and highlight the need for improved spatial audio rendering techniques that adapt to the dynamic and interactive nature of VR environments.

Dynamic spectral cues are important for sound localization in environments where the sound source or listener is moving, particularly to distinguish between sounds coming from the front and the back (Wightman et al., 1999). Head movements enhance localization accuracy by varying the interaural time and level differences perceived by the listener, which help refine spatial estimates and resolve directional ambiguities. Captured by HRTFs, the spectral changes allow the brain to track moving sound sources accurately. However, some studies claim that dynamic spectral cues do not aid localization when the movements are very small (McLachlan et al., 2023). Their experiments examined which acoustic cues listeners use to localize sounds when making small head movements (less than $\pm 10^\circ$ from the center) revealing that even minor yaw rotations (horizontal movements) significantly reduce the front-back confusion rate, while pitch rotations (vertical movements) had little impact.

This work is the first stage of an ongoing study, in which we are investigating whether larger movements and dynamic spectral cues contribute to more accurate localization. This poster presents the first results of a pilot experiment where we explore some specific movements with two amplitudes. We are using individual HRTFs compared with non-individual HRTFs to study the performance of the listener in dynamic situations for vertical and horizontal movements. Results show that the use of individual HRTFs reduce the front-back confusions (FBC) which provides a better audio localization.

Methods

Ten students from the School of Telecommunication Engineering of the University of Malaga participated in the experiment. All were right-handed and reported having normal hearing. All procedures were reviewed and approved by the Ethical Committee for Research in Malaga (Comité Ético de Experimentación de la Universidad de Málaga).

Participants used a Meta Quest 3 for visuals and Audio Technica ATH-R70x headphones for binaural audio playback with a Focusrite Scarlett 2i2 USB audio interface to deliver the audio stimuli. An application, specifically developed for this experiment using Unity rendered the visual scene, collected tracker data from the Head Mounted Display (HMD), and configured all visual and audio scenes. The virtual auditory scene has been rendered using the Binaural Rendering Tool (BRT) (Gonzalez-Toledo et al., 2023). An HRTF convolution-based simulation was used to generate the direct acoustic path, and room reverberation was not simulated. The BRT communicates with the experiment application via OSC. The sound stimulus was a white noise. Individual HRTFs were measured for all participants using the Earfish system (Reijniers et al., 2023).

The experiment considered an active dynamic listening scenario where participants could move their heads to receive dynamic cues from static sound sources. This makes the situation dynamic due to the listeners' movements, even though the sound sources remain stationary in the space. Listener's movements were guided to follow two types of trajectories: horizontal rotations (yaw movements, around the vertical axis) or vertical rotations (pitch movements, around the interaural axis). The listener range of movements are ± 10 and ± 30 degrees.

Each participant listened to the sound stimuli with synthesized spatiality using their individualized HRTF and another HRTF taken from the most different participant. That is, among all the HRTFs measured with the Earfish system for all participants in the experiment, the one that differs the most from the participant's own measurement was chosen. To study the level of similarity between two HRTFs, we have calculated the Logarithmic Spectral Difference (LSD). The LSD quantifies how much the frequency response of an HRTF deviates from a reference HRTF (Gutierrez-Parera et al., 2022).

During the experiment, participants sat in a chair while wearing the HMD and headphones, holding one of the Meta Quest controllers in their right hand. The HMD displayed an empty space, with a sphere around the participant, where the radius is the distance to the sound sources, 1.5 meters. Virtual sound sources were positioned in 17 predetermined positions relative to the listener's head, resulting from the combination of 8 different azimuths (0° , 45° , 90° , 135° , 180° , 225° , 270° and 315°) and 2 different elevations (0° and 45°) and one last source placed in the north pole of the sphere (0° azimuth and 90° elevation).

The participants' task was to listen to a sound, while performing a specific one-sided rotation around either the yaw or the pitch axis and locate the sound as precisely as possible by pointing to its position with the handheld controller. The experiment was structured in one session, each of them composed of 2 blocks (one for each HRTF) with 68 trials (17 source positions x 4 trajectories), randomly sorted. There was a previous training in which the participant had visual feedback, indicating after each trial the position of the sound source.

Results

To assess the localization performance, errors due to front and back confusions (FBC) have been analyzed. The axis connecting both ears, known as the interaural axis, is used as a reference to calculate these FBC, which are analyzed as azimuth deviations. We obtained the called mirror source by inverting the point selected by the participant (the prediction) with respect to the interaural axis. When the azimuth difference between the source and the prediction is greater than the azimuth difference between source and mirror source, it can be asserted that a front-back confusion has occurred. Then the rate (in %) of FBCs was calculated.

Results have been analysed as a function of the HRTF (individualized and most different) and head movement (pitch and yaw), using a 2-ways ANOVA. The analysis of the front-back confusion rate revealed significant differences for the HRTF condition ($F(1, 8) = 8.7$, $p = 0.018$) and head movement condition ($F(1, 8) = 14.6$, $p = 0.005$).

By observing the charts in Figure P 5, a lower percentage of errors can be seen when using an individualized HRTF. Pitch movements show a reduced error rate with individualized HRTFs, which indicates that the spectral cues of the individualized HRTF help in the source localization task. Average values suggest that horizontal head trajectories (yaw movements) are more effective in reducing front-back confusions.

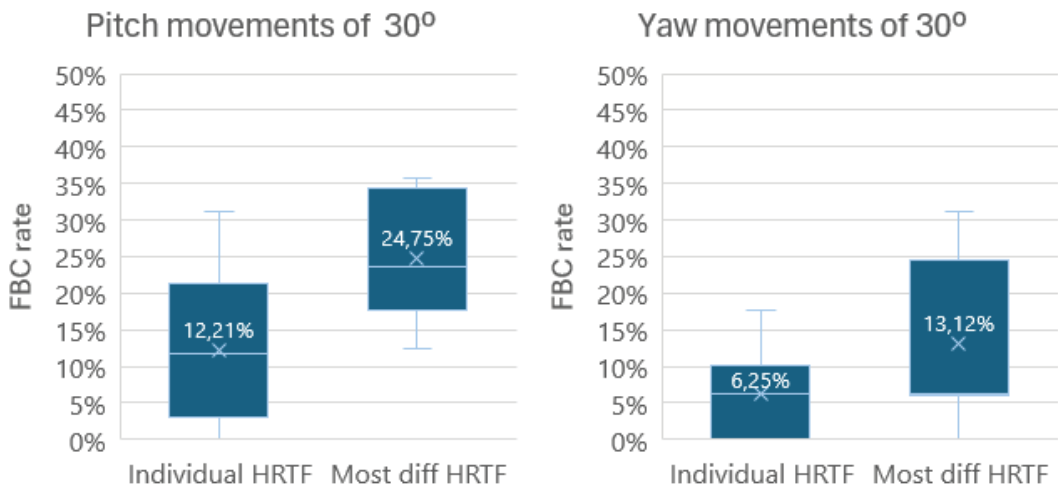


Figure P. 5. Front-back confusion rate for yaw and pith movements.

Discussion

This study is a pilot experiment to examine the impact of head movements and HRTF individualization on sound localization accuracy in virtual environments. The performed experiment included small movements (10°) where we found that spectral cues have much less influence than binaural cues on dynamic localisation, in line with the results of (McLachlan et al., 2023). However, the results also show that individualized HRTFs significantly reduce front-back confusion for larger movements (30°), for both horizontal (yaw) and vertical (pitch) head movements, highlighting the importance of spectral and dynamic cues in enhancing spatial audio perception. Individualized HRTFs capture the listener's unique spectral filtering effects, allowing better perception of the spatial sound information and improving the immersive auditory experience in VR settings.

The study found that yaw rotations significantly reduced FBC rates compared to pitch rotations. This is understandable given that yaw movements generate varying ITD and ILD, which are strong perceptual cues for resolving front-back ambiguities. But it is interesting to detect that the use of the own HRTF (individual spectral cues) introduces a significant improvement also for yaw movements.

Despite the promising findings, this pilot study has limitations, including a small sample size and the absence of fully dynamic scenarios. In addition, future studies should include static localization for comparison, as we need to clarify if the advantage we have found when using individual HRTFs is due to static perception or there are interactions with the dynamic nature of the test. Finally, we should allow more listener head trajectories and explore situations where both the listener and sound sources are in motion.

Acknowledgements

This work has been partially supported by the H2020 project SONICOM (grant 101017743), the Spanish National Project SONIX (grant PID2023-152547NB-I00), and the postdoctoral program APOSTD (grant CIAPOS/2021/132) by the Generalitat Valenciana and European Social Fund.

References

- Bormann, K. (2005). Presence and the utility of audio spatialization. *Presence: Teleoperators and Virtual Environments*, 14(3), 278–297. <https://doi.org/10.1162/105474605323384645>
- Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., & Reyes-Lecuona, A. (2019). 3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation. *PLoS one*, 14(3), e0211899. <https://doi.org/10.1371/journal.pone.0211899>
- Cuevas-Rodriguez, M., González-Toledo, D., Gutierrez-Parera, P., & Reyes-Lecuona, A. (2023). An experiment replication on dynamic 3D sound localisation using auditory virtual reality. *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*. Turin, Italy.
- Gaveau, V., Coudert, A., Salemmé, R., Koun, E., Desoche, C., Truy, E., ... & Pavani, F. (2022). Benefits of active listening during 3D sound localization. *Experimental Brain Research*, 240(11), 2817–2833. <https://doi.org/10.1007/s00221-022-06456-x>
- González-Toledo, D., Molina-Tanco, L., Cuevas-Rodríguez, M., Majdak, P., & Reyes-Lecuona, A. (2023). The Binaural Rendering Toolbox. A virtual laboratory for reproducible research in psychoacoustics. *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*. Turin, Italy.
- Gorzal, M., Allen, A., Kelly, I., Kammerl, J., Gungormusler, A., Yeh, H., & Boland, F. (2019, March). Efficient encoding and decoding of binaural sound with resonance audio. In *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society.
- Gutierrez-Parera, P., Lopez, J. J., Mora-Merchan, J. M., & Larios, D. F. (2022). Interaural time difference individualization in HRTF by scaling through anthropometric parameters. *Eurasip Journal On Audio, Speech, And Music Processing*, 2022(1), 9. <https://doi.org/10.1186/s13636-022-00241-y>
- McLachlan, G., Majdak, P., Reijnen, J., Mihocic, M., & Peremans, H. (2023). "Dynamic spectral cues do not affect human sound localization during small head movements". *Frontiers in Neuroscience*, 17, 1027827.
- Meta Quest Oculus Spatializer, "Oculus Audio SDK." Accessed: Sep. 06, 2024, <https://developer.oculus.com/documentation/native/audio-spatializer-features/>
- Reijnen, J., Peremans, H., Bart Partoens, B., and Fons de Mey. Earfish. <https://www.earfish.eu/>, 2023. Accessed: 2024-07-24
- Valve_Software, "Steam Audio SDK." Accessed: Sep. 06, 2024. <https://valvesoftware.github.io/steam-audio/>
- Wightman, F. L., and Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.* 105, 2841–2853. <https://doi.org/10.1121/1.426899>
- Xie, B. (2013). *Head-related transfer function and virtual auditory display*. J. Ross Publishing.

Evaluation of content for the integration and education of virtual and augmented reality in the field of additive manufacturing

Torsten Döhler, Moritz Heinrich, Christopher Kracht, Klarissa Koppusch, Fabian Grund, Frank Briesemeister, Eckart Wolf, Andreas Foitzik, Rene Krenz-Baath, and Andrea Böhme

Wildau University of Applied Sciences, Wildau, Germany

Corresponding author: tdoehler@th-wildau.de

Keywords: Virtual Reality, Augmented Reality, Additive Manufacturing, Teaching, Inclusion, Digitalisation, STEM Education

Introduction

The combination of virtual reality (VR) and augmented reality (AR) with additive manufacturing technologies opens innovative opportunities for an essential reform of teaching in technical and engineering disciplines. The goal of the project is to utilize XR technologies to facilitate the transfer of knowledge to heterogeneous groups of people in the technical, STEM, and high technology fields. The target groups include students from heterogeneous backgrounds, dual students and employees of companies as well as employees of the skilled trades in Germany, capital region Berlin/Brandenburg. The aim is to make complex STEM theories and practices accessible to these target groups through new, contemporary, innovative educational concepts in conjunction with content from additive manufacturing, especially metal 3D printing. The principle of clarity is prioritized over abstraction.

Methodology

In the period from June to the end of July 2024, a comprehensive survey was conducted among students, employees and external experts in the environment of the Wildau University of Applied Sciences in order to obtain information related to advertising, approaches or perspectives for the "MINT'oVation" project, for example, depending on age, gender and education. A total of 146 valid responses were analyzed, which provided information on the experiences with additive manufacturing and VR/AR, as well as on the needs and interests of the participants and possible advantages and disadvantages of various questions. The project

aims to use mainly experience-based methods such as animation-supported VR, AR and haptics in combination with additive manufacturing using 3D printing and other microscopic processes as well as analytics and image processing, thus benefiting training in general as well as the increasing shortage of skilled labor.

Results

The results of the survey suggest that the majority of respondents are very interested in using VR/AR in combination with the visualization of additive manufacturing technologies in teaching. Optical visualization of digital 3D additive manufacturing processing was identified as a key benefit by 75% of respondents, followed by improved learning and training opportunities (68.1%) and interactive design processes (43.1%). The most important, for example, desired functionalities in a VR/AR application include 3D model viewing (71.4%), simulation of manufacturing processes (69.5%) and training and tutorials (55.2%). Further correlations in the overall context were identified and analyzed (Wildau et al., 2024).

Challenges

The biggest challenges to integrating VR/AR into additive manufacturing, according to respondents, are the comparatively high costs (58.1%), high technical complexity (52.4%) and access to suitable and compatible hardware (48.6%). These and other barriers must be overcome in order to ensure successful implementation and broad acceptance.

The results of the survey can be summarized as follows

The results of the survey demonstrate the significant potential and interest in the integration of VR/AR into additive manufacturing in the higher education sector, especially in a large state like Brandenburg in Germany. In order to realize the full potential, further development and investment in user-friendly and cost-effective VR/AR technologies and combinatorial approaches are required. These technologies could significantly improve the teaching and understanding of complex manufacturing processes and thus contribute to the training of qualified specialists especially to educate and encourage those targeted groups. In addition, regional development in Brandenburg will be supported by the promotion of local labor markets and the identification of participants with the region.

Specific objectives of the entire "MINT`oVation" project

The specific objectives of the entire "MINT`oVation" project are to promote interactive and integrative learning experiences through the use of augmented reality (AR) and virtual reality (VR). The visualization of complex concepts, such as 3D metal printing, enables all learners to be involved in interactive and integrative learning. The immersive experience made possible by VR can promote memory (Schweiger et al., 2022) by ensuring that what has been learnt is retained for longer and understood in greater depth than without the use of VR technology, systematically shown and known as the "CA-MIL", cognitive affective model of immersive learning model in Figure P 6.

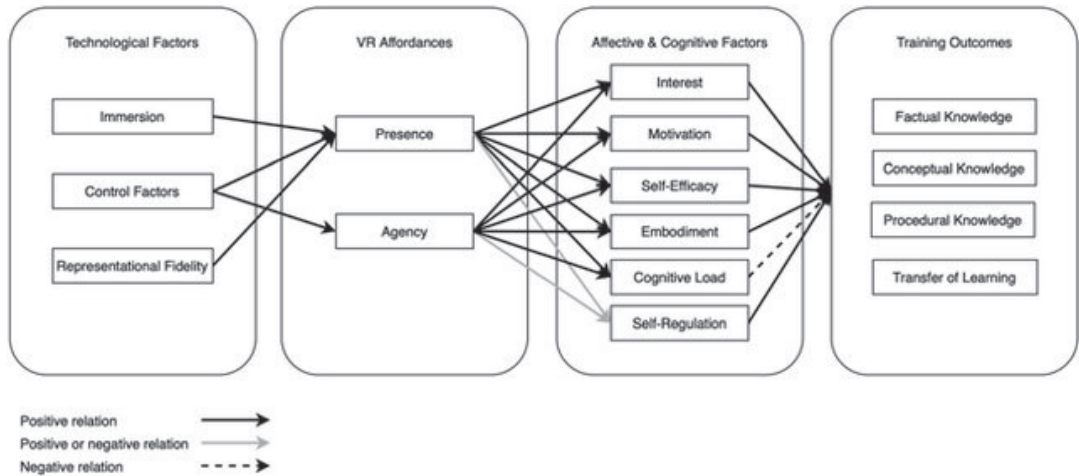


Figure P. 6. Graphical representation of the cognitive affective model of immersive learning (CA-MIL) by Eric Richter [Virtual Reality in Teacher Education from 2010 to 2020]. (Note: Adapted from the "Fig. 2 Overview of the CAMIL" (Makransky and Petersen 2021))

One example of the use of virtual reality technologies in academic teaching is London South Bank University (LSBU), where stereoscopic simulations are used to improve engineering students' understanding of aerodynamics^{13, 14}. The approach described here makes the learning process worthwhile and the learning content accessible. At the same time, participants are given access to complex technical knowledge in a simple and appealing way. The engineering company Renishaw offers workshops in which children and young people can transform their ideas into physical products using 3D printing¹⁵. These experiences motivate young learners to take an interest and apply for engineering and other STEM subjects. The LSBU lab is equipped with high-end 3D printers, a CAD suite and VR/AR headsets, allowing students to experience objects at previously impossible scales and develop a deeper understanding of complex engineering principles. The accessibility and interactivity of VR increases student (AR and VR, 2024) engagement more effectively than traditional teaching methods. The technologies mentioned allow a focus on various aspects of engineering, which deepens the understanding of this discipline. The concept of "blended learning" (BL) aims to make education more inclusive and more equal in terms of opportunities. The combination of online and face-to-face offerings in blended learning programs allows for a differentiated design of learning content, learning times and teaching methods, especially for people with family obligations from industry and trade. The individualized and unconditional approach to learners is a fundamental aspect of the didactic concept. A study shows that the use of 3D printing technologies in higher education promotes student motivation and engagement. This is exemplified by the "Physics

¹³ <https://www.lsbu.ac.uk/about-us/news/lsbu-virtual-engineering>, 2024

¹⁴ <https://www.imeche.org/news/news-article/into-the-third-dimension-transforming-education-has-never-felt-so-real>, 2024

¹⁵ <https://www.renishaw.com/en/renishaws-stem-education-programme-goes-virtual—45661>, 2024

of Materials" course Fernandes & Simoes, 2019). Conducting demonstrations on various 3D printing materials and other examples has proven to be an effective learning aid. Teachers can thus encourage student engagement and enrich the learning process. In the discussion of the BL concept, the low physical presence of the participants and the need for increased basic technical knowledge is considered a disadvantage. Nevertheless, the experience already gained with BL can be directly integrated as a basis for the development of new AR and VR learning environments. Supplemented learning (SL) with accessible content, such as 3D printing, represents innovations in the education sector. The annotation of recordings, for example with subtitles and audio descriptions, improves accessibility for people with various learning needs, limitations and disabilities. Additive manufacturing promotes competencies and problem-solving skills through 3D modelling, tactile learning and enables adaptable educational models, including for visually or otherwise impaired students. Stone et al., 2020.

Summary

In higher education institutions, the use of 3D printed designs as open resources leads to easier accessibility to technology and product as well as the possibility of continuous improvement through iteration. Additive manufacturing is particularly suitable for the implementation of inclusive learning strategies, for example in interdisciplinary service-learning courses. It also enables the use of augmented reality (AR) and virtual reality (VR) in educational contexts. The concept promotes creativity and inclusion in educational processes and supports tactile learning for students with disabilities. It therefore makes a significant contribution to the development of skills among teachers and students.

Acknowledgements:

Supported by funds from the European Union and the State of Brandenburg.

References

- Fernandes, S., & Simoes, R. (2019). Using 3D Printing as a Strategy for Including Different Student Learning Styles in the Classroom. <https://doi.org/10.4018/978-1-5225-7018-9.ch009>
- Stone, B., Kay, D., Reynolds, A., & Brown, D. (2020). 3D Printing and Service Learning: Accessible Open Educational Resources for Students with Visual Impairment. *International Journal of Teaching and Learning in Higher Education*, 32(2), 336-346.
- Schweiger, et al. (2022, May) Learning success at school through augmented and virtual reality? A quantitative synopsis of impact studies on the use of virtual reality in primary and secondary schools, University of Augsburg.
- Wildau, T.H., Döhler, T., Heinrich, M. et. al. (18.7.2024) Survey "Results report questionnaire", Virtual and augmented reality using the example of additive manufacturing in teaching, survey period: 24.06.24 - 30.07.24,

Towards User-Centered XR Solutions for Smart Agriculture: A Work-in-Progress Design Approach

Alexandra Plexousaki, Asterios Leonidis, Maria Korozi,
and Constantine Stephanidis

Institute of Computer Science (ICS), Foundation for Research and Technology - Hellas (FORTH), Greece

Corresponding author: aplex@ics.forth.gr

Keywords: Extended Reality (XR), Agriculture, Digital Twins, Internet of Things

Introduction

As agriculture faces increasing pressures from climate change, there is a growing need for innovative technologies to enhance productivity, resilience, and sustainability. In the last few years, the agricultural domain has slowly started to adopt modern technologies in its practices, resulting in the emergence of more domain-specific research topics such as Agriculture 5.0, referring to a sustainable agricultural production system (Fountas et al., 2024), or Smart Farming Verdouw et al., 2021). Immersive technologies are emerging as transformative tools in this domain, offering new ways to optimize agricultural practices. However, despite their growing popularity, Extended Reality (XR) technologies remain relatively nascent in agriculture, livestock farming, and aquaculture. There is a clear need for future research that integrates XR with complementary technologies like IoT, robotics, teleoperation, and advanced interaction controllers to unlock its full potential Anastasiou et al., 2023). Similarly, Digital Twin (DT) technologies mark a new phase in smart farming, where their application as a central management tool decouples physical processes from planning and control, thereby enhancing operational efficiency and data-driven decision-making Verdouw et al., 2021).

This research aims to bridge the gap between advanced immersive technologies and practical agricultural applications. As immersive devices with enhanced computational power continue to evolve, they offer a unique opportunity to transform promising concepts into scalable, real-world solutions for the agricultural sector. Despite the abundance of innovative ideas, many remain stuck at the proof-of-concept stage, without achieving widespread adoption or impact. Combining XR technologies and digital twins to create realistic 3D representations of agricultural environments offers farmers an immersive experience, allowing them to engage with and manage systems in a more intuitive and efficient manner [9]. In this work, we aim to explore transforming theoretical models into practical solutions that drive efficiency and innovation in agricultural practices.

Related Work

A growing body of research highlights the potential of technologies such as augmented reality (AR), virtual reality (VR), and the Internet of Things (IoT) to enhance decision-making, resource management, and operational efficiency in agriculture. The extensive study in (Anastasiu et al., 2023), explores how XR can influence key production areas such as decision-making and operational activities in these sectors, proving that XR technology offers significant benefits for education, training, and system development, enhancing human capabilities and experiences across these domains.

Practices in Augmented Reality (AR) for agriculture share a common focus on enhancing decision-making and user guidance in farming activities. These approaches utilize AR to provide real-time visual aids and guidance, helping users make informed decisions during tasks such as crop management. Whether it's assisting gardeners by overlaying virtual instructions on actual fields (Okayama and Miyawaki, 2013) (Figure P 7b), guiding farmers to specific areas for soil sampling (Huuskonen and Oksanen, 2018), or offering in-situ visualization of field conditions (Zheng and Campbell, 2019), the goal is to streamline complex agricultural processes. Recent studies highlight the use of simulation and visualization technologies to provide immersive and realistic experiences. A common practice is simulating specific conditions in Virtual Reality (VR) environments to educate farmers about seasonal issues or environmental conditions. For example, the authors in Kim et al. (2019 and 2021) simulate aerodynamic conditions inside a piglet house to examine the effects of various airflow or ventilation settings on livestock (Figure P 7a).

Forecasting and prediction play a crucial role in modern agriculture by enabling proactive management and improving decision-making processes. Recent research emphasizes the use of visualization techniques to provide farmers with interactive tools that offer insights into future conditions. These technologies allow for better monitoring (Xi et al., 2019) and management of agricultural environments, enhancing efficiency and responsiveness across various farming operations. For instance, the authors in Rahman et al. (2021) use machine learning models to predict key water quality parameters like dissolved oxygen, pH, and temperature. These 24-hour forecasts (Figure P. 7) help farmers shift from reactive to proactive management, allowing them to anticipate poor conditions and take action to prevent issues

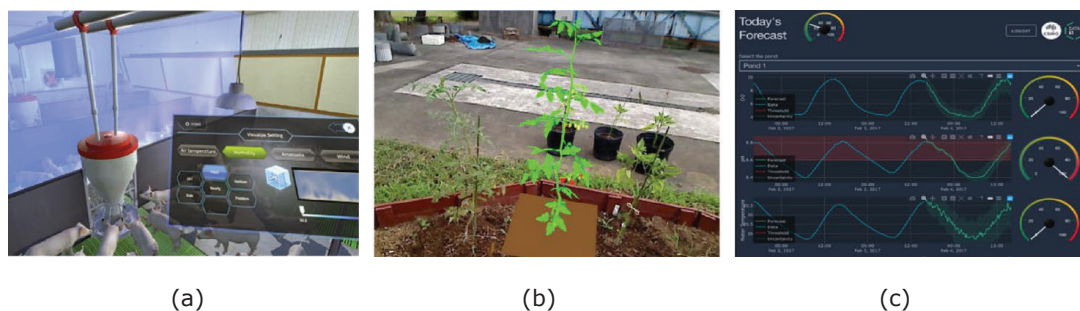


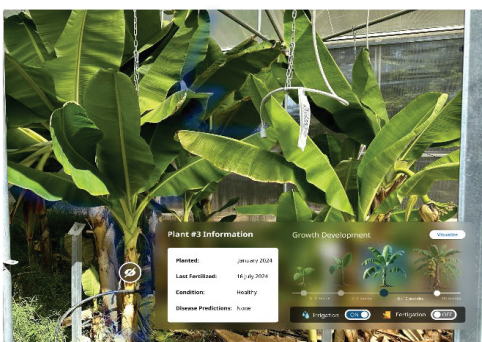
Figure P. 7. (a) Aerodynamic condition simulation in VR [4], (b) Plant growth visualization in AR [5], (c) Water quality forecast dashboard [7]

Discussion

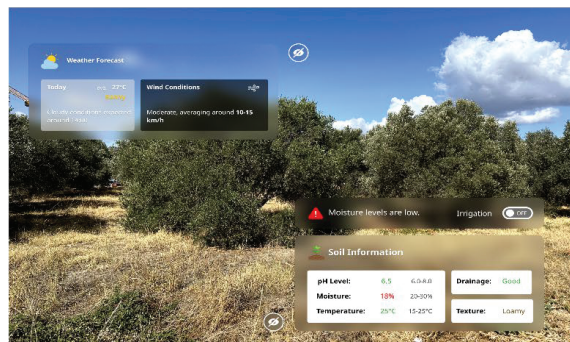
Considering the challenges highlighted by recent research trends in the field of XR in Agriculture, the most prevalent solutions can be organized into seven (7) categories based on their core functionalities:

1. **Context-sensitive information visualization** provides dynamic insights tailored to specific contexts, such as animal identification in a field.
2. **Data visualization** is primarily used for presenting animal and environmental data through dashboards or contour plots, to monitor, for example, crop health.
3. In the realm of **data collection**, much of the research has concentrated on supporting in-situ data acquisition and real-time monitoring of animal behaviour, including movement and feeding patterns.
4. The **forecasting, prediction visualization, and insights** category emphasizes real-time visualization of environmental fluctuations, and predictive analytics in agricultural practices.
5. Offering **guidance and support** in tasks such as planting, fertilizing, and field management is crucial for enhancing precision and operational efficiency in agriculture.
6. **Simulation** plays a vital role in modeling environmental control systems or plant growth, delivering an immersive and interactive experience.
7. Lastly, **digital twins** are instrumental in generating actionable data and insights that can optimize agricultural practices, for example in greenhouses or farming processes.

In response to the growing need for technologically advanced solutions in agriculture and addressing the challenges outlined in current research, we initiated the design of an XR tool aimed at leveraging emerging XR technologies to support agricultural practices in a user-centered way. By following the design process proposed by the authors in Stephanidis et al. (2024), we developed a tool focused on enhancing user experience during the monitoring of agricultural environments, specifically using AR devices in outdoor settings. Figure P 8 presents an in-progress view of our tool being utilized with an AR device in the field. Our approach seeks to improve decision-making precision and assist in predicting and preventing undesired situations, thereby offering a more efficient means of managing agricultural tasks.



(a)



(b)

Figure P. 8. View of our XR tool both in an indoor and outdoor setting. (a) Plant monitoring inside a greenhouse, (b) Information viewing in an outdoor olive tree field

Acknowledgements

This work is supported by the FORTH-ICS internal RTD Program 'Ambient Intelligence and Smart Environments'. Moreover, the authors would like to thank Antonis Katzourakis for his assistance in preparing the poster artwork.

References

- Anastasiou, E., Balafoutis, A. T., & Fountas, S. (2023). Applications of extended reality (XR) in agriculture, livestock farming, and aquaculture: A review. *Smart Agricultural Technology*, 3, 100105. <https://doi.org/10.1016/j.atech.2022.100105>.
- Fountas, S., Espejo-García, B., Kasimati, A., Gemtou, M., Panoutsopoulos, H., & Anastasiou, E. (2024). Agriculture 5.0: Cutting-edge technologies, trends, and challenges. *IT Professional*, 26(1), 40-47. <https://doi.org/10.1109/MITP.2024.3358972>.
- Huuskonen, J., & Oksanen, T. (2018). Soil sampling with drones and augmented reality in precision agriculture. *Computers and electronics in agriculture*, 154, 25-35. <https://doi.org/10.1016/j.compag.2018.08.039>.
- Kim, R. W., Kim, J. G., Lee, I. B., Yeo, U. H., & Lee, S. Y. (2019). Development of a VR simulator for educating CFD-computed internal environment of piglet house. *biosystems engineering*, 188, 243-264. <https://doi.org/10.1016/j.biosystemseng.2019.10.024>.
- Kim, R. W., Kim, J. G., Lee, I. B., Yeo, U. H., Lee, S. Y., & Decano-Valentin, C. (2021). Development of three-dimensional visualisation technology of the aerodynamic environment in a greenhouse using CFD and VR technology, Part 2: Development of an educational VR simulator. *Biosystems Engineering*, 207, 12-32., <https://doi.org/10.1016/j.biosystemseng.2021.02.018>.
- Okayama, T., & Miyawaki, K. (2013). The "Smart Garden" system using augmented reality. *IFAC Proceedings Volumes*, 46(4), 307-310. <https://doi.org/10.3182/20130327-3-JP-3017.00070>.
- Rahman, A., Xi, M., Dabrowski, J. J., McCulloch, J., Arnold, S., Rana, M., ... & Adcock, M. (2021). An integrated framework of sensing, machine learning, and augmented reality for aquaculture prawn farm management. *Aquacultural Engineering*, 95, 102192. <https://doi.org/10.1016/j.aquaeng.2021.102192>.
- Slob, N., Hurst, W., Van de Zedde, R., & Tekinerdogan, B. (2023). Virtual reality-based digital twins for greenhouses: A focus on human interaction. *Computers and Electronics in Agriculture*, 208, 107815. <https://doi.org/10.1016/j.compag.2023.107815>.
- Stephanidis, C., Leonidis, A., Korozi, M., Kouroumalis, V., Adami, I., & Ntoa, S. Design for Intelligent Environments. In *Human-Computer Interaction in Intelligent Environments* (pp. 1-43). CRC Press.
- Verdouw, C., Tekinerdogan, B., Beulens, A., & Wolfert, S. (2021). Digital twins in smart farming. *Agricultural Systems*, 189, 103046. <https://doi.org/10.1016/j.agsy.2020.103046>.
- Xi, M., Adcock, M., & Mcculloch, J. (2019, November). An end-to-end augmented reality solution to support aquaculture farmers with data collection, storage, and analysis. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry* (pp. 1-2). <https://doi.org/10.1145/3359997.3365721>.
- Zheng, M., & Campbell, A. G. (2019, October). Location-based augmented reality in-situ visualization applied for agricultural fieldwork navigation. In *2019 IEEE Int. Symp Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (pp. 93-97). <https://doi.org/10.1109/ISMAR-Adjunct.2019.00039>.

AR-enabled digital twin framework for automated shotcrete planning and operation

Rehan Khan, Christian Martinez, Mohith Sunkara, and Rahul Tomar

DigitalTwin Technology GmbH, Cologne, Germany

Corresponding author: rehan.khan@digitaltwin.technology

Keywords: Digital Twin, Shotcreting, BIM/CIM, AR, IoT Sensors.

Introduction

This poster presents a digital twin (DT) framework that will realize a construction Digital Twin environment through a visualization and real-time control dashboard using the research results and developed technologies from integrating other digital tools in the ongoing research project. The project aims towards Construction 4.0 by automating particularly laborious construction tasks in all phases of shotcrete application. The DT and simulation environment will allow lifelong monitoring of the constructed/repared site from start to finish with smart, low-energy internet-of-things (IoT) sensors, construction process diagnostics & data visualization through tools such as dashboard, 4D planning, shotcrete simulation and robot accessibility.

Digital Twin Framework

In the first step to creating a DT, the digital representation of a site to be shotcreted is done through building/construction information models (BIM/CIM) and point cloud scans. This is done through an adaptive modeling methodology and a robot construction information model (rCIM) data model to facilitate interoperability between different systems of the digital twin. *“Adaptive model” would adequately include all relevant DT representations that can be changed through a modular approach as per the desired DT requirements for simulation and visualization of the construction process (shotcreting) based on an end-user centered approach.* The IFC based data model includes the necessary classes and metadata linked to the geometric BIM model to be used by the IRR (Inspection-Reconnaissance mobile manipulator) and SFR (Shotcrete and Finishing mobile manipulator) robots during shotcreting operation through the different phases. This geometric and information model integrates into the DT to create an accurate geometrical representation and the associated metadata of a construction site.

The **robotic simulation environment** provides the simulations of the robotic systems allowing for the interactive 3D simulation of the actors at the construction site, including kinematic, dynamic and sensor simulation. Furthermore, the simulation environment simulates the shotcreting process (adhesion, cohesion & rebound effects), which is sent to the DT platform through heat maps. In the DT platform,

these heat maps are converted to 3D meshes that are overlaid on the existing BIM model to twin the actual construction site conditions.

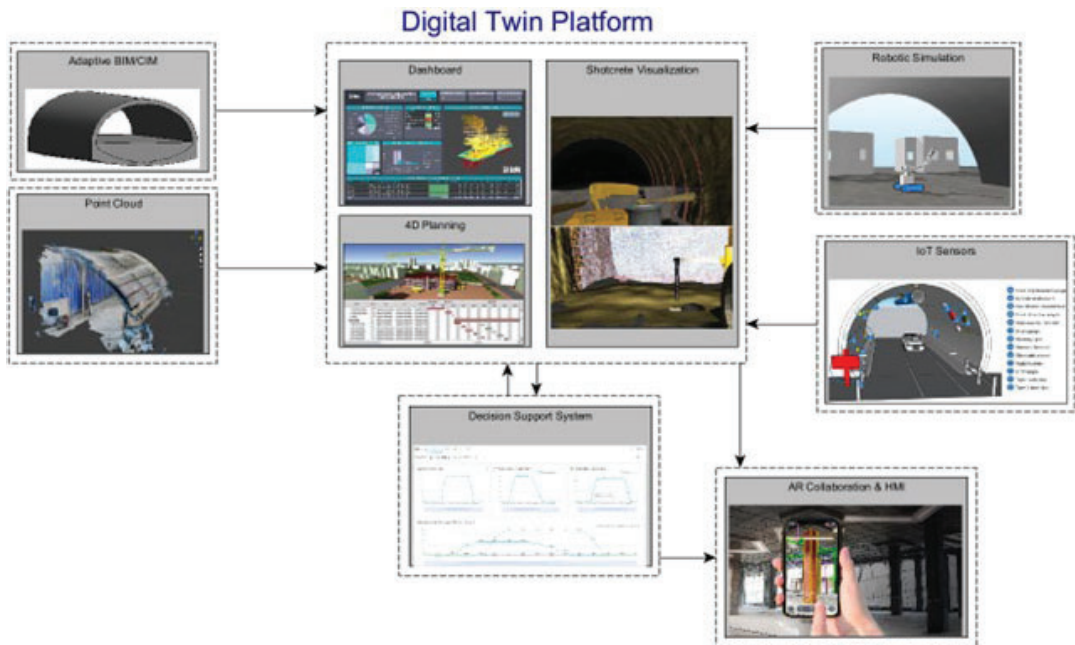


Figure P. 9. DT framework for automated shotcrete planning and operation.

The **IoT sensors** will be implanted during the preparatory phase to allow for the Structural Health Monitoring of the construction site. Low energy consumption sensors measuring temperature, humidity, corrosion, vibration, etc., are integrated into the DT platform through a Neo4j database to link the sensors modeled in the BIM model to their actual physical location on site. The DT platform-IoT data integration allows the user to visualize historical and current time series data alongside the alerts based on sensed data when the threshold of a particular parameter is surpassed by visually highlighting the sensor or concerned area in the digital model.

The **decision support system (DSS)** allows different actors in the domain to analyze data in a holistic fashion. An essential element of the DSS is to serve the needs of different types of users based on the dynamic analysis of heterogeneous historic and streaming data. The DSS provides data to the DT platform through an application programming interface (API) integration to visualize the relevant data in the digital replica to support decision making.

For the **human machine interface (HMI)**, an AR-enabled application will be developed to increase the operator's situation awareness during construction activities. It will provide notification, possible annotations on the constructed area, and inferences regarding the next robot-human action. Additionally, the DT platform provides data (decisions and BIM model) on the mobile device through AR.

The **DT platform** aims to enable visualization, simulation and a real-time control dashboard during the planning and operation phases of the shotcrete application through AR integration.

AR integration application

BIM Integration

AR systems will be integrated with BIM models, allowing real-time access to project data, structural models, and tolerances. Workers and engineers can overlay BIM data onto the construction site through AR glasses or MR headsets, helping them ensure that the shotcreting follows the intended design specifications. Robotic shotcreting could adopt similar AR overlays for precision, where the system would guide the robot based on BIM data, ensuring material is applied as designed.

Digital Twin Bi-directional Control

For any DT system, bi-directional control/communication between the physical asset and the digital replica is essential. AR systems will collect data from sensors, cameras, and the robotic machinery during the shotcreting process, providing feedback to the DT for optimization. This data can be used to refine future operations and improve shotcreting performance.

Conclusion and future works

We present an AR-enabled digital twin framework for automated shotcrete planning and operation. Our DT platform provides a visualizable dashboard integrated with 4D planning, 3D shotcrete visualization, decision support system, AR collaboration and HMI solutions. The DT framework will be tested on diverse demo sites - ground support walls, piles & beams, posttensioned bridge box and tunnels - to validate the automated solution in the shotcrete application domain. Furthermore, we will explore the applicability of our DT framework in both the construction of new infrastructure and the repair of existing infrastructure.

Acknowledgements

This work has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101058731.

References

- Khan, R., Tomar, R. & Ibrahim, A. (2023). Adaptive BIM/CIM for Digital Twinning of Automated Shotcreting Process. In: International Conference on Construction Logistics, Equipment, and Robotics. Springer, pp. 59-68. https://doi.org/10.1007/978-3-031-44021-2_7
- Samadi, M., Waspe, R. & Schlette, C. (2024). Towards shotcrete process simulation to support robotic operation. In: Proceedings of the 41st ISARC. I.A.A.R.C, pp. 707-713. <https://doi.org/10.22260/ISARC2024/0092>

Enhancing XR Accessibility: Automated Lip Reading and Captioning with Audio-Visual Speech Recognition Integration in Augmented Reality

Hrishikesh Mulay¹, Sam Redfern², and Eleni Mangina¹

¹University College Dublin, Dublin, Ireland, ²University of Galway, Galway, Ireland²

Corresponding author: hrishikesh.mulay@ucdconnect.ie

Keywords: Augmented Reality, XR Accessibility, Computer Vision, Artificial Intelligence, Machine Learning.

Abstract

Immersive technologies are poised to revolutionize education and professional training, offering interactive experiences and accelerated learning opportunities. Despite their potential benefits, accessibility remains a critical concern. Research suggests that the use of captions in eXtended Reality (XR) environments improves the accessibility and control for users with variety of backgrounds. Automated Lip Reading (ALR), integrated into Audio-Visual Speech Recognition (AVSR) systems, presents a promising solution to enhance accessibility in XR environments, particularly in demanding conditions such as noisy settings. However, its implementation in XR has not yet been contemplated.

This work in progress explores the development of an Augmented Reality (AR) architecture leveraging a cross-modal AVSR system supported by a Large Language Model (LLM) for real-time captioning. A Unity-based prototype integrates Azure AI Speech service with the Visual-HuBERT ALR architecture, addressing latency issues inherent in Python-dependent deep learning development. By employing a multi-platform approach using Unity and Python with User Datagram Protocol (UDP) communication, significant latency reductions have been achieved.

Preliminary results demonstrate a 75% reduction in processing time compared to Python-only implementations, highlighting the feasibility of real-time ALR implementation in XR. Future efforts will focus on enhancing noise robustness through LLM-based multimodal fusion and deploying the architecture on HoloLens 2 for live captioning and retrospective conversation analysis in noisy environments. This research aims to make XR environments more accessible and comprehensible, particularly benefiting users with hearing impairments.

Introduction

It is believed that immersive technologies will have a significant impact on the education industry in coming years (Fact.MR, 2021). Interactive classroom experience with activity-based learning assistance is a key attraction of XR in education (Dick, 2021). In professional sectors, many employers have employed XR technologies for workplace training and lifelong learning purposes. According to a 2020

study conducted by the professional services company PwC, training employees in Virtual Reality is four times faster than in the classroom (Eckert and Mower, 2020). As XR technology revolutionizes workplaces and schools, limited case studies address the accessibility of XR for individuals with disabilities. Approximately 15% of the global population lives with some form of disability (The World Bank, 2023). Thus, the increasing influence of XR in education has instigated an urgency to explore avenues to make XR environments more accessible for individuals with disabilities.

Computer Vision researchers have explored Automated Lip Reading (ALR) applications in diverse fields, including education, welfare for Deaf and Hard of Hearing (DHH) individuals, security and speech recognition enhancement. Cutting-edge ALR algorithms consistently outperform human lip-reading performance (Shi et al., 2022; Assell et al, 2016). Therefore, in audio-visual speech recognition (AVSR) systems, ALR is used to complement audio-based speech recognition to boost its accuracy, making it more reliable in noisy environments (Anwar et al., 2023). Despite AVSR's noise robustness and potential in improving XR accessibility through captioning, its implementation in XR has not been contemplated yet.

In this work in progress, we discuss our progress in developing an Augmented Reality (AR) based architecture, which will employ a cross-modal AVSR system with Large Language Model support for sentence correction to display captions in the user's field of vision. So far, we have developed a Unity-based prototype that implements Azure AI Speech service along with a state-of-the-art ALR architecture titled Visual-HuBERT (Shi et al., 2022) within an AR environment.

Methodology

Audio-Visual Hidden Unit BERT (AV-HuBERT) is a state-of-the-art self-supervised framework for understanding speech that learns by both seeing and hearing people speak (Shi et al., 2022). Its visual-only architecture is titled Visual-HuBERT. Generally, an Automated Lip Reading (ALR) system consists of three stages: Lip Localization/ROI Extraction, Feature Extraction, and Classification. State-of-the-art deep learning ALR and audio-visual speech recognition (AVSR) architectures, such as HuBERT architectures, rely heavily on Python tools and libraries to perform these steps. A key challenge in implementing these architectures in Augmented Reality (AR) is that most AR development is conducted using frameworks like Unity, which do not natively support Python. This limitation contributes to latency issues due to the critical dependence of algorithms on various Python libraries, which in some cases lack equivalents in languages supported by AR frameworks, such as C# (Ott et al., 2019). Additionally, commonly used libraries, such as OpenCVSharp (a C# wrapper for OpenCV), are not supported by Unity, necessitating the exploration of Unity-compatible OpenCV solutions. Thus, we opted for a multi-platform solution using Unity and Python with User Datagram Protocol (UDP) communication to achieve enhanced control over the quality and speed of the system. Significant development effort was put into creating custom implementations to migrate the pre-processing steps to Unity.

To address these issues, firstly, an AR prototype in Unity with integrated speech recognition capabilities was developed. Despite AVSR algorithms like AV-HuBERT offering both visual and audio functionalities, their dependence on Python libraries necessitated an alternative approach. A cross-modal solution was implemented, involving the use of Microsoft Azure Speech SDK in parallel with Visual-HuBERT architecture. Azure Speech SDK integrates seamlessly with Unity and demonstrates reported Word Error Rates as low as 5% at a latency of 500 milliseconds to 1 second. Furthermore, we implemented the lip localization steps in live within Unity using OpenCV for Unity¹ framework. This feature records the region around lips in real time from the camera feed, which is then passed to the

base model using UDP sockets for feature extraction and classification, and a final predicted output is received within Unity.

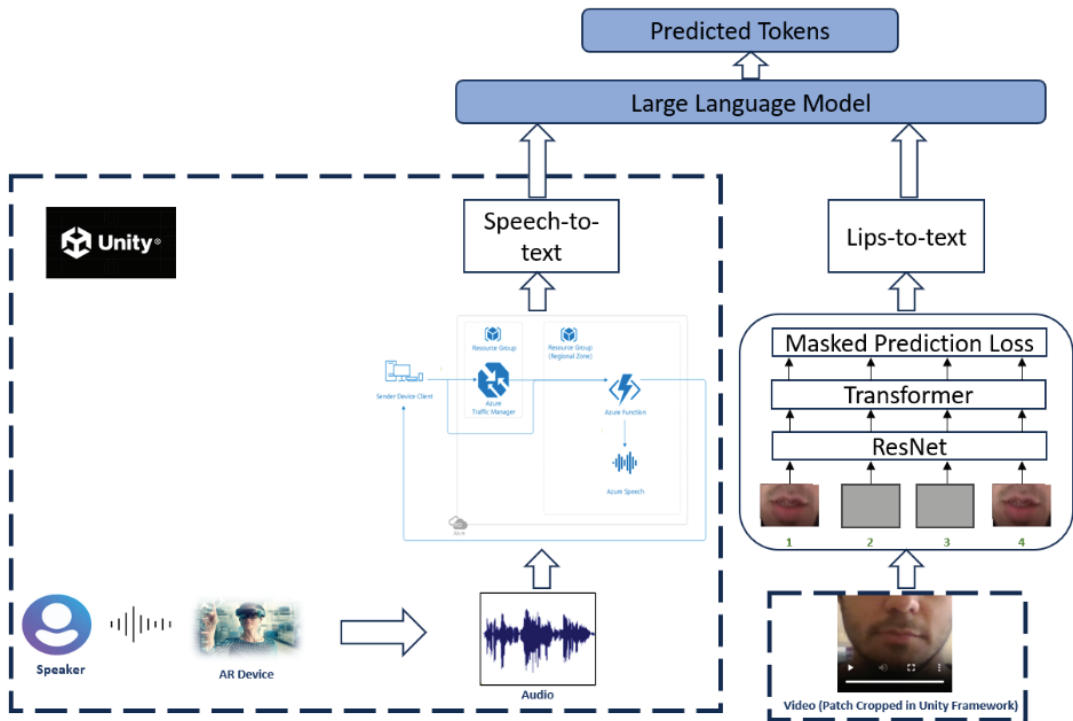


Figure P. 10. Proposed System Architecture.

Results

The Python-based prototype, which processes entire frames for landmark detection and lip localization, exhibited a lag of 81 seconds for a 10-second video at 30 frames per second (fps). In contrast, the Unity-based lip localization approach performs all preprocessing steps in real-time, reducing the total processing time to about 20 seconds for the same video length and frame count. This represents a substantial improvement of 75% over the first iteration. While the Azure Speech API's accuracy was consistently 100% across all tests, Table P. 1 shows the comparison of Word Error Rates (WERs) and latencies of the Python-only version and our multi-platform approach using Unity and Python.

These experiments were conducted on a laptop equipped with an 8GB NVIDIA 3060 RTX Graphics Processing Unit (GPU). While a lag of 20 seconds for processing 300 frames is not ideal at this point in time, GPUs are advancing rapidly and becoming more affordable. According to the Stanford Artificial Intelligence Index Report 2023 (Maslej et al., 2023), since 2021, the median floating-point operations per second (FLOP/s) speed has nearly tripled, and since 2003, it has increased roughly 7,000 times. The report also states that the median FLOP/s per U.S. dollar of GPUs in 2022 was 1.4 times greater than in 2021 and 5,600 times greater than in 2003, demonstrating a doubling in performance every 1.5 years. Given the rapid advancements and increasing affordability of hardware over the past few years, the prototype clearly shows the potential to achieve real-time or near-real-time implementation of Automated Lip Reading in the realm of eXtended Reality.

Table P. 1 Performance Comparison between Python-Only implementation (V1.0) and Unity+Python implementation (V2.0) (Latency calculated in seconds)

Video Length	V1.0 WER	V1.0 Latency	V2.0 ALR WER	V2.0 ALR Latency
10 sec	50%	81.46	43.80%	20.13
3 sec	50%	34.64	0%	9.55
6 sec	22.20%	51	11.10%	13.03
9 sec	21.40%	73.22	21.40%	18.12
12 sec	35.30%	86.43	29.40%	21.27
15 sec	40%	92.83	40%	23.03

Future Work

In the future, this research intends to implement a functionality utilizing a Large Language Model (LLM) for multimodal fusion to compare two sentences: one generated by Automatic Speech Recognition (ASR) and the other by Automated Lip Reading (ALR). In the presence of noise interference, the ALR system is expected to enhance speech recognition; however, its accuracy may vary. This LLM-based functionality will systematically compare the outputs from both ASR and ALR systems and endeavor to predict the most accurate sentence possible.

Furthermore, we plan to deploy this architecture on the HoloLens 2, providing a live captioning system based on audio speech recognition, with the capability to record conversations in noisy environments. This will enable users to capture and document conversations in noisy environments. Subsequently, by employing Automated Lip Reading (ALR), users will be able to attain a more accurate understanding of the spoken content. Due to the limited computational power of the HoloLens 2, an edge computing approach will be adopted to perform heavy computational tasks on cloud servers. This system allows for retrospective analysis, enabling users to revisit and review conversations for enhanced comprehension. In the future, once this system is ready, we plan to conduct user studies to evaluate its effectiveness and suitability for real-world applications.

References

- Anwar, M., Shi, B., Goswami, V., Hsu, W. N., Pino, J., & Wang, C. (2023). Muavic: A multilingual audio-visual corpus for robust speech recognition and robust speech-to-text translation. arXiv preprint arXiv:2303.00628.
- Assael, Y. M., Shillingford, B., Whiteson, S., & De Freitas, N. (2016). Lipnet: End-to-end sentence-level lipreading. arXiv preprint arXiv:1611.01599.
- Dick, E. (2021). The promise of immersive learning. Information Technology & Innovation Foundation, Washington, DC. <https://itif.org/publications/2021/08/30/promise-immersive-learning-augmented-and-virtual-reality-potential/>.
- Eckert, D., & Mower, A. (2020). The effectiveness of virtual reality soft skills training in the enterprise: a study. PwC Public Report.
- Fact.MR (2021) - Augmented reality (AR) in education market. Available at: <https://www.factmr.com/report/ar-in-education-market>
- Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., ..., & Perrault, R. (2023). Artificial Intelligence Index 2023 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA.
- Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., & Auli, M. (2019). Fairseq: A fast, extensible toolkit for sequence modeling. In Proc Conf. NAACL-HLT 2019 (pp. 48-53). Minneapolis, MN, USA.
- Shi, B., Hsu, W. N., Lakhota, K., & Mohamed, A. (2022). Learning audio-visual speech representation by masked multimodal cluster prediction. arXiv preprint arXiv:2201.02184. <https://doi.org/10.48550/arXiv.2201.02184>
- The World Bank (2023) Disability Inclusion Overview. Available at: <https://www.worldbank.org/en/topic/disability>.

Psychometric Properties of the Greek Versions of VIMSSQ, FMS, and SSQ for Assessing Cybersickness

Vasiliki Stefanouli¹, Dimitrios Diamantis², Eirini Cholopoulou², Chrysoula Lazou³,
Asimakis Kanellopoulos¹, Dimitris Iakovidis², and Nikolaos Strimpakos^{1,4}

¹*Department of Physiotherapy, School of Health Sciences, University of Thessaly, Lamia, Greece*

²*Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece*

³*Department of Computer Science, Democritus University of Thrace, Kavala, Greece*

⁴*Honorary Research Associate, Division of Musculoskeletal & Dermatological Sciences,
University of Manchester, UK*

Corresponding author: vstefanouli@uth.gr

Keywords: Cybersickness, Motion Sickness, Validation, Reliability, Greek version

Abstract

Background

Cybersickness, a specific form of visually induced motion sickness, has emerged as a prevalent issue in virtual reality (VR) environments (McCauley and Sharkey et al., 1992; Biswas et al., 2024). As VR technology continues to expand across various fields—ranging from entertainment to education, healthcare, and professional training—cybersickness can undermine user comfort, acceptance, and the overall potential of immersive experiences. Cybersickness symptoms, such as nausea, disorientation, dizziness can deter individuals from engaging fully with VR applications (Laviola, 2000; Biswas et al., 2024). To mitigate these effects, reliable tools for assessing susceptibility and severity of cybersickness are essential. In this context, three primary tools are widely recognized for their ability to measure motion sickness and in many cases cybersickness (Chang et al., 2020): the Simulator Sickness Questionnaire (SSQ), the Fast Motion Sickness Scale (FMS), and the Visually Induced Motion Sickness Susceptibility Questionnaire (VIMSSQ) – Short Form. The SSQ is extensively validated and widely recognized for its ability to capture a broad range of simulator sickness symptoms, making it a commonly used tool for post-exposure assessment (Kennedy et al., 1993). The FMS is a quick and efficient real-time scale (from 0 to 20) that tracks nausea during VR exposure (Keshavarz and Hecht, 2011). The short form of the VIMSSQ is specifically designed to measure susceptibility to visually induced motion sickness, with a focus on prior experiences with visual stimuli (Golding et al., 2021). Given the limited availability of translated and validated tools in Greek for assessing motion sickness—particularly cybersickness—this study offers a comprehensive framework for evaluating cybersickness. These tools contribute to ongoing efforts to enhance the accessibility and comfort of VR experiences for Greek users.

Purpose

The purpose of this study was to translate, adapt, and to evaluate the test-retest reliability and validity of the SSQ, FMS, and VIMSSQ questionnaires in the Greek language and culture.

Methods

The translation and cross-cultural adaptation of the questionnaires were conducted following the internationally accepted guidelines (Beaton et al., 2021). A total of 64 participants (34 females and 30 males) were recruited to assess the validity and reliability of these tools. Upon arrival at the laboratory in the Department of Physiotherapy, participants first completed a demographic questionnaire and the VIMSSQ. This was followed by the SSQ and the Motion Sickness Assessment Questionnaire (MSAQ) to provide baseline measurements. The MSAQ had already been validated in Greek (Kousoulis et al., 2016). Participants then took part in a 3-minute and 34-second VR roller coaster simulation (EPIC Roller Coaster, moderate difficulty) using the Oculus Quest 2. Cybersickness levels were assessed every minute using the FMS scale (ranging from 0 to 20). During the simulation, all participants were seated (Figure P. 11). After the VR session, participants reported their peak sickness rating on the FMS and completed the SSQ and MSAQ again. To assess test-retest reliability, 40 participants repeated the procedure one week later, completing the SSQ and FMS again, while 45 participants completed the VIMSSQ for a second time. Data were tested for normality using the Kolmogorov–Smirnov test. Test-retest reliability was evaluated using the Spearman's Rank-Order Correlation (non-parametric data). For concurrent validity, the results from all tools were correlated with the MSAQ (Spearman's Correlation). Statistical analyses were conducted using IBM SPSS Statistics Software (version 29.0). The study received approval from the ethics committee (approval number 11909/24/ΤΦΣΚΘ - 12/06/24).

Results

There were no notable differences between the original and translated versions, apart from one item, which was resolved by the four translators. The results demonstrated a strong correlation (Spearman's Rank-Order Correlation) between the two administrations of each questionnaire, indicating very good test-retest reliability. Specifically, there was a very strong correlation between SSQ(1) and SSQ(2) ($r = 0.782$, $p < 0.001$), VIMSSQ(1) and VIMSSQ(2) ($r = 0.764$, $p < 0.001$), and FMS(1) and FMS(2) ($r = 0.717$, $p < 0.001$), reflecting consistent responses across the two time points.

A Spearman's correlation revealed significant positive correlations between the VIMSSQ Total, SSQ Total, FMS Peak, and the validated MSAQ. VIMSSQ Total was moderately correlated with MSAQ Total ($r = 0.593$, $p < 0.001$), and strongly correlated with SSQ Total ($r = 0.654$, $p < 0.001$). SSQ Total showed a very strong correlation with MSAQ Total ($r = 0.841$, $p < 0.001$), and both SSQ Total and FMS Peak were strongly correlated ($r = 0.631$, $p < 0.001$). FMS Peak also showed a strong correlation with MSAQ ($r = 0.620$, $p < 0.001$) and moderate correlation with VIMSSQ ($r = 0.456$, $p < 0.001$). These results indicate a strong correlation among all tools, supporting the reliability and validity of VIMSSQ, SSQ, and FMS.



Figure P. 11. Participants during the roller coaster simulation

Conclusion

The VIMSSQ, SSQ, and FMS demonstrate very good test-retest reliability and show statistically significant positive correlations with the MSAQ, ranging from moderate to very strong. Therefore, these tools are considered reliable and valid for assessing cybersickness, particularly in VR environments, for

the Greek population. However, further research with larger sample sizes and more diverse VR conditions is recommended to strengthen these findings and explore the broader applicability of these tools in different contexts.

References

- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Biswas, N., Mukherjee, A., & Bhattacharya, S. (2024). "Are you feeling sick?" A systematic literature review of cybersickness in virtual reality. *ACM Computing Surveys*. <https://doi.org/10.1145/3670008>
- Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual reality sickness: a review of causes and measurements. *International Journal of Human-Computer Interaction*, 36(17), 1658-1682. <https://doi.org/10.1080/10447318.2020.1778351>
- Golding, J. F., Rafiq, A., & Keshavarz, B. (2021). Predicting individual susceptibility to visually induced motion sickness by questionnaire. *frontiers in Virtual Reality*, 2, 576871.
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3), 203-220. https://doi.org/10.1207/s15327108ijap0303_3
- Keshavarz, B., & Hecht, H. (2011). Validating an efficient method to quantify motion sickness. *Human factors*, 53(4), 415-426. <https://doi.org/10.1177/0018720811403736>
- Kousoulis, P., Pantas, A., Alevetsovitis, G., & Fydanaki, O. (2016). Psychometric properties of the Greek version of the gianaros motion sickness assessment questionnaire. *Aerospace medicine and human performance*, 87(11), 954-957. <https://doi.org/10.3357/AMHP.4540.2016>
- LaViola Jr, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32(1), 47-56. <https://doi.org/10.1145/3441234>
- McCaughey, M. E., & Sharkey, T. J. (1992). Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoperators & Virtual Environments*, 1(3), 311-318. <https://doi.org/10.1162/pres.1992.1.3.311>

eXtended Reality Content Prototyping through On-Device Solutions: An Experimental Platform based on Google's MediaPipe Face-Tracking and Unity

Stavroula Zoi^{1,2}, and George Caridakis¹

¹Intelligent Systems Research Lab, Department of Cultural Technology & Communication, University of the Aegean, Mytilene, Greece

²School of Fine Arts, Department of Visual Arts, Multimedia Laboratory, Athens, Greece

Corresponding author: s.zoi@aegean.gr, voula.zoi@gmail.com

Keywords: eXtended Reality (XR) Content, Face Tracking, On-Device Machine Learning, Google's Mediapipe, Unity

Introduction

Human motion tracking technologies, e.g. eye, face, hand, and body tracking, play an important role in the eXtended Reality (XR) ecosystem of technologies, as they establish new spatial relations between the user and the device. However, these technologies are mostly accessible to content creators through commercial XR headsets, which are not always practical for supporting all aspects of the creative process, such as freedom of movement, facial expressiveness, self-observation, and intuitive interactions. Furthermore, their dedicated SDKs impose certain restrictions, such as limited access to raw camera data or proprietary tracking algorithms, which may prevent from exploring customized forms of content. On the other hand, advancements in computer vision and machine learning technologies have made tracking functionalities even more accessible and cost-effective. They are now possible using low-cost web cameras and smartphone cameras, through on-device inferencing. Personal devices, particularly smartphones, are integral to daily life, enabling creators to use them as observation and capturing tools. With their growing computational capabilities, integrated sensors, and advanced camera systems, smartphones are now capable of handling complex content, thus establishing a significant role in the XR creative ecosystem. Considering the above, an experimental platform is presented based on Google's MediaPipe and Unity. It enables to prototype different forms of XR content by leveraging spatial relationships between the user, the device and the physical space, as defined through face-tracking functionalities and smartphone sensors. As the possibilities within this creative framework are limitless, this study presents a case of a spherical content prototype.

The XRiScope v1.0 experimental platform

The XRiScope v1.0 experimental platform is implemented with Google’s MediaPipe (Lugaresi, 2019) and Unity¹⁶. MediaPipe is an open-source and cross platform solution that supports real-time, on-device machine learning pipelines, including face and iris tracking, without relying on external hardware or cloud-based computing. The integration of MediaPipe to Unity 6 was based on the implementation found in (GitHub, 2024). As MediaPipe is a data-centric, cross platform solution, it allows deployment across different platforms leveraging options available through Unity. Inference Mode (CPU or GPU) can be selected; GPU option is available when deploying to smartphone devices.

Among available perception pipelines, this study focuses on face and iris tracking using the 478 canonical landmark model employed by MediaPipe. It includes 468 landmarks across the human face, and 10 landmarks for the irises. (Google, n.d.). Landmarks are normalized to [0.0, 1.0] by the image width and height respectively. However, they can be represented as a point with x, y, z coordinates within MediaPipe’s 3D metric space. It is a right-handed orthonormal 3D coordinate space, assuming a virtual perspective camera located at the space origin and pointed in the negative direction of the Z-axis. (Google Developers, 2020)

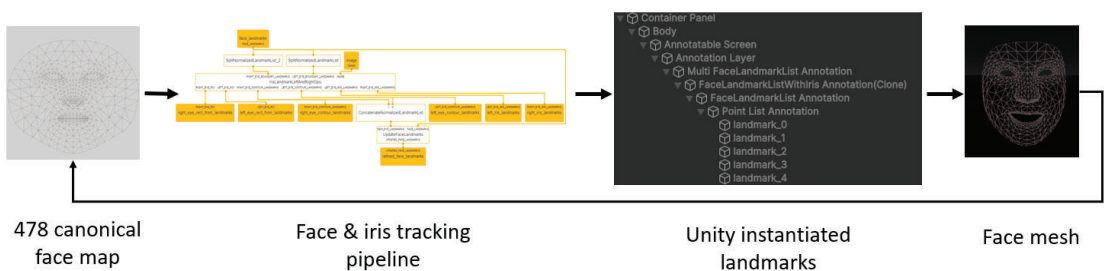


Figure P. 12. Landmark Identification as instantiated game objects and real-time hollow mask creation.

Face and Iris tracking pipeline is available within Unity 3D editing space through scenes: Iris Tracking, Face Landmark Detection, and Face Mesh. Face and irises landmarks are provided to Unity Hierarchy, in real-time, as instantiated game objects, only when the whole face is properly tracked (otherwise they appear deactivated). They are properly named through a dedicated script, corresponding to the numbered canonical face map. In this way, it is possible to trace their position, according to subject expressions and create a hollow mask mesh (Figure P. 12).

Physical and Virtual Space are conceptualized using different coordinate systems (2D and 3D) that need to be aligned, requiring transformations between these metric spaces to manage different content structures. Figure P. 13 illustrates associated concepts, while Table P. 2 summarizes the implemented modules for parameter configuration and for supporting related real-time functionalities.

Experimentations and preliminary results

Experimentation and prototyping possibilities are endless, in this creative framework. One case is the definition of a *Spherical Content Prototype* that unfolds around the user leveraging smartphone

¹⁶ <https://unity.com/>

gyroscope sensor. Facial reference enables the definition of perspective and distance-based functionalities (e.g. content zones, zoom) and touchless interactions defined based on parameters of Table P. 2, or relations between landmarks (e.g. identified expressions).

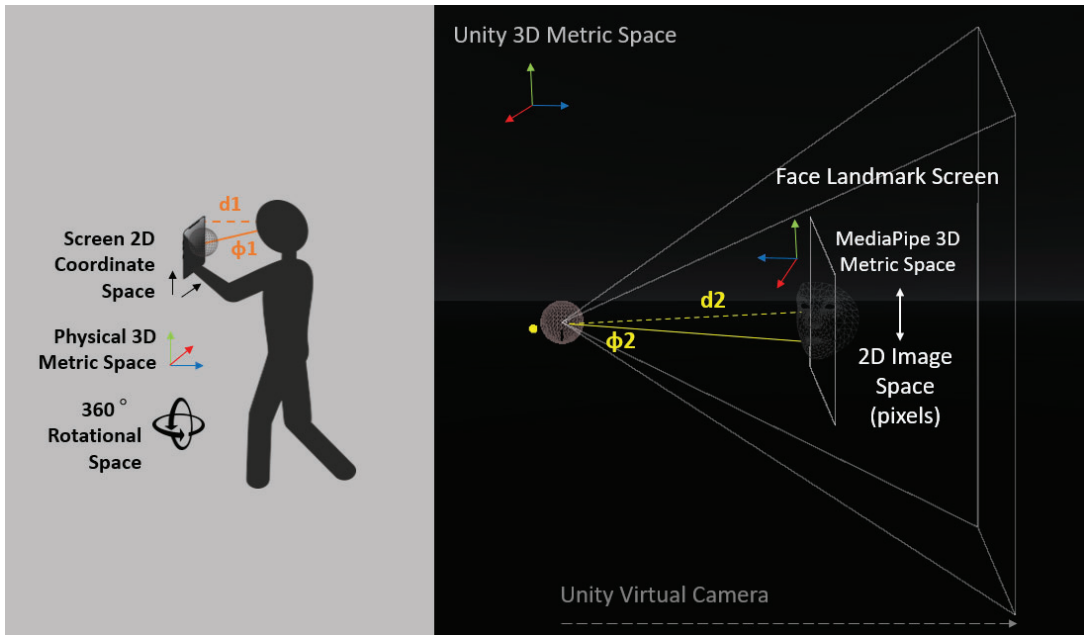


Figure P 13. Physical, Virtual & MediaPipe Metric Spaces

Table P. 2. Modules for parameter configuration and implementation of related real-time functionalities.

Module	Parameter Configurations	Real-Time Functionalities
Physical Camera Config	Front / Back Enabled Video Stream Visible / Invisible	Switch Front / Back (to be defined)
Landmark Identification	Landmark Numbered IDs (consistent with Canonical Face Map)	Landmark IDs Update (based on MediaPipe Live Data Stream) Landmark Relations (e.g. gaze point)
Face Landmark Screen View	Image Size in Pixels (Width, Height) Invisible / Visible in Virtual Space & Scale Face Landmarks / Face Mesh View Overlay (Fullscreen / Fixed Window) / Virtual World Placed Flipped based on Physical Camera (Front / Back) / Virtual Camera View	MediaPipe to Unity Metric Space Transformations Distance (d2) & Angle (φ_2) from Virtual Camera Calculation Distance (d1) from Physical Camera Calculation (Vakunov, 2020)
Physical Spatial Manager	Distance (d1) & Angle (φ_1) Gyroscope Sensor for 360° View 360° Content Zones Definition based on d1	d1 & φ_1 Visualization 360° Content Zone Transitions based on d1 Facial Reference Spatial Interactions
Unity Virtual Camera Manager	d2, φ_2 Field of View (FoV) Position, Orientation based	$d_2 = f(d_1)$, $\varphi_2 = g(\varphi_1)$ Zoom In / Zoom Out based on d1 (FoV / Camera Movement Zoom) 3D Screen Effect based on d1, φ_1 Touchless Actions Trigger based on d1, φ_1 , d2, φ_2 and Raycasting

Figure P. 14 illustrates a structural visualization of this prototype. It is based on a 3D sphere model with flipped normals that defines the inside of a fully spherical environment. The Virtual Camera is positioned at the center of this spherical space and it can be rotated based on gyroscope input, allowing the user to fully inspect the environment. Given a distance-based spherical content zone, game objects are positioned or instantiated (e.g. on mesh vertices, as illustrated through the wireframe shader of Figure 3, or at certain spherical coordinates). Their properties can be affected based on current *View Area*, and actions can be triggered based on *Angle of View*. Angle of View is visualized through the yellow line (sight vector) connecting gaze point and Virtual Camera center point.

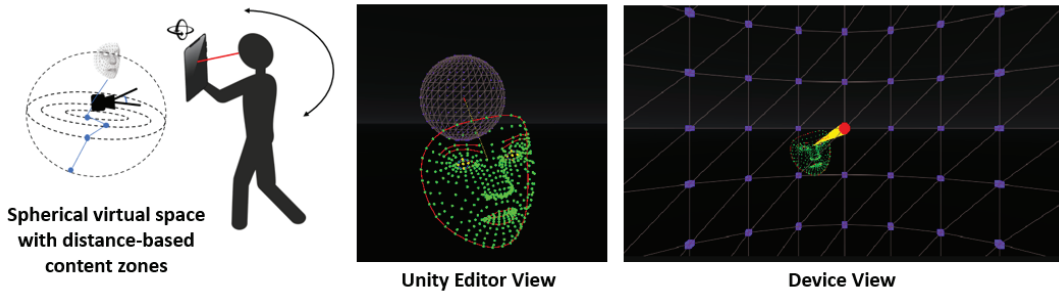


Figure P. 14. Spherical Content Prototype Visualization.

Future work

The Spherical Content Prototype will be expanded, and new ones will be explored, including additional perception pipelines (e.g. cartesian cubic prototype using object detection to identify physical space references), in order to study more complex spatial relationships between user, device, physical space and content. The goal is to lead to reusable content templates. In addition to on-device solutions, the proposed platform serves as a tool for studying data centric approaches, contributing to cross-platform XR content standardization.

Acknowledgements

This research is part of Stavroula Zoi's postdoctoral work under the supervision of Associate Professor George Caridakis, at the University of the Aegean.

References

- GitHub. (2024). MediaPipe Unity plugin. Retrieved in September 2024, from <https://github.com/homuler/MediaPipeUnityPlugin>
- Google. (n.d.). Face landmark detection guide. Retrieved in September 2024, from https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker
- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.
- Sokal, K. (2020). MediaPipe 3D Face Transform. Retrieved in September 2024, from <https://developers.googleblog.com/en/mediapipe-3d-face-transform/>
- Vakunov, A., & Lagun, D. (2020). MediaPipe Iris: Real-time Iris Tracking & Depth Estimation. Google Research. Retrieved in September 2024, from <https://research.google/blog/mediapipe-iris-real-time-iris-tracking-depth-estimation/>

Dynamic City-Scale Digital Twins for Extended Reality

Tariqul Islam, Jyothi Pudota, and Rahul Tomar

DigitalTwin Technology GmbH, Cologne, Germany

Corresponding author: tariqul.islam@digitaltwin.technology

Keywords: Robust Localization, Sensor Alignment, City-Scale Digital Twin, Extended Reality

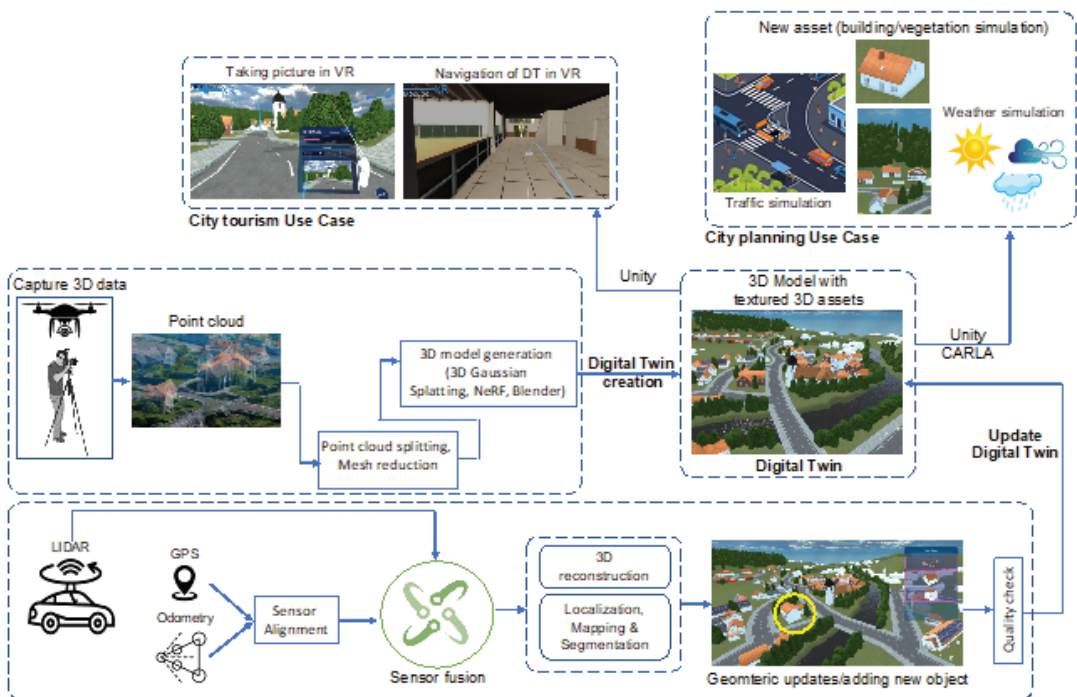


Figure P 15. Process diagram showing the digital twin (DT) creation, update, simulation, city planning, city tourism use cases and underlying technologies in each step of the DT.

Introduction

Digital Twins (DTs) are evolving digital counterparts of physical objects, continuously updated with real-time data. Unlike static 3D models, DTs dynamically reflect real-world changes and are particularly valuable in scenarios where physical replicas are not feasible. In the EU-funded “DIDYMOS-XR” project (Didymos-XR, 2024), we have been developing city-scale DTs for Extended Reality (XR) applications, with

a focus on DT creation, city planning, and city tourism. These systems are designed to enhance urban planning through integrated data and advanced simulations, leveraging dynamic, highly accurate digital representations of urban areas. Our approach incorporates cutting-edge technologies for 3D reconstruction, sensor alignment, localization and mapping, and rendering to address various challenges in creating, storing, streaming, and manipulating DTs in XR.

The process of creating city-scale DTs begins with collecting 3D data using LIDAR sensors or cameras, which is then merged with geospatial information from GIS databases. Initial point clouds are refined into high-fidelity 3D assets, including textures, vegetation, and infrastructure like roads. Methods like 3D Gaussian splatting (Kerbl, 2023) and Neural Radiance Field (NeRF) (Mildenhall, 2021) are employed for model generation, with further refinements done using tools such as Blender (Blender, 2024). To ensure the fidelity and relevance of the DT, periodic updates are required using data from vehicles/drones equipped with cameras, GPS, and odometry sensors. These updates leverage techniques like sensor fusion and alignment, object detection, and compression algorithms to ensure smooth model transmission.

Once the 3D model is created, it is integrated into simulation environments like Unity3D (Unity3D, 2024) and Carla (Dosovitskiy, 2017) to analyze urban dynamics. This enables a wide range of simulations, including weather conditions, smart infrastructure, and traffic patterns. City planners can explore “what if” scenarios-adjusting traffic flow, introducing new structures, or simulating autonomous vehicles. Furthermore, city tourism applications can also use XR environments to offer location-based experiences. Remote experience, via VR, lets users explore the city, checking real-time data on traffic, crowd levels, and environmental conditions, while also accessing shop and museum hours. A tourist planner tool could help organize groups and optimize tours to avoid traffic and mobility issues. The process diagram in Figure P 15 shows the technologies and processes used in this work.

The underlying technologies that enable the immersive exploration, navigation, simulation, and rendering of the DT in XR are described in the following section.

Descriptions of the technologies developed for city-scale digital twins

A range of technologies plays a key role in the development of DTs, city planning, and city tourism applications, including 3D scene reconstruction, sensor alignment, fusion, localization, scene understanding, rendering, and XR. In the following sections, we outline the technologies developed in this work.

3D reconstruction

Creating a city's digital twin starts with 3D data acquisition, such as point clouds from the target location. The point cloud is then pre-processed, involving tasks like splitting and mesh reduction, to accelerate the development of a detailed 3D model featuring enhanced assets (e.g., buildings, traffic lights, smart city elements). Following this, 3D model generation, which includes buildings, road networks, and vegetation with textures and geometry, can be done using two methods: automated AI techniques like 3D Gaussian splatting and NeRF, or manual methods using tools such as Blender. Our approach combines both, where the initial model is created using 3D Gaussian splatting and NeRF, with further refinements made in Blender. This streamlines the model development and ensures accurate integration of the processed point clouds into the overall pipeline.

Sensor alignment, fusion, and localization

Here, we analyzed and processed non-visual sensor data such as GPU and Odometry for integration into the DT model to improve the overall semantics and accuracy of the map. For the sensor data alignment, after converting the GPS and Odometry data into a single coordinate space (cartesian space), we applied outlier removals and then developed an adaptive Monte Carlo Localization (MCL) (Dellaert, 1999) by leveraging unsupervised machine learning method for dynamic adjustment of the particles, reduced computational overhead and improved localization accuracy. Here, future GPS locations of vehicles inside a scene/model are predicted based on the GPS and Odometry data. Our low-variance re-sampling method and improved motion and pose models helped achieve better computational efficiency and better estimation of particle motion and pose.

In addition to the non-visual sensor-based localization, we also developed a hybrid self-localization approach by leveraging visual sensors' data where we introduce Image-based localization (IBL) to the SLAM pipeline (particularly with H-SLAM (Georges Younes, 2024)) for addressing the global localization issues found in SLAMs. We are working with two approaches - IBL and NeRF-based localization, which takes geo-referenced images as input. We developed a Unity-based virtual camera setup to supply geo-reference images to F-SLAM to test the overall localization accuracy.

The localization outputs from the SLAM-based approach and non-visual sensors are fused to improve the overall localization in the city planning and city tourism use cases.

2.3 Rendering for XR environments

We worked towards creating adaptive and scalable XR rendering for different content, integrating data based on point cloud geometry and the user's navigation area. These implementations are intended to meet users' needs in a personalized manner (Arvanitis, 2023). We are working on catering the 3D model's contents inside the DT, for XR environments, to diverse user profiles, including their capabilities, preferences, and experience levels, while also being compatible with a range of devices such as heads-up displays, smartphones, and tablets, across various environments, featuring different lighting and weather conditions. We are looking into key areas such as saliency maps corresponding to point cloud geometry, user distance, and user's viewpoint. Additionally, we are addressing technical aspects such as accurate shadow casting and occlusion handling to enhance navigation within the DT in XR setups.

Conclusion and future works

In this work, we present our efforts in developing, updating, and visualizing dynamic city-scale digital twins for DT creation, interaction, city planning, and tourism, aimed at enabling large-scale XR applications. Additionally, we highlight the enabling technologies employed in creating these digital twins, such as 3D scene reconstruction, sensor alignment and fusion, localization, and rendering, with a particular focus on those crucial to the discussed use cases. Looking ahead, we aim to further enhance and integrate these technologies to ensure seamless creation, manipulation, and immersive exploration of city-scale DTs in XR.

Acknowledgements

This work has received funding from the European Union under Grant Agreement No. 101092875.

References

- Arvanitis, G., & Moustakas, K. (2023). Digital Twins and the City of the Future: Sensing, Reconstruction and Rendering for Advanced Mobility.. ITS2023: Intelligent Systems and Consciousness Society.
- Blender. (25. Sep. 2024). Blender - a 3D modelling and rendering package. Retrieved: <http://www.blender.org> .
- Didymos-XR. (25. Sep. 2024). Didymos-XR. Retrieved from Didymos-XR: <https://didymos-xr.eu/> .
- Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017, October). CARLA: An open urban driving simulator. In Conference on robot learning (pp. 1-16). PMLR. <https://doi.org/10.48550/arXiv.1711.03938>
- Younes, G., Khalil, D., Zelek, J., & Asmar, D. (2024). H-SLAM: Hybrid direct-indirect visual SLAM. Robotics and Autonomous Systems, 104729. <https://doi.org/10.48550/arXiv.2306.07363>
- 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. <https://doi.org/10.48550/arXiv.2308.04079>
- 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.
- Dellaert, F., Fox, D., Burgard, W., & Thrun, S. (1999, May). Monte carlo localization for mobile robots. In Proceedings 1999 IEEE international conference on robotics and automation (Cat. No. 99CH36288C) (Vol. 2, pp. 1322-1328). IEEE.
- Unity3D. (25. Sep. 2024). Retrieved from <https://unity.com> .

Voice-based Interactions in Immersive Applications – Cognitive and Affective Dimensions of User Experience

Olga Chatzifoti, and Maria Roussou

National and Kapodistrian University of Athens, Greece

Corresponding author: olchatz@di.uoa.gr

Keywords: Natural Input Modalities, Voice-Based Interactions, User Experience, Immersive Technologies, Human-Computer Interaction

Introduction

This research explores the cognitive and affective impact of voice as a natural input modality in immersive experiences and examines if and how voice-based interactions influence the users' critical and emotional engagement with the content. This work advances current UX research by moving beyond usability to explore more nuanced and complex user experience constructs. The research is based on design, development, and evaluation in user studies of three artifacts ranging from reading linear narratives to engaging in scripted and unscripted dialogues with AI-driven NPCs. AI models for speech transcription, text matching, and natural language generation are employed. Research is contextualized in the medium of interactive narratives because of the importance that text (and its verbalization) holds on the user experience.

Motivation

Voice in HCI is of high interest due to its perceived value from a user perspective (Seaborn et al., 2022), its contemporaneity as a technological achievement (Alharbi et al., 2021), and its complexity as a neurophysiological process. Advances in automated speech transcription (ASR) and natural language understanding and generation (NLU/NLG) have increased research on usability, satisfaction, and performance (Aslam & Zhao, 2024; Monteiro et al., 2021; WeiB et al., 2018). However, there is limited work on emotion and engagement. These constructs, despite their importance, present a methodological challenge in HCI. Yet, they are integral to the user experience in immersive spaces, because of correlations with presence (Riva et al., 2007) and embodiment (Kilteni et al., 2012). This work aims to help bridge these gaps.

Related Work

Research on voice interactions in video games has noted interaction patterns (Zargham, Friehs, et al., 2024) and highlighted the influence of social context in both single and multiplayer settings (Zargham, Dratzidis, et al., 2024). In immersive environments, studies have shown that voice interactions

can amplify emotional effects on user experience within immersive narratives (Osking & Doucette, 2019). However, further validation is needed due to the limited scale and design of this initial experiment. Additionally, the experience of voice interactions appears to correlate with the level of immersion: users found speech interfaces awkward in non-immersive settings (Carter et al., 2015), less preferable in AR (Pourmemar & Poullis, 2019), but highly preferable in fully immersive VR (Osking & Doucette, 2019; WeiB et al., 2018, p. 2).

Research Approach and Methodology

Core research questions (RQs) are: “Is using voice as an input modality in immersive experiences influencing...”

- RQ1 “...the user’s sense of place and/or plausibility illusion?”
- RQ2 “...the user’s emotional response to the content?”
- RQ3 “...the user’s critical engagement with the content?”

The study employs a cascading and iterative approach. Three experiments with increasingly complex interactions will be performed: from voice to speech to conversation, framed in corresponding interactive narrative schemas respectively: from linear to branching to emerging narrative, as illustrated in Figure P 16.

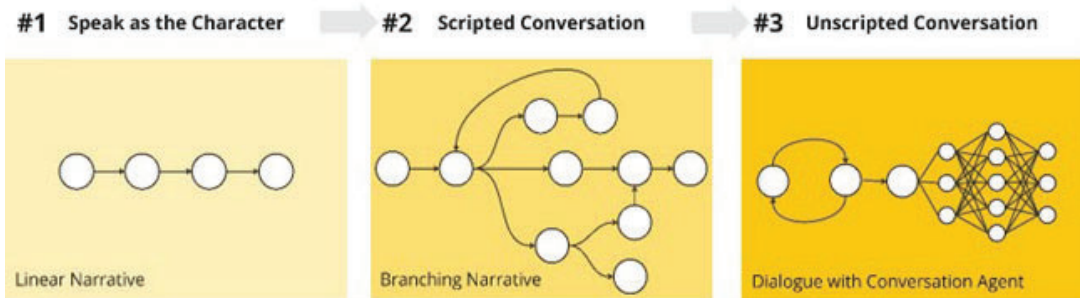


Figure P 16. Incremental experiment design

Experiment #1 is a single-user experience of reading out loud a text story with a linear flow in an immersive environment. Following Allison’s et al. work on design patterns (Allison et al., 2018) this design falls under the “Speak as a character” pattern in the “Diegetic framing” category.

Experiment #2 will expand to a branching narrative using dialogue interactions with NPC(s). The design falls under the “Scripted conversation” pattern in the “Dialogue structure” category (Allison et al., 2018). The user will be presented with their options in a multiple-choice format and will need to read their choice out loud.

Experiment #3 will investigate free-form input implementing the design pattern “Unscripted Conversation” in the “Dialogue structure” category. This will require the design and development of an AI conversation agent (CA). The investigation is focused on the user experience of free-form speech, and not on the emergent properties of the narrative or the qualities of the agent.

Current State & Next Steps

The first research artifact, developed in Unity using OpenXR, incorporates an open-source ML model for transcription combined with text matching for improved accuracy and latency¹. Experiment #1 is

planned for late autumn 2024 with results expected by February 2025 and experiment #2 is scheduled for summer 2025.

Expected Contributions

Expected outcomes include design guidelines for voice-based interactions in immersive experiences, accounting for synergetic and/or conflicting patterns, technological dependencies, and future research recommendations.

Challenges & Limitations

Adopting a design and development approach for the research artifact, compared to using an existing work, allows for better scoping of features but requires addressing the artistic dimensions of the medium. Evaluating user experience in narratives (Roth & Koenitz, 2016) and voice interactions from an HCI perspective (Seaborn & Urakami, 2021) present methodological complexities. Self-reporting presents challenges due to its subjectivity, particularly for nuanced constructs like presence (Slater, 1999), but exploring biometric tracking (e.g., eye tracking, EEG) also offers complexities. Variability in affective responses and sensitivity to environmental factors have been noted by previous research (Riva et al., 2007).

Conclusion

Transitioning to voice interactions as an interaction paradigm in immersive experiences involves more than usability considerations; it requires an understanding of the interplay between cognition, affect, and user experience. The complexity of voice-based interactions, coupled with the subtle nuances of these constructs, suggests a step-by-step approach for this research. The findings could inform design guidelines for integrating voice-based interactions into immersive media, as well as highlight future research directions.

Acknowledgments

This work has been partly supported by the European Union-funded research project VOXReality. Special thanks are extended to fellow PhD candidates, Abderrahmane Issam and Paweł Mąka, and the Department of Advanced Computing Sciences of the University of Maastricht for their support on machine learning technologies.

References

- Alharbi, S., Alrazgan, M., Alrashed, A., Alnomasi, T., Almojel, R., Alharbi, R., Alharbi, S., Alturki, S., Alshehri, F., & Almojil, M. (2021). Automatic Speech Recognition: Systematic Literature Review. *IEEE Access*, 9, 131858–131876. <https://doi.org/10.1109/ACCESS.2021.3112535>
- Allison, F., Carter, M., Gibbs, M., & Smith, W. (2018). Design Patterns for Voice Interaction in Games. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 5–17. <https://doi.org/10.1145/3242671.3242712>
- Allison, F., Newn, J., Smith, W., Carter, M., & Gibbs, M. (2019). Frame Analysis of Voice Interaction Gameplay. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3290605.3300623>

- Aslam, F., & Zhao, R. (2024). Voice-Augmented Virtual Reality Interface for Serious Games. 2024 IEEE Conference on Games (CoG), 1–8. <https://doi.org/10.1109/CoG60054.2024.10645616>
- Brusk, J., & Björk, S. (n.d.). *Gameplay Design Patterns for Game Dialogues*.
- Carter, M., Allison, F., Downs, J., & Gibbs, M. (2015). Player Identity Dissonance and Voice Interaction in Games. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 265–269. <https://doi.org/10.1145/2793107.2793144>
- Kilteni, K., Groten, R., & Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21(4), 373–387. https://doi.org/10.1162/PRES_a_00124
- Monteiro, P., Goncalves, G., Coelho, H., Melo, M., & Bessa, M. (2021). Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics*, 27(5), 2702–2713. <https://doi.org/10.1109/TVCG.2021.3067687>
- Osking, H., & Doucette, J. A. (2019). Enhancing Emotional Effectiveness of Virtual-Reality Experiences with Voice Control Interfaces. In D. Beck, A. Peña-Rios, T. Ogle, D. Economou, M. Mentzelopoulos, L. Morgado, C. Eckhardt, J. Pirker, R. Koitz-Hristov, J. Richter, C. Gütl, & M. Gardner (Eds.), *Immersive Learning Research Network* (pp. 199–209). Springer International Publishing.
- Pourmemar, M., & Poullis, C. (2019). Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality. *Proceedings of the 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*. <https://doi.org/10.1145/3359997.3365693>
- Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective Interactions Using Virtual Reality: The Link between Presence and Emotions. *CyberPsychology & Behavior*, 10(1), 45–56. <https://doi.org/10.1089/cpb.2006.9993>
- Roth, C., & Koenitz, H. (2016). Evaluating the User Experience of Interactive Digital Narrative. *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, 31–36. <https://doi.org/10.1145/2983298.2983302>
- Seaborn, K., Miyake, N. P., Pennefather, P., & Otake-Matsuura, M. (2022). Voice in Human-Agent Interaction: A Survey. *ACM Computing Surveys*, 54(4), 1–43. <https://doi.org/10.1145/3386867>
- Seaborn, K., & Urakami, J. (2021). Measuring Voice UX Quantitatively: A Rapid Review. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–8. <https://doi.org/10.1145/3411763.3451712>
- Slater, M. (1999). Measuring Presence: A Response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5), 560–565. <https://doi.org/10.1162/105474699566477>
- WeiB, Y., Hepperle, D., SieB, A., & Wolfel, M. (2018). What User Interface to Use for Virtual Reality? 2D, 3D or Speech—A User Study. *2018 International Conference on Cyberworlds (CW)*, 50–57. <https://doi.org/10.1109/CW.2018.00021>
- Zargham, N., Dratzidis, L., Alexandrovsky, D., Friehs, M., & Malaka, R. (2024). Gaming with Etiquette: Exploring Courtesy as a Game Mechanic in Speech-Based Games. *International Journal of Human-Computer Interaction*, 1–19. <https://doi.org/10.1080/10447318.2024.2387901>
- Zargham, N., Friehs, M. A., Tonini, L., Alexandrovsky, D., Ruthven, E. G., Nacke, L. E., & Malaka, R. (2024). Let's Talk Games: An Expert Exploration of Speech Interaction with NPCs. *International Journal of Human-Computer Interaction*, 1–21. <https://doi.org/10.1080/10447318.2024.2338666>

Impact of group synchrony, postural stability and agency on cybersickness symptoms

Julia Ayache¹, Camille Pistre¹, Pierre Jean¹, Anna Katharina Hebborn²,
Didier Stricker², and Benoît G. Bardy¹

¹*EuroMov Digital Health in Motion, Université de Montpellier, France*

²*German Research Center for Artificial Intelligence (DFKI), Germany*

Corresponding author: julia.ayache@umontpellier.fr

Keywords: Cybersickness, Postural Stability, Group Synchrony

Overview

Virtual Reality (VR) is a developing technology for computer-mediated communication (CMC). Nevertheless, challenges remain that are impeding the adoption of VR for social interactions. One of them is cybersickness, characterized by symptoms like motion sickness, such as nausea, dizziness, and eyestrain. Around 60% of users report symptoms of cybersickness when they first immerse themselves in virtual reality (Stanney et al., 2020).

According to the postural instability hypothesis (Riccio & Stoffregen, 1991), cybersickness could result from postural instability due to alterations of postural control. Therefore, greater postural instability increases the likelihood and severity of motion sickness. Interestingly, postural stability is often studied in single participants, even though an important part of our social life is composed of interactions, involving two or more individuals. Yet, only a few studies have investigated the impact of social interactions on postural stability. Varlet et al. (2011) demonstrated that visual interaction alone was sufficient to synchronize participants' postural dynamics and Gueugnon et al. (2016) found that when participants were explicitly instructed to synchronize their arm movements, their postural coordination stability increased. Overall, these studies indicate that intentional social synchronization leads to greater postural stability compared to spontaneous synchronization or the absence of social interactions. Consequently, group synchrony, characterized by the temporal coordination of movement between individuals, could increase postural stability and, therefore, reduce cybersickness.

The present study investigated this hypothesis by conducting a study in VR setting exploring the role of intentional group synchronization on postural stability and cybersickness symptoms. Groups of participants (7 triads) completed oscillatory movements with their right arm in **SOLO**, **TOGETHER** and **SYNCHRO** conditions, manipulating visual coupling and synchronization instructions.

Methods

A virtual environment was developed using Unreal Engine and the MetaXR plugin. Participants ($N = 21$) provided their informed consent and choose a gender-matched avatar before entering the

environment. Participants' upper body movements (i.e., finger and head motion) were tracked and mapped onto virtual avatars. For the **SOLO** and **TOGETHER** conditions, participants were instructed to perform the motion at their "natural" frequency. However, they were seeing each other in the **TOGETHER** condition. For the **SYNCHRO** condition, participants were instructed to achieve group synchronization. The order of conditions was counterbalanced between groups and each condition consisted of 5 trials lasting 1 min, for a total duration of 5 min per condition.



Figure P 17. Illustration of the virtual environment and avatar representations

- Group synchrony and individual contributions to synchrony were extracted using the Kuramoto model of coupled oscillators (Alderisio et al., 2017; Bardy et al., 2020).
- Postural stability was computed from the normalized length of head motion on the anteroposterior and mediolateral directions, capturing the interpersonal influences on individual body sway (Laroche et al., 2022; Tallon et al., 2013).
- Cybersickness symptoms were evaluated before and after each VR immersion. Participants completed the Simulator Sickness Questionnaire (SSQ, Bouchard et al., 2007).
- Sense of agency (SoA), the feeling of being in control of one's actions, was measured after each VR immersion using the Sense of Agency Scale (SoA, Hurault et al., 2020)

Mixed models were performed to assess the role of the experimental conditions on (i) group synchrony scores, (ii) postural stability, (iii) cybersickness symptoms and sense of agency. Spearman's correlations explored the associations between individual synchrony scores, postural stability, cybersickness symptoms and sense of agency.

Results

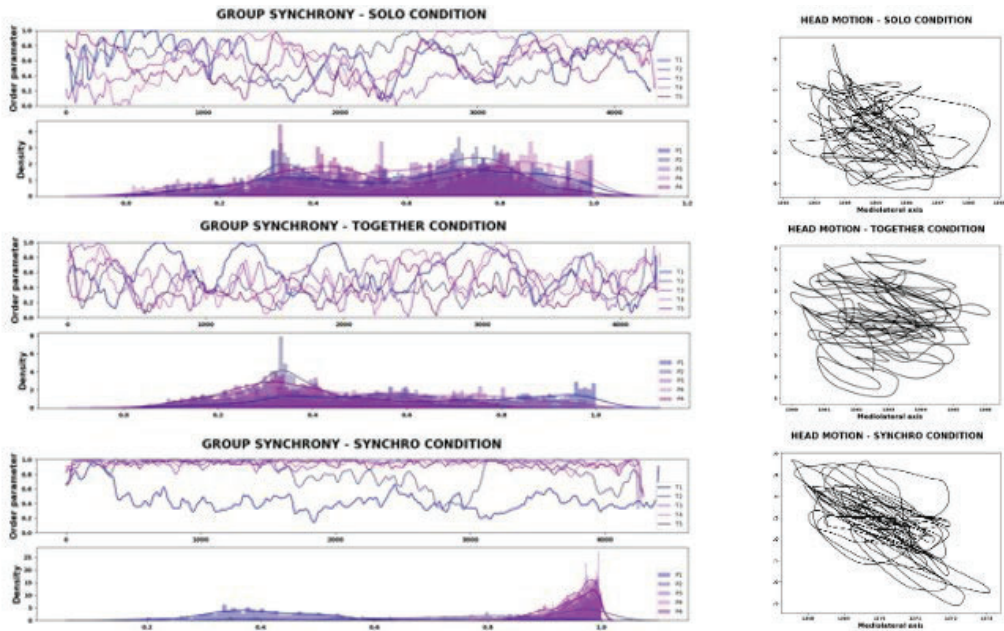


Figure P 18. Graphical representations of the impact of the experimental conditions SOLO (top), TOGETHER (middle) and SYNCHRO (bottom) on group synchrony scores and body sway.

The experimental conditions influenced group synchrony scores ($F(2, 96) = 107.36, p < .001$) and postural stability ($F(2, 310) = 3.28, p = .039$). When participants could see each other and were instructed to synchronize (SYNCHRO condition), group synchrony scores increased (adjusted $p < .001$). In contrast, body sway tended to decrease (adjusted $p = .055$). However, there was no effect of the experimental manipulation on cybersickness symptoms ($F(2, 58) = 0.17, p = .847$) and sense of agency ($F(2, 58) = 0.71, p = .494$).

Spearman's correlations showed that individual synchrony scores tended to display negative associations with body sway ($p = .212$). However, there were no significant associations of cybersickness symptoms with body sway ($p = .911$) or individual synchrony scores ($p = .937$), but a tendency for a negative association with the sense of agency ($p = .093$).

Table P. 3. Spearman's correlations between synchrony, body sway and cybersickness

	Synchrony	Body sway	Cybersickness	Sense of Agency
Synchrony	-			
Body sway	-.16	-		
Cybersickness	.01	.01	-	
Sense of agency	<.01	-.01	-.21	-

Discussion

This study replicated previous findings showing the impact of visual coupling and instructions to synchronize on group synchrony (Alderisio et al., 2017; Bardy et al., 2020). Furthermore, this study expanded previous findings observed in dyads, showing (ii) an impact of group synchrony on postural stability (Gueugnon et al., 2016; Varlet et al., 2011). However, there was no impact of synchrony nor body sway on cybersickness symptoms.

Importantly, there was an overall low report of cybersickness symptoms in this experiment and the length of body sway did not capture the dynamic fluctuations of postural coordination. Nevertheless, the association between the sense of agency and cybersickness suggests that the successful binding of actions and their consequences could prevent the occurrence of cybersickness symptoms. This finding paves the way for exploring the impact of collective movement on the temporal binding between actions and their consequences.

To conclude, this exploratory study did not confirm the postural stability hypothesis but suggested that synchronized social interactions in a VR environment do not generate cybersickness symptoms. Instead, synchrony could stabilize postural coordination through interpersonal influences, paving the way for social interventions in clinical populations.

References

- Alderisio, F., Fiore, G., Salesse, R. N., Bardy, B. G., & Bernardo, M. D. (2017). Interaction patterns and individual dynamics shape the way we move in synchrony. *Scientific reports*, 7(1), 6846. <https://doi.org/10.1038/s41598-017-06559-4>
- Bardy, B. G., Calabrese, C., De Lellis, P., Bourgeaud, S., Colomer, C., Pla, S., & Di Bernardo, M. (2020). Moving in unison after perceptual interruption. *Scientific reports*, 10(1), 18032. <https://doi.org/10.1038/s41598-020-75182-4>
- Bouchard, S., Robillard, G., & Renaud, P. (2007). Revising the factor structure of the Simulator Sickness Questionnaire. *Annual review of cybertherapy and telemedicine*, 5(Summer), 128-137. <https://doi.org/10.1037/t73606-000>
- Gueugnon, M., Salesse, R. N., Coste, A., Zhao, Z., Bardy, B. G., & Marin, L. (2016). Postural coordination during socio-motor improvisation. *Frontiers in psychology*, 7, 1168.
- Hurault, J. C., Broc, G., Crône, L., Tedesco, A., & Brunel, L. (2020). Measuring the sense of agency: a French adaptation and validation of the sense of agency scale (F-SoAS). *Frontiers in Psychology*, 11, 584145. <https://doi.org/10.3389/fpsyg.2016.01168>
- Laroche, J., Tomassini, A., Volpe, G., Camurri, A., Fadiga, L., & D'Ausilio, A. (2022). Interpersonal sensorimotor communication shapes intrapersonal coordination in a musical ensemble. *Frontiers in human neuroscience*, 16, 899676. <https://doi.org/10.3389/fnhum.2022.899676>
- Riccio, G. E., & Stoffregen, T. A. (1991). An ecological theory of motion sickness and postural instability. *Ecological psychology*, 3(3), 195-240. https://doi.org/10.1207/s15326969eco0303_2
- Stanney, K., Lawson, B. D., Rokers, B., Dennison, M., Fidopiastis, C., Stoffregen, T., ... & Fulvio, J. M. (2020). Identifying causes of and solutions for cybersickness in immersive technology: reformulation of a research and development agenda. *International Journal of Human-Computer Interaction*, 36(19), 1783-1803. <https://doi.org/10.1080/10447318.2020.1828535>
- Tallon, G., Blain, H., Seigle, B., Bernard, P. L., & Ramdani, S. (2013). Dynamical and stabilometric measures are complementary for the characterization of postural fluctuations in older women. *Gait & posture*, 38(1), 92-96. <https://doi.org/10.1016/j.gaitpost.2012.10.021>
- Varlet, M., Marin, L., Lagarde, J., & Bardy, B. G. (2011). Social postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 473.

Concept for a short-term training program to introduce the process of additive manufacturing in the space of Extended Reality (xR)

Moritz Heinrich, Torsten Döhler, Christopher Kracht, Klarissa Koppusch, Fabian Grund, Frank Briesemeister, Eckart Wolf, Andreas Foitzik, Rene Krenz-Baath, and Andrea Böhme

Wildau University of Applied Sciences, Wildau, Germany

Corresponding author: moheinrich@th-wildau.de

Keywords: Virtual Reality, Augmented Reality, Additive Manufacturing, Teaching, Inclusion, Digitalization, STEM Education

Introduction

The new xR (extended reality) technologies are advancing every year and are being used in more and more fields of application. The use of the technology has great potential in the teaching sector, especially in connection with online learning, which has been used more and more since the coronavirus. "MINT'oVation" is a project to present the processes of **additive manufacturing** by using xR for education. The integration of modern technologies in the classroom helps to convey and consolidate knowledge in an understandable way and brings with it a wide range of applications that appeal equally to different groups of students (Wehrmann and Zender, 2023). The whole project is structured in a short-term and a long-term part.

An overview of the short-term part is given below. This part is designed to train participants with haptic and visual xR processes to learn the physical principles of additive manufacturing in an immersive way. The results will be used as a preparation for the long-term part.

Methodology

The short-term measure of the project is divided into a total of five blocks. The individual teaching units serve to impart knowledge and familiarize students with the general AR and VR environment as well as the basics of metal technology in connection with additive manufacturing processes. The short-term measure will initially introduce a new, supportive learning method in the STEM subjects in the university TH Wildau (Germany). Based on a survey conducted from June to the end of July 2024 among staff, students and external collaborators of TH Wildau, the most wanted Aspects are (1) visualize various 3D models with interaction (71.4 %), (2) the use of xR technologies for learning advanced subjects (68.1 %), and (3) use as a preparatory training for a long-term program in this area (55.2 %) (Wildau et al., 2024).

The first teaching unit will consist of a coherent introductory presentation that introduces xR technologies, additive manufacturing using metal 3D printing and, in this context, different crystal lattice structures as the physical basis of metal printing.

The second block of the short-term activity is designed **to introduce** students and interested parties to xR technologies in a playful way. Using AR glasses or AR applications for mobile devices for a larger number of participants, the aim is to assemble a jigsaw puzzle that is haptically available as a real 3D print. A "devil's knot" puzzle is used as an introductory object. Participants should first try to solve the puzzle on their own before using the xR technologies to help them. Within the application, the parts of the devil's knot are highlighted in different colors in the form of instructions and the assembly of the knot is illustrated step by step. The combination of a simple puzzle with simple xR technologies is an optimal, interactive learning method to ensure a successful start in the subject area.

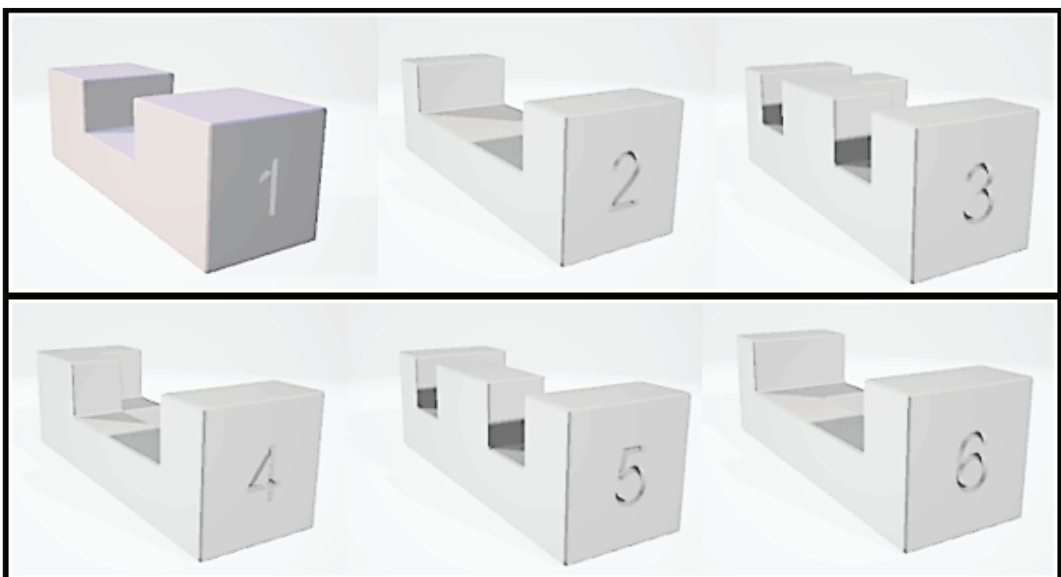


Figure P. 19. Numbered components of a 3D devil's knot

The third lesson focuses on illustrating metal lattice structures and their displacement, visualized using a combination of xR technology and magnetic spheres. The magnetic spheres show the structure of the atoms. The elastic and plastic deformation of metals and the effects on their internal structure are shown on film and can then be experienced haptically on the magnets. Step-by-step instructions on how to construct different shapes and the internal motion sequences when an external force is applied, combined with the actual sensory impression, are intended to improve understanding of the physical effects.

The fourth unit also deals with metal grid structures. The physical object is a specially designed cube structure with pins on half of the adjacent sides and guide grooves on the other sides (Figure P 20). Due to the special nature of the cube, a wide range of surface displacements are possible. Printed QR codes are used to implement an xR video that visualizes the relationship between the atomic lattice and the corresponding lattice structure. The basic knowledge about the physical processes during deformation is not only expanded and made visible, but also linked to reality.

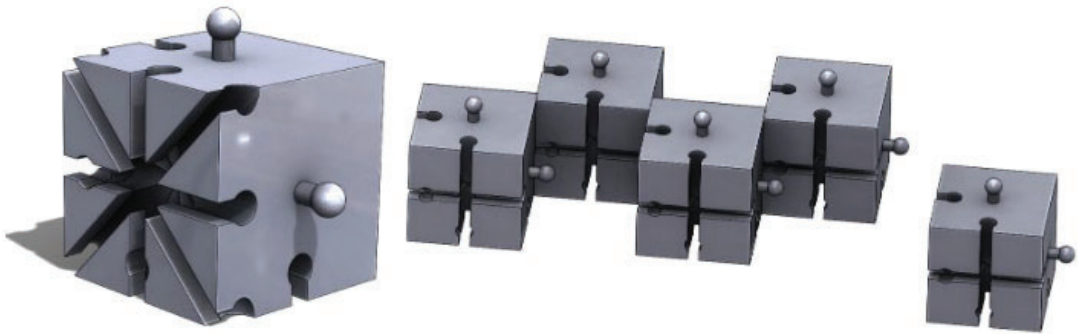


Figure P 20. Specially designed cube to illustrate the 3D grid structure with guides for moving entire levels.

The final part of the short-term measure aims to evaluate the participants' learning and interaction experiences. To do this, they fill out a digital form in the xR context, which is used to evaluate the measure.

The focus is then placed on preparing participants for a more advanced long-term program. This is intended to provide an in-depth introduction to metal physics, additive manufacturing and 3D metal printing. The aim is to consolidate the knowledge acquired and develop practical skills that are important in the STEM subjects and in industry.

Results

The introduction of the short-term measure aims to familiarize students and other interested parties with both additive manufacturing and xR technologies, given the increasing development of these technologies. The teaching unit will lay the foundations for the use of Extended Reality in teaching at the university and promote the expansion of the technology into other teaching areas. Furthermore, interactive courses will make studying STEM subjects more attractive and facilitate access to knowledge. The project aims to promote inclusion at the university and make the content of STEM subjects accessible to everyone¹⁷.

Within the overall project, a long-term method is also being developed to demonstrate the processes of design, production and quality control of metal 3D printing in extended reality. The short-term measures are also intended to motivate interested parties to participate in further learning units in order to promote the further integration of extended reality in production, research and, above all, teaching.

Challenges

The main challenges in creating appropriate learning content for xR technologies in additive manufacturing are the high technical complexity and access to suitable hardware for teaching. The use of xR glasses is also always associated with high costs. The learning content needs to be precise and easy for students to understand, while still providing a complete and physically correct representation of the complex content and metal structures. The content of each phase must be designed and

¹⁷ <https://integranxt.com/blog/extended-reality-xr-in-stem-learning/>

programmed as a unified learning experience to maximize student learning. These and other hurdles must be overcome for the successful implementation and acceptance of xR technologies in the teaching environment (Obeidallah et al., 2023).

Summary

The MINT'ovation project focuses on the use of Extended Reality (xR) for teaching the relatively new technology of additive manufacturing, especially for metal, to future-proof STEM subjects through innovative and practical teaching methods. In a five-part course, participants are progressively introduced to xR technologies and 3D metal printing. This methodology effectively combines theory and practice: Interactive learning units promote a deep understanding of physics and make complex content clear and tangible. The short-term course prepares the participants for an in-depth long-term course that provides advanced knowledge of metal physics, 3D printing and its industrial applications. The aim of the project is to make STEM education more attractive and sustainable, especially for groups of people who find it difficult to access these subjects, through the use of variable modern technologies (Stone et al., 2020).

Acknowledgements:

Supported by funds from the European Union and the State of Brandenburg.

References

- Obeidallah, R., Ahmad, A.A., & Qutishat, D. (2023). Challenges of Extended Reality Technology in Higher Education: A Review, *International Journal of Emerging Technologies in Learning*.
<https://doi.org/10.3991/ijet.v18i14.39871>
- Stone, B., Kay, D., Reynolds, A., & Brown, D. (2020). 3D Printing and Service Learning: Accessible Open Educational Resources for Students with Visual Impairment. *International Journal of Teaching and Learning in Higher Education*, 32(2), 336-346.
- Wehrmann, F., & Zender, R. (2023). Potenziale von Virtual Reality für inklusive Schulbildung.
<https://doi.org/10.18420/delfi2023-66>
- Wildau, T.H., Döhler, T., Heinrich, M. et. al. (2024). Survey "Results report questionnaire", Virtual and augmented reality using the example of additive manufacturing in teaching, survey period: 24.06.24 - 30.07.24.

A multimodal AR System: Enhancing Hiking Experiences

Leyi Hong, Wenjun Hou, Qian Li, and Meiyu Lv

Beijing University of Posts and Telecommunications, Beijing, China

Corresponding author: hwj1505@bupt.edu.cn

Keywords: Augmented Outdoor Sports, Multimodal Interaction, Augmented Reality and mobile devices, UI for XR headsets

Introduction

Amid escalating health consciousness, hiking has emerged as a prevalent pastime. Despite this, hikers encounter challenges such as equipment portability, extended treks, and dynamic environmental conditions. To tackle these, we've investigated incorporating Augmented Reality (AR) into outdoor expeditions, devising a multimodal AR system aligned with outdoor AR design principles. This system leverages both AR glasses and custom hiking poles for an enhanced interactive experience via diverse interaction modes and intuitive information display.

For an engaging AR experience, managing and rendering information is key to virtual element depiction. Trigger-based AR technologies include marker-, location-, and object-based systems, along with their hybrids (Edwards-Stewart et al., 2016). Studies have shown that outdoor AR interface visibility can be markedly improved through color blending (Ang, S. Y et al., 2020) and correction techniques (Gabbard et al., 2020). Moreover, multimodal interactions, integrating diverse inputs like voice and physical gestures with environmental awareness and gaze tracking (Zhao et al., 2023), are instrumental for complex AR tasks, enhancing virtual environment engagement.

Nevertheless, the domain of hiking, characterized by prolonged and extensive natural environment exposure, has remained underserved in terms of tailored research and interaction design for AR applications. Therefore, this paper proposes an AR-based optimization system for outdoor hiking experiences, introduces its implementation and evaluation results, and provides reference content for the design of AR systems in outdoor sports scenarios.

Method

The cardinal research goal is to explore the application of AR technology in outdoor hiking scenarios. Unlike the application of two-dimensional user interfaces in ergonomics, layout, and user experience, AR user interface design must take into account more factors such as the environment, space, and interaction entities (Krauß et al., 2021). At the same time, the unique scene characteristics of outdoor sports, such as dynamically changing background textures, lighting conditions, and environmental colors, pose challenges to the application of AR technology. Against this backdrop, the project has

formulated the following methods to achieve rational research, design, and development of outdoor AR systems.

- User research to define system functions.
- Develop AR-UI strategies for environmental adaptability.
- Establish Multimodal interaction standards for natural user experiences.
- Prototype development in Unity.
- Pilot testing and usability assessments.
- Iterative refinement based on feedback.

Implementation

We have constructed an AR hiking enhancement system prototype, leveraging Unity 2022 and the Oculus Quest Pro. The system is primarily designed to operate on AR glasses (Figure P 21), supplemented by a modified interactive hiking trekking pole (replaced by the right-hand controller in this project). The AR glasses are equipped with eye-tracking sensors that allow users to interact with virtual objects through gaze, as depicted in Figure P 22, reducing the reliance on physical movements. To enhance interaction, gesture controls have been integrated for additional engagement with objects initially selected via gaze (Figure P 23). Acknowledging the common use of trekking poles among hikers, an alternative to gesture control has been devised: a pole controller featuring physical buttons and a scroll wheel for command input through radio frequency signals.

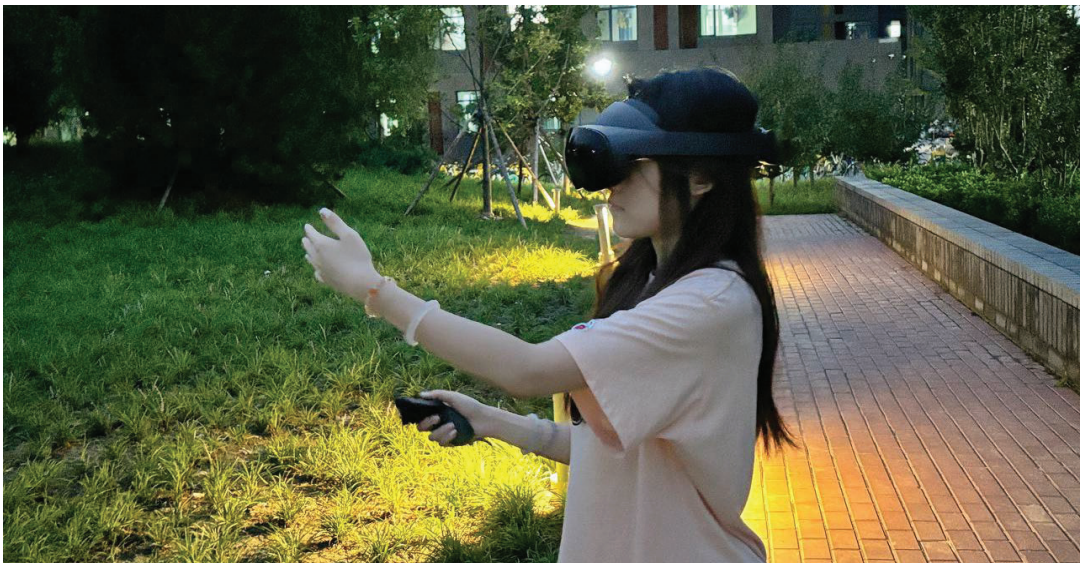


Figure P 21. System testing in natural environments



Figure P 22. Interact with virtual objects by gazing



Figure P 23. Gesture recognition evokes flat maps

Test and Evaluation

User testing and focus group discussions were conducted to assess the system's usability and user-friendliness in outdoor settings. The findings revealed a preference for trekking pole interaction, which is more suited to the outdoor context. While gesture interaction was appreciated for its intuitiveness, it was found to be limiting in complex terrains and could cause fatigue. Gaze interaction was universally accepted. The system exhibited high user-friendliness in outdoor adaptability, operational naturalness, information recognizability, layout rationality, and natural integration, aligning with the design strategies

and development principles. However, some participants noted challenges with distance perception, indicating a need for improvement in future iterations. Overall, the system design met the expectations and needs of the participants.

Discussion and Conclusion

This study developed an outdoor AR hiking experience optimization system using Unity and the Oculus Quest Pro, aiming to enhance users' perception and experience in natural environments through on-site navigation, environmental perception, and health monitoring. User testing confirmed the rationality and usability of the system design, revealing the usability challenges and user preferences for various modalities of interaction in outdoor scenarios. Despite hardware limitations, the research provides valuable insights into the application of AR-UI in outdoor hiking contexts. With advancements in technology and hardware, the system is anticipated to expand to more complex outdoor hiking or other sports scenarios in the future.

References

- Ang, S. Y., & Ng, G. W. (2020, December). A study on the effect of the real-world backgrounds using colour blending technique on optical see-through AR user interface design. In 2020 6th International Conference on Interactive Digital Media (ICIDM) (pp. 1-4). IEEE.
- Edwards-Stewart, A., Hoyt, T., & Reger, G. (2016). Classifying different types of augmented reality technology. *Annual review of cybertherapy and telemedicine*, 14, 199-202.
- Gabbard, J. L., Smith, M., Tanous, K., Kim, H., & Jonas, B. (2019). AR drivesim: An immersive driving simulator for augmented reality head-up display research. *Frontiers in Robotics and AI*, 6, 98. <https://doi.org/10.3389/frobt.2019.00098>
- Krauß, V., Boden, A., Oppermann, L., & Reiners, R. (2021, May). Current practices, challenges, and design implications for collaborative AR/VR application development. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-15). <https://doi.org/10.1145/3411764.3445335>
- Zhao, G., Orlosky, J., Gabbard, J., & Kiyokawa, K. (2023). HazARdSnap: gazed-based augmentation delivery for safe information access while cycling. *IEEE Transactions on Visualization and Computer Graphics*.

Immersive Cocooning – Competition between Control and Convenience

Wolfgang G. Schäfer, and Gioia Volkmar

ZHAW, Institute of Marketing Management, Winterthur, Switzerland

Corresponding author: wolfgang.schaefer@zhaw.ch

Keywords: Artificial Intelligence, Cocooning, Virtual Reality.

Abstract

The new framework of "immersive cocooning" of consumers in virtual worlds shows the tension between self-determined and externally determined cocooning. We point out the threat of mental prisons through virtual platforms, which may jeopardize the self-determination of consumers. This research is the response to a call by the EU Commission, where it has already been successfully presented as a draft. The presentation as a poster at EuroXR 2024 should further support this ongoing research and develop it into a publication.

Introduction

In consumer research, the metaphor of the nomad (Bardhi and Eckhardt 2017; Bardhi, Eckhardt and Arnould 2012) is a central research topic to examine consumption in the context of the dynamic and complex world of our liquid modernity (Bauman 2000). Current research on digital nomadism explores consumers' desire for security, which has emerged from experiences of loss, such as pandemics (Atanasova, Eckhardt and Husemann 2024). In this study, digital nomadism is defined as *"individuals leverage remote and digital working to serially migrate between different locations, seeking [...] a better quality of life"* (2024, 1244).

This contribution, however, proposes a new conceptual framework that combines digital nomadism and the search for security - immersive cocooning. The framework locates consumers within the sphere of influence of virtual reality platforms to illustrate the influence of such platforms on them as well as consumers' search for safety, security, and separation from the world outside virtual reality.

Our research has been motivated by the observation that viewers of the movie "Avatar" found that computer-generated worlds to be so beautiful that they became depressed when returning to their usual physical world (Spurr 2022). This observation is like the phenomenon of cocooning (Slaughter and Grigore 2015; Zalega 2018). Cocooning is the act of withdrawing into a private space and thus excluding oneself from the rest of the world – a form of escapism.

The novelty of this framework is that consumers can withdraw into computer-generated worlds as if into an immersive cocoon - a cocoon in which they can still move around like digital nomads. Moreover, we also recognize that due to the high density of recordable data points in virtual worlds, the platforms that run them have complex databases that allow for the "Manufacturing [of] Customers" (Zwick and Denegri Knott 2009). We have concerns that, on the basis of such data, this phenomenon previously initiated by consumers can gradually be actively pursued by platforms and their client corporations, e.g. by creating products based on customer behaviour recorded in virtual reality (Gao and Liu 2023; Haleem et al. 2022; Khan et al. 2022).

Moreover, the use of AI in virtual worlds, e.g. AI-based avatars, may lead consumers to develop trusting relationships with such representations of AI (Huang, Kim and Lennon 2022). Such stimulation by AI in virtual worlds can lead to a loss of autonomy for consumers (Gonçalves et al. 2024; Kim 2022), which we understand as externally determined cocooning in the sense of the mental prison mentioned above. The protection of this autonomy is not only central to the protection of consumers as representatives of democratic societies (Bjørlo, Moen and Pasquine 2021), but, in our view, fundamentally challenges competition in virtual worlds under the influence of generative AI.

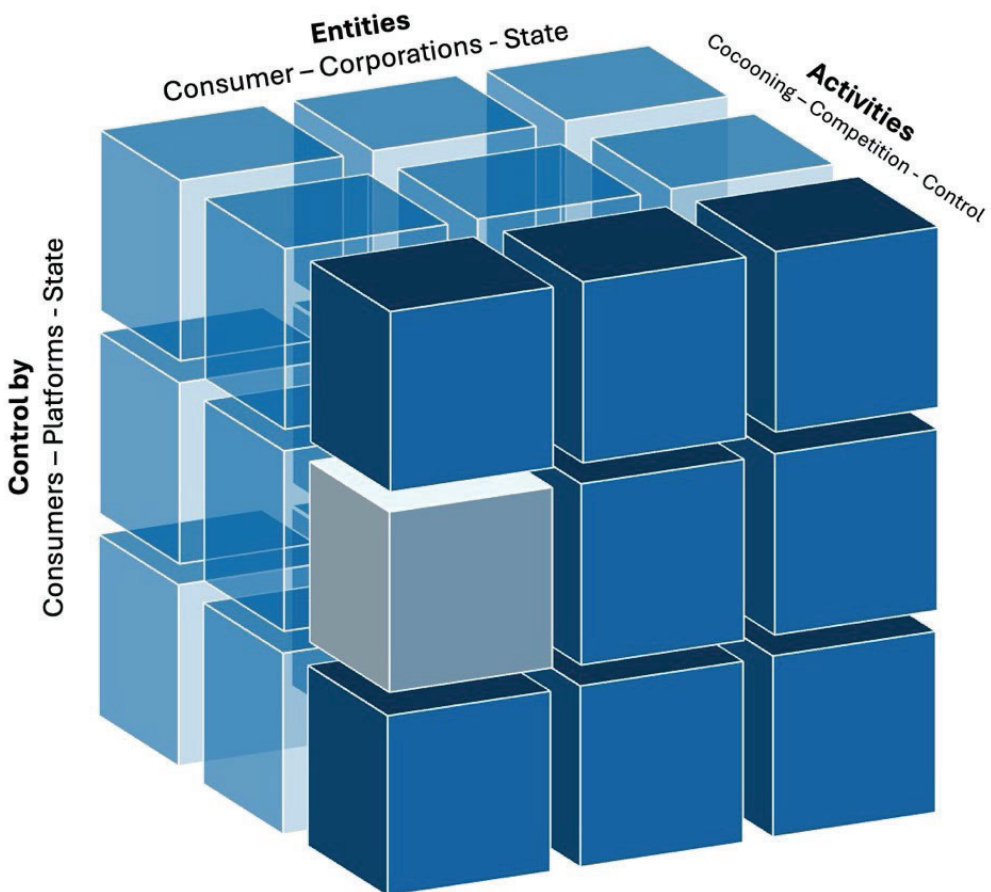


Figure P 24. Immersive Cocooning Cube

Three axes - entities, activities and control - define the dimensional space of this framework. From the point of view of practice theory (Reckwitz 2002; Røpke 2009), the activities of the actors involved are to be placed in relation to their power and control from the point of view of power theory (Foucault 1995; Schwan and Shapiro 2011). In short, the aim is to show which actors can exercise or are subject to power and control in this structure. In addition to consumers, the actors involved are companies and the State as a legislative and controlling entity, with companies referring to platforms, media, (advertising) companies and other organisations. The framework also addresses key issues such as barriers to entry, standards, and the influence of generative AI that can be used to customize entire worlds and operate robotic avatars.

1st axis: Entities. In this framework, consumers consume virtual worlds as experiences. Consumers are characterized by the consumption of products and services, whereby this consumption can be an experience for them, as well as social integration or classification (Holt 1995). Those virtual worlds are operated by companies as platforms on which other companies offer products and services to consumers and influence them with advertising content. And finally, we see the State as an entity that can control commercial competition in virtual worlds with the legal resources at its disposal.

2nd axis: Activities. We describe a consumer activity we could observe as 'cocooning', in which consumers surround themselves in a virtual world with both self-selected and commercially provided experiences, protecting themselves from the uncertainties of a physical lifeworld (Blackshaw 2010, 147). This activity can involve the formation of close social groups that stabilise the virtual consumer identity (Belk 2014; Nechvatal 2001) through shared social values and consumption practices (Habuchi 2005). By convenience, we mean activities that are brought to consumers in virtual worlds by corporations, facilitate consumption through the use of AI technologies (Ameen et al. 2021; He and Zhang 2023; Thakur, Bandyopadhyay and Datta 2023) and can therefore also influence cocooning from the outside. Competition in turn comprises activities between corporations, e.g. a platform, companies advertising on the platform or companies that have yet to access a platform, and the virtual worlds available there (Rietveld and Schilling 2021; Weiss and Schiele 2013).

3rd axis: Control. The control exercised by technologies over consumers (Beckett 2012) as well as by consumers over technologies (Pizzi, Scarpi and Pantano 2021) is an area of tension that has been analysed in detail yet. Consumer self-control can be understood as a mental prison in the Foucauldian sense, as in virtual worlds the collection of data can holistically encompass all consumer activities without the need to consciously identify a guardian. This mental prison, in turn, can be controlled by platforms, and the use of AI, e.g. for the customised design of entire virtual worlds ("Canopy" 2023) or the control of avatars ("Convai" 2024), can intensify such control.

The tension that can arise between the consumer's externality and self-determination in immersive cocooning is a broad research agenda that demands further exploration using new research methods such as immersive netnography (Kozinets 2023). Hence, future research will collect and analyse empirical data to further refine this framework. In addition to expert interviews, this will include interviews with users of metaverse platforms, netnographic observations in the sense of a netnography (Kozinets 2019, 2024), and empirical reports on the use of metaverse platforms. As we anticipate that immersive experiences and worlds will increasingly define our lives, we hope to make a useful contribution with this framework. The aim is to publish the results in an academic journal.

References

- Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. *Computers in human behavior*, 114, 106548. <https://doi.org/10.1016/j.chb.2020.106548>
- Atanasova, A., Eckhardt, G.M., & Husemann, K.C. (2024). Liquid Consumer Security. *Journal of Consumer Research*, 50(6), 1243–64. <https://doi.org/10.1093/jcr/ucad047>
- Bardhi, F., & Eckhardt, G. M. (2017). Liquid Consumption. *Journal of Consumer Research*, 44(3), 582–97. <https://doi.org/10.1093/jcr/ucx050>
- Bardhi, F., Eckhardt, G. M., & Arnould, E. J. (2012). Liquid relationship to possessions. *Journal of Consumer Research*, 39(3), 510–529. <https://doi.org/10.1086/664037>
- Bauman, Z. (2000), *Liquid Modernity*, Cambridge: Polity Press. <https://doi.org/10.5565/rev/athenead/v1n0.10>
- Beckett, A. (2012). Governing the Consumer: Technologies of Consumption. *Consumption Markets & Culture*, 15(1), 1–18.
- Belk, R. (2014). Digital Consumption and the Extended Self. *Journal of Marketing Management*, 30(11–12), 1101–18.
- Bjørlo, L., Øystein, M., & Pasquine, M. (2021). The Role of Consumer Autonomy in Developing Sustainable AI: A Conceptual Framework. *Sustainability*, 13(4), 2332. <https://doi.org/10.3390/su13042332>
- Blackshaw, T. (2010). Nostalgia. in *Key Concepts in Community Studies*, London: Sage, 145–50.
- Canopy (2023), <https://canopy.procedural-worlds.com/files/file/32-ai-game-development-toolkit-gaia-ml/>
- Convai (2024), <https://www.convai.com> .
- Foucault, M. (1995), *Discipline & Punish*, New York, NY: Vintage Books.
- Gao, Y., & Liu, H. (2023). Artificial Intelligence-Enabled Personalization in Interactive Marketing: A Customer Journey Perspective. *Journal of Research in Interactive Marketing*, 17(5), 663–80.
- Gonçalves, A. R., Pinto, D. C., Shuqair, S., Dalmoro, M., & Mattila, A. S. (2024). Artificial intelligence vs. autonomous decision-making in streaming platforms: A mixed-method approach. *International Journal of Information Management*, 76, 102748. <https://doi.org/10.1016/j.ijinfomgt.2023.102748>
- Habuchi, I. (2005). Accelerating Reflexivity. in *Personal, Portable, Pedestrian*, ed. Mizuko Ito, Daisuke Okabe, and Misa Matsuda, London: The MIT Press, 165–82. <https://doi.org/10.7551/mitpress/5309.003.0011>
- Haleem, A., Javaid, M., Asim, M., Qadri, M.A., Singh, R.P., & Suman, R. (2022). Artificial Intelligence (AI) Applications for Marketing: A Literature-Based Study. *International Journal of Intelligent Networks*, 3, 119–32.
- He, A-Z., Zhang, Y. (2023). AI-Powered Touch Points in the Customer Journey: A Systematic Literature Review and Research Agenda. *Journal of Research in Interactive Marketing*, 17(4), 620–39. <https://doi.org/10.1108/JRIM-03-2022-0082>
- Holt, D. B. (1995). How Consumers Consume: A Typology of Consumption Practices. *Journal of Consumer Research*, 22(1), 1–16. <https://doi.org/10.1086/209431>
- Huang, R., Kim, M., Lennon, S. (2022). Trust as a Second-Order Construct: Investigating the Relationship between Consumers and Virtual Agents. *Telematics and Informatics*, 70, 101811.
- Khan, S., Tomar, S., Fatima, M., Khan, M.Z. (2022). Impact of Artificial Intelligent and Industry 4.0 Based Products on Consumer Behaviour Characteristics: A Meta-Analysis-Based Review. *Sustainable Operations and Computers*, 3, 218–25. <https://doi.org/10.1016/j.susoc.2022.01.009>
- Kim, W. (2022). Shopping with AI: Consumers' Perceived Autonomy in the Age of AI. in *Human-Centered Artificial Intelligence*, Elsevier, 157–71.
- Kozinets, R.V. (2019). *Netnography: The Essential Guide to Qualitative Social Media*”, 3rd ed., London: Sage.

- Kozinets, R.V. (2023). Immersive Netnography: A Novel Method for Service Experience Research in Virtual Reality, Augmented Reality and Metaverse Contexts. *Journal of Service Management*, 34(1), 100–125.
- Kozinets, R.V. (2024). Researching AI chatbots, platforms and the metaverse: understanding today's netnography. ch. 17, In: Belk, R.W. & Otnes, C., *Handbook of Qualitative Research Methods in Marketing*, Glos, UK: Edward Elgar Publishing Ltd., 185–196.
- Nechvatal, J. (2001). Towards an Immersive Intelligence. *Leonardo*, 34(5), 417–22. <https://doi.org/10.1162/002409401753521539>
- Pizzi, G., Scarpi, D., & Pantano, E. (2021). Artificial Intelligence and the New Forms of Interaction: Who Has the Control When Interacting with a Chatbot? *Journal of Business Research*, 129, 878–90. <https://doi.org/10.1016/j.jbusres.2020.11.006>
- Reckwitz, A. (2002). Toward a Theory of Social Practices. *European Journal of Social Theory*, 5(2), 243–63. <https://doi.org/10.1177/13684310222225432>
- Rietveld, J., & Schilling, M.A. (2021). Platform Competition: A Systematic and Interdisciplinary Review of the Literature. *Journal of Management*, 47(6), 1528–63. <https://doi.org/10.1177/0149206320969791>
- Røpke, I. (2009). Theories of Practice — New Inspiration for Ecological Economic Studies on Consumption. *Ecological Economics*, 68(10), 2490–97.
- Schwan, A., & Shapiro, S. (2011). *Foucault's Discipline and Punish, How to read*, London: Pluto Press. <https://doi.org/10.22439/fs.v0i15.3997>
- Slaughter, S., & Grigore, G. (2015). Exploring the Post-Recession Experiences of the "Cocooning" Consumer: The Importance of Core Values and Rituals. *Journal of Promotional Communications*, 3(1), 242–57.
- Spurr, A. (2022). The Avatar Effect – And How to Beat It. FM 96.5 -Christian Living, <https://www.96five.com/stories/faith/christian-living/2022/the-avatar-effect-and-how-to-beat-it/>.
- Thakur, S. S., Bandyopadhyay, S., & Datta, D. (2023). Artificial intelligence and the metaverse: Present and future aspects. In *The Future of Metaverse in the Virtual Era and Physical World* (pp. 169-184). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-29132-6_10
- Weiss, T., & Schiele, S. (2013). Virtual Worlds in Competitive Contexts: Analyzing ESports Consumer Needs. *Electronic Markets*, 23(4), 307–16. <https://doi.org/10.1007/s12525-013-0127-5>
- Zalega, T. (2018). Consumption Cocooning in Consumer Behaviour of Polish Seniors (Report from Own Research). *Handel Wewnętrzny*, 4(375), 410–22.
- Zwick, D. & Denegri Knott, J. (2009), Manufacturing Customers. *Journal of Consumer Culture*, 9(2), 221–47. <https://doi.org/10.1177/1469540509104375>

Architecture proposal of a VR operating procedure system for construction training based on BIM and QHSE rules

Barbara Schiavi¹, Vincent Havard², Karim Beddiar³, and David Baudry²

¹ VINCI Construction, Nanterre, 92000, France

² LINEACT, CESI, Saint-Étienne-du-Rouvray, 76800, France

³ LINEACT, CESI, Nantes, 44000, France

Corresponding authors: barbara.schiavi@live.fr, vhavard@cesi.fr, kbeddiar@cesi.fr, d baudry@cesi.fr

Keywords: Architecture, VR, BIM, QHSE rules, SWRL, web semantic, training system, AEC

Abstract

This research fosters collaboration in developing VR training simulations for construction procedures using semantic web technology. It introduces SWRP, an intuitive, web-based graphic editor for QHSE rule editing, addressing the difficulties that construction experts face with SWRL. The main contribution is the VROPTS architecture, which integrates BIM, VR, and QHSE rules, as detailed in the results section, offering the potential to revolutionize collaborative construction and training practices through VR.

Introduction

In recent advancements within the construction industry, the integration of Building Information Modeling (BIM) with augmented and virtual reality (AR/VR) technologies has been a focal point of innovation (Hadavi & Alizadehsalehi, 2024). (Schiavi et al., 2022b) propose a comprehensive review that shows a significant gap in the literature regarding worksite operations and risk prevention and safety training using VR tools. Addressing this, (Schiavi et al., 2021a) developed a semi-automatic system for generating operational procedure scenarios in VR, leveraging 4D BIM and procedural knowledge. This system, underpinned by a UML model, facilitates the creation of realistic VR simulations for construction site training, allowing for the easy customization of procedural scenarios.

A key enhancement of this system is the incorporation of Quality, Health, Safety, and Environment (QHSE) considerations into the ifcBuildingElement, fostering the automatic generation of these aspects in new scenarios. (Schiavi et al., 2022a) presented an innovative no-code tool, the Virtual Reality Training Scenario Editor (VRTSE), that empowers experts to intuitively add, modify, or remove tasks within these scenarios, bridging the gap between BIM data and VR training with a focus on QHSE rules.

To further this endeavor, (Schiavi et al., 2021b) introduced an architecture that ensures interoperability between BIM operational procedures and VR simulations, anchored by the Safety BIM

Ontology (SBO). This ontology-based system maps BIM components to safety concepts and employs SWRL rules for conducting safety inspections. The efficacy of the SBO was validated through the development of an automatic safety risk identification and prevention mechanism, with its effectiveness confirmed via three distinct use cases.

Recognizing the necessity for a more accessible approach to rule management within ontologies, this research paper introduces a web-based graphical editor designed to simplify the creation and modification of SWRL rules for those not well-versed in semantic web technologies, named SWRP (Semantic Web Rules Platform). The fundamental works that influenced its user-friendly interface and the extensive system for modifying QHSE rules has been explored. SWRP serves as a bridge, connecting the ontology—rich with QHSE knowledge expressed through semantic web technology—with various users, including QHSE experts and design engineers. This connection could foster, collaborative, and real-time interaction within the VR training simulation environment.

The main contribution based on these findings is a final architecture that integrates BIM, VR and QHSE rules, enabling seamless data exchange for creating VR training simulations. This system, called VROPTS - Virtual Reality Operating Procedure Training System, is presented in the result section of this paper.

Related works

In the literature, various papers have studied the integration of web semantic technologies with VR, such as advanced visualization of molecular data in immersive environments (Trellet et al., 2018), VR modeling for Assisted Living, and VR training and simulations (Spoladore et al., 2017). Ontologies can serve as knowledge bases for creating VR training procedures in manufacturing (Havard et al., 2017). A proposal for semantic modeling of VR training in industrial contexts was made (Flotyński & Walczak, 2017), and a study developed a VR training application using an IFC ontology to prevent hazards by creating random experiences (Dris et al., 2019).

However, rule editing with a graphics tool is necessary. Therefore, different methods of visualizing SWRL rules graphically have been reviewed, to enhance the understanding of the language through graphical representation. Also, the creation of graphical editors for making, modifying and removing SWRL rules has been studied.

A comparison for each of the papers reveals: the different approaches examined, what kind of data architecture is used, which technologies are employed, and whether the solutions offered are usable or reproducible with the information given. A focus was made on four research papers that enable graphical editing of SWRL rules: (Orlando et al., 2012) have created a Web-Protected plugin for visualizing, creating and editing SWRL rules; (Bak et al., 2013) suggested a graphical visualization of an ontology and SWRL rules before and after reasoning as well as a graphical tool for SWRL rules creation; (Fill et al., 2017; Pittl & Fill, 2020) have developed a visual modeling language for SWRL; (Bolock et al., 2020) have designed a system of connecting boxes containing SWRL concepts for creating SWRL rules.

In this regard, one of the crucial features of a graphical editor of SWRL rules is the validation of links between entities (directions and types), the cardinality check and the consistency check. A vital feature that is missing in all the graphical editors studied in this work is the ability to edit a rule that has been previously created in the ontology or created via the editor interface. The functionality provided by (Fill et al., 2017; Pittl & Fill, 2020) for exporting SWRL rules to other platforms such as Protégé is noteworthy, but it has been used only for checking ontology and SWRL rules outside the graphical editor as their editor does not allow this. The ability to add atoms to rules in SWRL notation is a useful feature for

representing properties with more than two parameters, for example. However, it cannot be used by anyone not familiar with SWRL. Thus, the scientific challenge is to propose a graphic editor for QHSE rule editing, on BIM ontology, to be used in VR training simulations. The outcomes of this platform are discussed in the following section.

The web-based platform

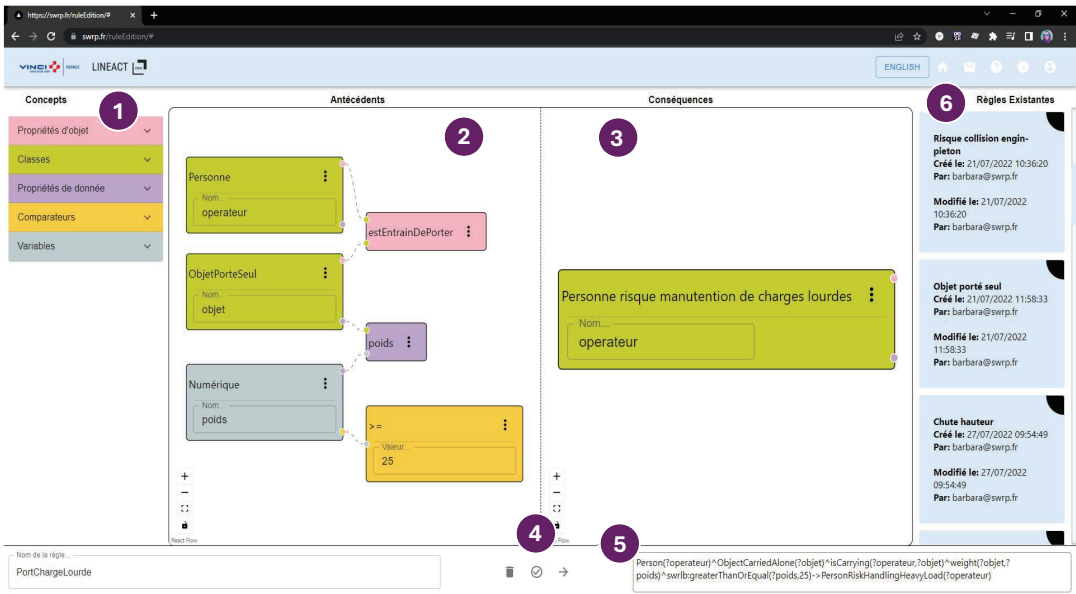


Figure P 25. SWRP interface, with a QHSE rule, indicating as an antecedent that if “A single person carrying an object weighing 25 kg or more” therefore “A person risks handling heavy loads”.

Managing complex formal systems in industrial and construction settings requires applications that non-expert users of the semantic web can use. The developed graphical SWRL rule editor, named SWRP (Semantic Web Rule Platform) and shown in Figure P 25, follows the state of the art and has these features: See all entities with their hierarchy; Create rules graphically; Edit rules from the ontology or the interface; Delete rules; Show rules from the ontology; Rename entities; Check links between entities (directions and types); Check cardinality and consistency.

To clarify, the numbers displayed on the SWRP interface in Figure P 25 correspond to these functionalities: the concept blocks (1) are in the menus and can be dragged and dropped into the rule's antecedent (2) or consequence (3). After filling in the blocks and naming the rule, a check button (4) shows the rule in SWRL (5) and send it to the ontology server. The SWRL rule lets language experts verify that the GUI-edited rule matches the intended SWRL rule. Previous rules (6) are in the right panel and can be edited or deleted.

Results

The graphical editor SWRP has been integrated in a final system architecture, called VROPTS (Virtual Reality Operating Procedure Training System), enabling interoperability between BIM, VR and QHSE rules and presented in Figure P 26. A use case has been studied and validated based on the QHSE rule shown in the Figure P 25.

The different areas of expertise are represented in the VROPTS. These knowledge bases feed and manage the simulation structure. They are supported by two editors, one for editing procedure scenarios (VRTSE) and one for graphically editing QHSE rules (SWRP). This system architecture makes it easy to add, modify or delete knowledge bases as required. Indeed, the ontology (SBO) can be gradually adjusted to match the learner's improvement. This allows the training to be tailored to each user's individual requirements.

Future iterations of the system could independently assess the trainee's level, allowing for automatic updates to the ontology requirement. Moreover, the VR simulation could be adapted to other ontologies, for example for off-site assembly operations from industrial prefabrications, DfMi, etc. (Li et al., 2015).

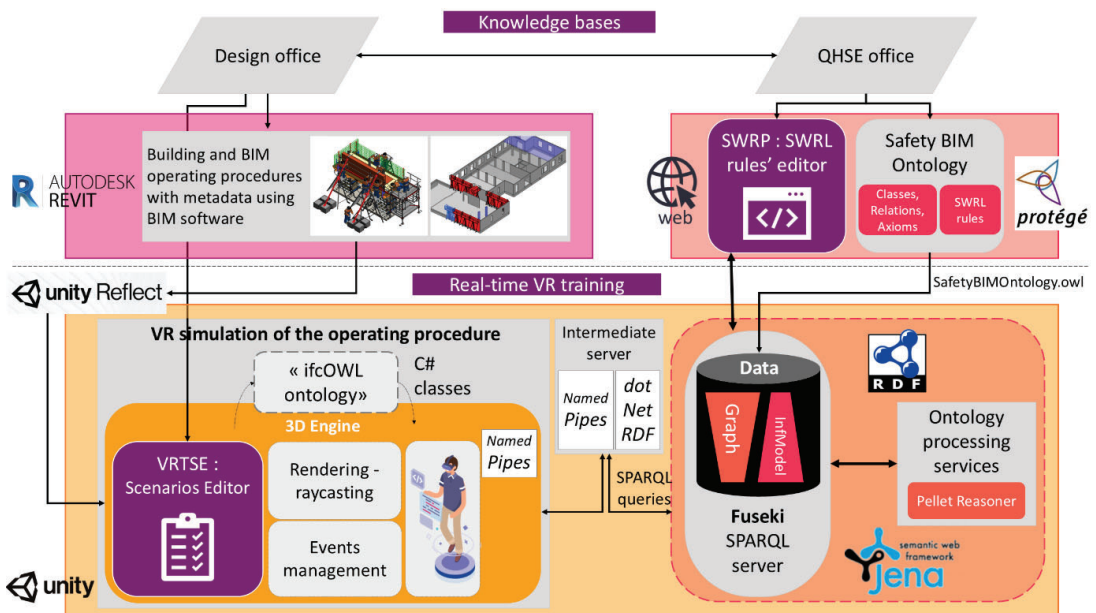


Figure P 26. Proposed architecture of an interoperable system between BIM - VR and QHSE rules

References

- Bak, J., Nowak, M., & Jedrzejek, C. (2013, July). Graph-based Editor for SWRL Rule Bases. In RuleML (2).
- Dris, A.-S., Lehericey, F., Gouranton, V., Arnaldi, B. (2019). OpenBIM Based IVE Ontology: An Ontological Approach to Improve Interoperability for Virtual Reality Applications. In: Mutis, I. and Hartmann, T. (eds.) Advances in Informatics and Computing in Civil and Construction Engineering. pp. 129-136. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-00220-6_16.
- El Bolock, A., Mohamed, I., Herbert, C., & Abdennadher, S. (2020). A Visual Rule Generation Tool for SWRL. In RuleML+ RR (Supplement) (pp. 58-72).
- Fill, H.-G., Pittl, B., & Honegger, G. (2017). A Modeling Environment for Visual SWRL Rules Based on the SeMFIS Platform. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10243 LNCS (pp. 452-456). https://doi.org/10.1007/978-3-319-59144-5_30

- Flotyński, J., Walczak, K. (2017). Ontology-Based Representation and Modelling of Synthetic 3D Content: A State-of-the-Art Review. *Comput. Graph. Forum.* 36, 329–353. <https://doi.org/10.1111/cgf.13083>.
- Hadavi A., & Alizadehsalehi S. (2024). From BIM to metaverse for AEC industry, *Automation in Construction*, 160, 105248. <https://doi.org/10.1016/j.autcon.2023.105248>.
- Havard, V., Benoit, J., Xavier, S., David, B. (2017). Inoovas - Industrial ontology for operation in virtual and augmented scene: The architecture. In: 4th International Conference on Control, Decision and Information Technologies, CoDIT 2017. pp. 300–305. IEEE (2017). <https://doi.org/10.1109/CoDIT.2017.8102608>.
- Havard, V., Jeanne, B., Lacomblez, M., & Baudry, D. (2019). Digital twin and virtual reality: a co-simulation environment for design and assessment of industrial workstations. *Production and Manufacturing Research*, 7(1), 472–489. <https://doi.org/10.1080/21693277.2019.1660283>
- Li, H., Lu, M., Chan, G., & Skitmore, M. (2015). Proactive training system for safe and efficient precast installation. *Automation in Construction*, 49, 163–174. <https://doi.org/10.1016/j.autcon.2014.10.010>
- Orlando, J. P., Rivolli, A., Hassanpour, S., O'Connor, M. J., Das, A., & Moreira, D. A. (2012). SWRL Rule Editor - A Web Application as Rich as Desktop Business Rule Editors. *Proceedings of the 14th International Conference on Enterprise Information Systems, 2 ISAS(SAIC/-)*, 258–263. <https://doi.org/10.5220/0003999402580263>
- Pittl, B., & Fill, H.-G. (2020). A visual modeling approach for the Semantic Web Rule Language. *Semantic Web*, 11(2), 361–389. <https://doi.org/10.3233/SW-180340>
- Schiavi, B., Havard, V., Beddiar, K., & Baudry, D. (2021a). Semi-automatic generation of virtual reality procedural scenarios for operation in construction based on 4D building information models. In *Proceedings of 21st International Conference on Construction Application of Virtual Reality* (pp. 104–111).
- Schiavi, B., Havard, V., Beddiar, K., & Baudry, D. (2021b). VR Simulation of Operating Procedure in Construction Based on BIM and Safety Ontology: A Proof of Concept. *EuroXR Conference 2021*, 126–141. https://doi.org/10.1007/978-3-030-90739-6_8
- Schiavi, B., Havard, V., Beddiar, K., & Baudry, D. (2022a). A VR training scenario editor for operation in construction based on BIM 4D and domain expert authoring. *Smart and Sustainable Built Environment*. <https://doi.org/10.1108/SASBE-06-2022-0125>
- Schiavi, B., Havard, V., Beddiar, K., & Baudry, D. (2022b). BIM data flow architecture with AR/VR technologies: Use cases in architecture, engineering and construction. *Automation in Construction*, 134, 104054. <https://doi.org/10.1016/j.autcon.2021.104054>
- Spoladore, D., Arlati, S., & Sacco, M. (2017). Semantic and Virtual Reality-Enhanced Configuration of Domestic Environments: The Smart Home Simulator. *Mob. Inf. Syst.* 1–15 (2017). <https://doi.org/10.1155/2017/3185481>.
- Trellet, M., Férey, N., Flotyński, J., Baaden, M., & Bourdot, P. (2018). Semantics for an Integrative and Immersive Pipeline Combining Visualization and Analysis of Molecular Data. *J. Integr. Bioinform.* 15, <https://doi.org/10.1515/jib-2018-0004>.

XR and haptic feedback for electrician training: Ferrexpo case study

Maryna Tymchenko

Sensorama Lab, Kyiv, Ukraine

Corresponding author: maryna.tymchenko@sensoramalab.com

Keywords: XR, Haptic Feedback, Electrician Training, Energy Sector, Simulation, Ferrexpo, Virtual Reality

Abstract

This paper presents the use of XR and haptic feedback for electrician training at Ferrexpo, a mining factory in Ukraine. In this training module, trainees interact with a simulated environment to practice oil switch operations. A haptic feedback suit simulates electric shock when users make critical errors, such as operating in a live compartment or causing a short circuit, providing real-time consequences for mistakes in exam mode. In addition to the training module, an instructor app allows for user management and performance tracking, ensuring thorough evaluation and skill development. This approach highlights the practical benefits of immersive technologies for safety-critical environments in the energy sector.

Recent studies in the energy sector have shown that XR and haptic feedback significantly improve trainee engagement, safety, and retention of complex technical skills. Bairaktarova et al. (2023) note that "the integration of VR, AR, and wearable and haptic devices into learning environments creates immersive user experiences, thus enhancing user engagement." Alqallaf and Ghannam (2023) found that virtual reality "effectively enhances experiential learning, fostering deeper engagement and improved knowledge retention through hands-on simulations" in photovoltaic energy education. These findings support the implementation of XR and haptic technologies for safety-critical tasks, such as those demonstrated in the Ferrexpo case.

References

- Alqallaf, N. & Ghannam, R. (2023). Immersive learning in photovoltaic energy education: A comprehensive review of virtual reality applications. *Solar* 2024, 4(1), 136-161.
- Bairaktarova, D., Valentine, A. & Ghannam, R. (2023). The use of extended reality (XR), wearable, and haptic technologies for learning across engineering disciplines. In book: *International Handbook of Engineering Education Research* (pp. 501-524).



Figure P 27. 3D visualisation of a room with switchgear compartments.

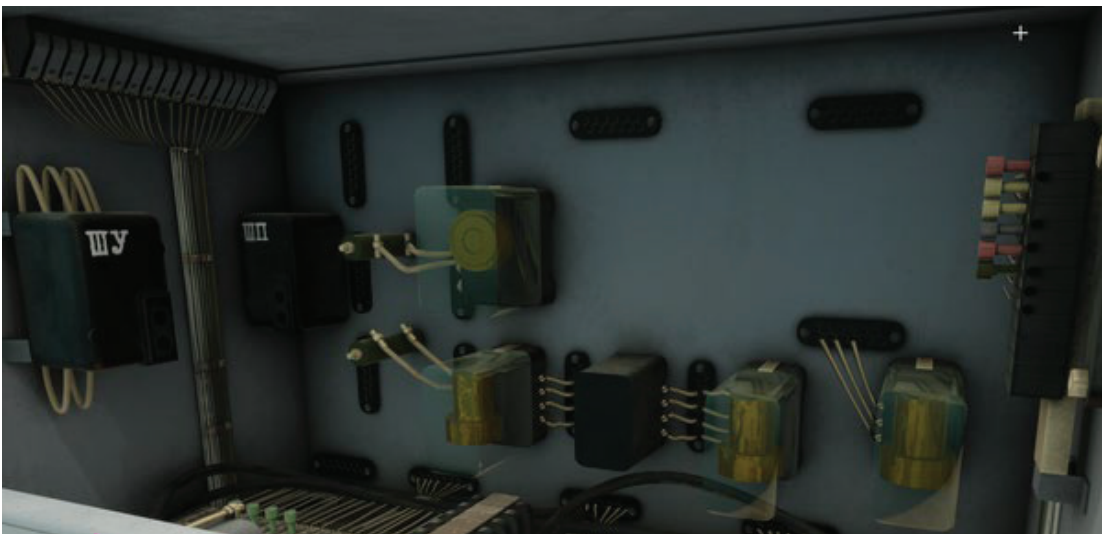


Figure P 28. The relay cabinet with control bus circuit breakers and power bus circuit breakers.

Beyond human imagination: The art of creating prompt-driven 3D scenes with Generative AI

Giulio Federico, Fabio Carrara, Giuseppe Amato, and Marco Di Benedetto

CNR-ISTI, Pisa, Italy

Corresponding author: giulio.federico@isti.cnr.it

Keywords: Generative AI, Computer Graphics, Denoising Diffusion Probabilistic Model, Gaussian Splatting, NeRF, Signed Distance Field, Video Reconstruction, Deep Learning, Machine Learning, Artificial Intelligence, Text-to-3D, Image-to-3D, Urban Environment, Score Distillation Sampling

Extended Abstract

The reconstruction of large-scale real outdoor environments is crucial for promoting the adoption of Extended Reality (XR) in industrial and entertainment sectors. This task often requires significant resources such as depth cameras, LiDAR sensors, drones, and others, alongside traditional data processing pipelines like Structure-from-Motion (SfM), which demand extensive computational resources, thus preventing real-time processing. Additional constraints arise from the limited accessibility to the aforementioned resources. While 3D laser scanners (e.g., LiDAR) are precise and fast, they are expensive, often bulky – especially the high-quality models – and their effectiveness is contingent on the type of environment being scanned. Depth sensors offer a more affordable and compact alternative; however, due to their limited range, they are ideal only for indoor settings. Photogrammetry, while capable of producing high-quality results at a lower cost, can be time-consuming and computationally intensive. It also suffers from limited accuracy, strong dependence on lighting conditions, and the need for numerous photos from various angles that can be not always easily accessible.

To address these limitations, we initially proposed a Spatio-Temporal Diffusion neural architecture (Federico et al., 2024), a generative architecture based on diffusion models. This solution integrates simple and cost-effective temporal information (a brief temporally ordered sequence of photographs) with spatial information (a rough approximation of the environment to be reconstructed) to rapidly reconstruct complex 3D environments, filling in missing or noisy information. The use of a neural architecture stems from the need to achieve real-time processing, while the application of generative artificial intelligence serves to compensate for the lack of information, arising from the absence of access to costly resources or situations where certain data are unattainable (e.g., an unreachable viewpoint). We also introduced a novel 3D representation termed the Most Informative Part (SDF_MIP), a modification of the well-known Signed Distance Field (SDF) that aims to symmetrically distribute positive and negative voxels — a requirement we identified as essential during network training, particularly for

outdoor environments. Our model comprises a two-stage network: the first stage fuses temporal and spatial information to generate the missing data Figure P 29, left), while the second stage converts the SDF_MIP representation back to SDF (Figure P 29, right). An optimal trade-off between reconstruction quality and execution speed was achieved using the DDIM scheduler.

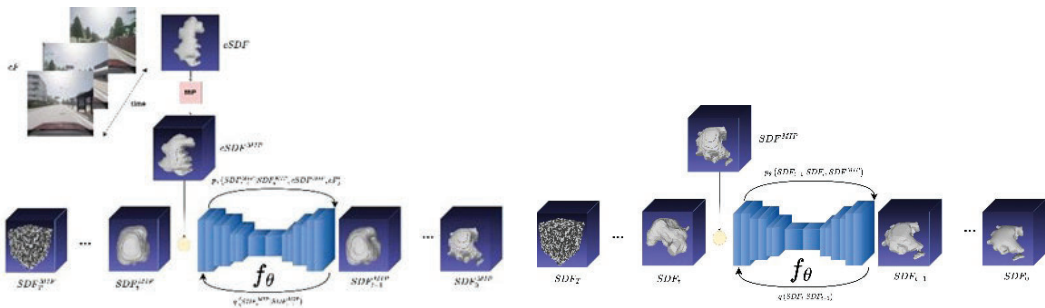


Figure P 29. Stage 1 (left) and Stage 2 (right) of our Spatio-Temporal Diffusion Neural Architecture.

Despite encouraging results, the reconstruction of complex environments required prohibitively long training times and high computational costs, failing to overcome the aforementioned limitations. Furthermore, the proposed solution does not scale well with the resolution and complexity of the target environment. We are currently developing a generative network architecture called Neural-Clipmap and an associated algorithm that alleviates training burdens, enables scalability, and minimizes computational and memory requirements. Specifically, the algorithm hallucinates and build a high quality version of a complex environment with a divide and conquer strategy by enhancing the structure of the underlying supporting octree, where each leaf is an atomic unit of computation (Figure P 30). Supported by the generative network (the diffusion one) (Ho et al., 2020), it determines whether a leaf node requires a coarsening operation (i.e., the input leaf is overly detailed and should be removed, refinement (i.e., the leaf requires further detailing), or no operation at all. The algorithm operates in two iterative phases (Figure P 30). In the first phase, for each leaf of the coarse octree, the generative network uses contextual information to modify it: a series of frames of color images of an hypothetical actor driving around a path, and the spatial neighbors of the leaf. The latter are taken at multiple levels. In particular, given a leaf, its spatial neighbors correspond to the nodes around it and, going up a level, those around the leaf's parent and so on up to the root. These two phases dynamically allow the octree levels to be pruned or increased. The generative network leverages the contextual information from our initial work (Federico et al., 2024), but it now employs a video vision encoder Arnab et al., 2021) to create a compact representation of the RGB frames. In this phase, we encountered challenges with training when using SDF or SDF_MIP, as the network struggled to appropriately correlate the contextual information. Consequently, we adopted the Triplane representation (Chan et al, 2022), inspired by recent successes in neural representations, yielding more promising preliminary results.

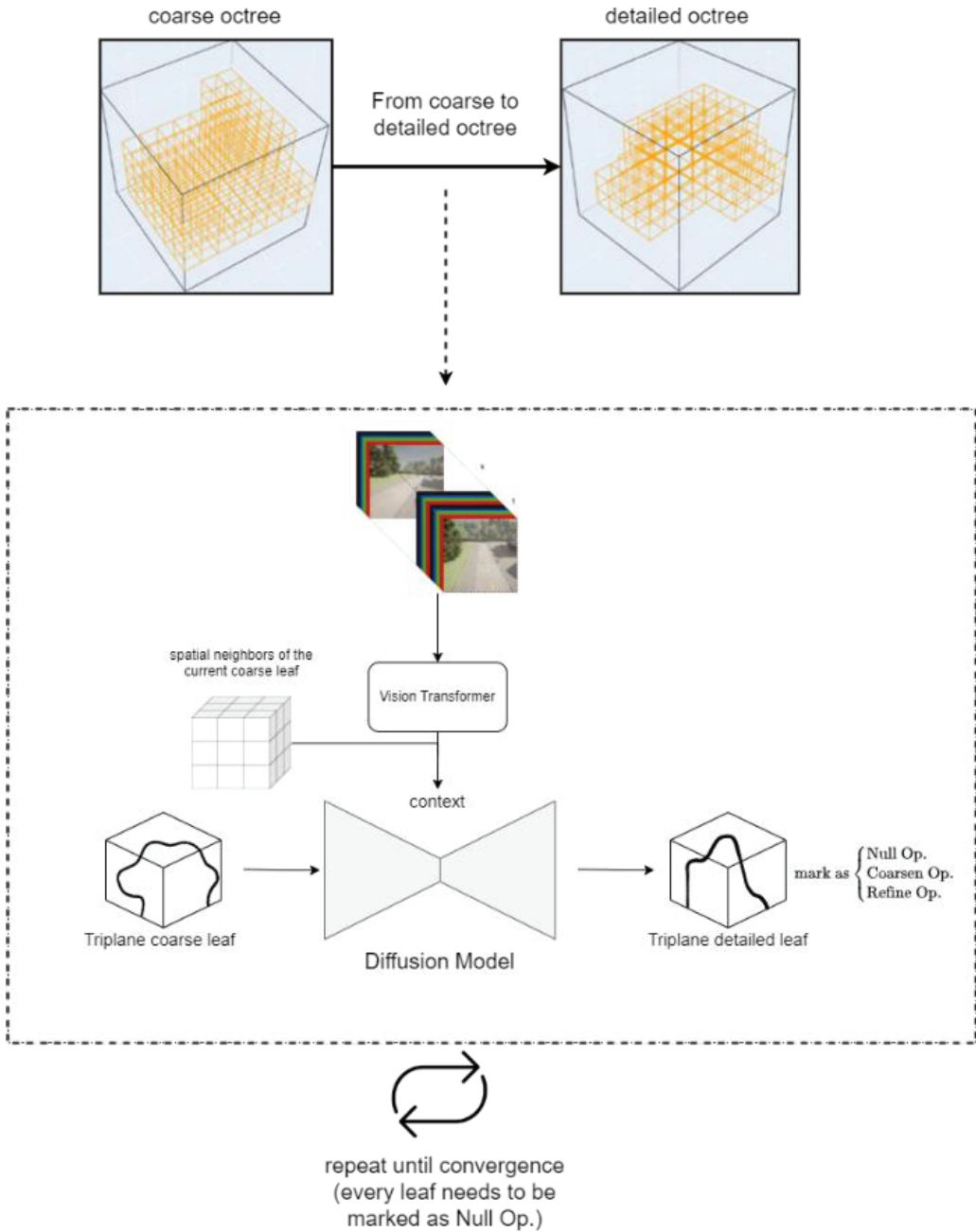
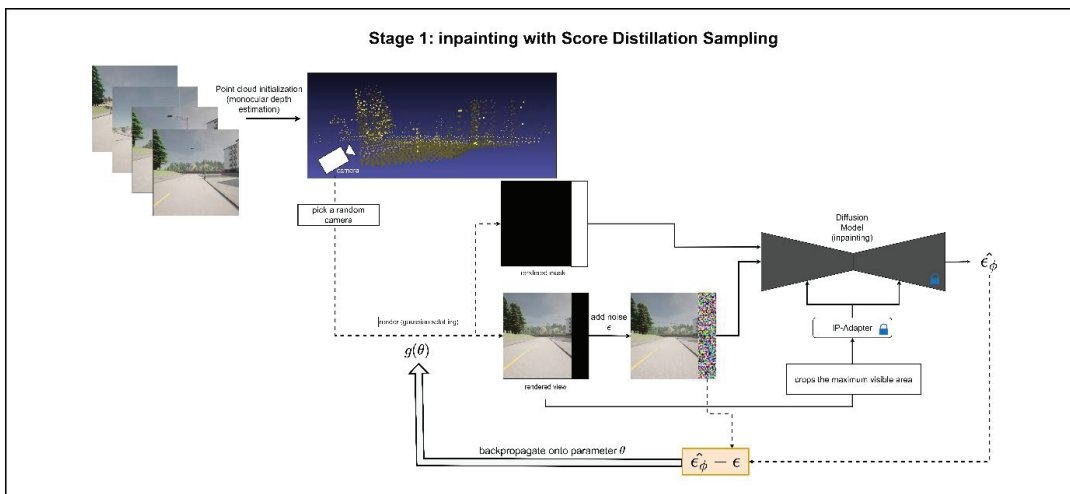


Figure P 30. The Neural-Clipmap algorithm progressively converts a coarse octree into a detailed octree.

However, generating the Triplane representation for every leaf of the octree and for each model in our dataset results in a frustrating delay. Additionally, increasing the model resolution exponentially raises the number of leaves, thereby contradicting our goal of easy scalability with resolution. Concurrently, we are exploring the use of 2D diffusion models (Ho et al., 2020) as priors for

reconstructing complex 3D environments. In this new attempt we minimize the input data by using only sequences of RGB images. Through a monocular depth estimator (Yang et al., 2024), we generate an initial point cloud of the environment, serving as the initialization for a 3D Gaussian Splatting representation (Kerbl et al., 2023). However, these points represent partial information, as the RGB image sequence covers only a portion of the environment. Thus, we employ 2D diffusion models as priors, utilizing a recent technique known as Score Distillation Sampling (SDS) (Poole et al., 2022) to reconstruct the missing information by moving the camera to points of interest. Unlike other works based on this technique (Lin et al., 2023; Tang et al., 2024; Liu et al., 2023) our challenge lies in starting from images rather than text and ensuring the consistency of the generated missing information with the existing data (termed "anchors"). SDS has been employed to generate simple 3D models and, despite this, suffers from various issues such as the Janus problem (Armandpour et al., 2023) difficulty in determining the appropriate guidance value, and overly saturated or blurred colors. To compensate for the lack of text guidance for the SDS, we used an image prompt adapter Ye et al., 2023). To tackle the issues of overly saturated or blurred colors, we devised a two-phase approach. The first phase uses the SDS (with an inpainting diffusion model) for initializing the missing parts, which may exhibit the aforementioned problems. The subsequent phase involves inpainting over the areas initialized by the SDS. For inpainting to work effectively, the missing parts require some initialization coherent with the real information available. Classical approaches include initializing the missing region with the average color of the real data, Perlin Noise, or, though slower, using an algorithm known as Patch Match (Connelly et al., 2009). We propose using SDS (Figure P 31, top) as a new method for initializing the missing areas and employing inpainting in the next phase (Figure P 31, below) to mitigate issues with saturated and blurred colors. Stage 1 is crucial not only for initializing the missing regions but also for performing multi-view inpainting in a manner consistent with other views, which would be unfeasible if starting directly from Stage 2. Furthermore, we could introduce a third Refine phase, where we add some noise to the rendered views and subsequently employ a standard diffusion model to denoise them, thereby eliminating any residual noise that may persist after Stage 2. For better quality results, the diffusion model for the refinement phase will likely be fine-tuned using the anchor images with the prior preservation technique (Ruiz et al., 2023)



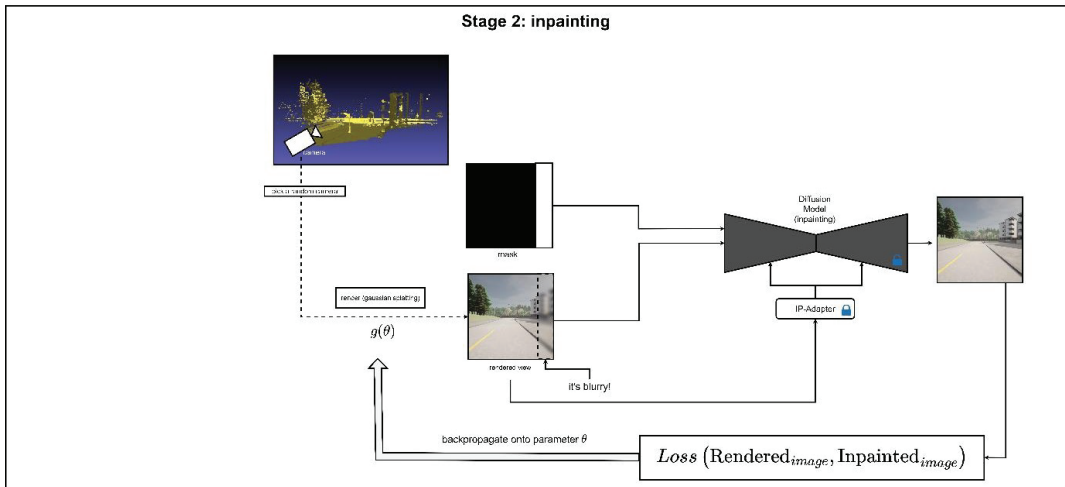


Figure P 31. Starting from a sequence of images (the anchors), a monocular depth estimator is used to initialize the point cloud and then gaussian splatting is used to render the views and optimize its parameters via Score Distillation Sampling (Stage 1, above). The inpainting model is then used to realistically fill the parts filled but blurred by Stage 1 (Stage 2, below)

References

- Armandpour, M., Sadeghian, A., Zheng, H., Sadeghian, A., & Zhou, M. (2023). Re-imagine the negative prompt algorithm: Transform 2d diffusion into 3d, alleviate janus problem and beyond. arXiv preprint arXiv:2304.04968. <https://arxiv.org/abs/2304.04968>
- Arnab, A., Deghani, M., Heigold, G., Sun, C., Lučić, M., & Schmid, C. (2021). Vivit: A video vision transformer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6836-6846).. <https://arxiv.org/abs/2103.15691>
- Barnes, C., Shechtman, E., Finkelstein, A., & Goldman, D. B. (2009). PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Trans. Graph., 28(3), 24. 1–11. <https://doi.org/10.1145/1576246.1531330>
- Chan, E. R., Lin, C. Z., Chan, M. A., Nagano, K., Pan, B., De Mello, S., ... & Wetzstein, G. (2022). Efficient geometry-aware 3d generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16123-16133). <https://arxiv.org/abs/2112.07945>
- Federico, G., Carrara, F., Amato, G., & Di Benedetto, M. (2024, June). Spatio-Temporal 3D Reconstruction from Frame Sequences and Feature Points. In Proceedings of the 2024 ACM International Conference on Interactive Media Experiences Workshops (pp. 52-64). Association for Computing Machinery, New York, NY, USA, 52–64. <https://doi.org/10.1145/3672406.3672415>
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851. <https://arxiv.org/abs/2006.11239>
- 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. <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>
- Lin, C. H., Gao, J., Tang, L., Takikawa, T., Zeng, X., Huang, X., ... & Lin, T. Y. (2023). Magic3d: High-resolution text-to-3d content creation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 300-309). <https://arxiv.org/abs/2211.10440>

- Liu, R., Wu, R., Van Hoorick, B., Tokmakov, P., Zakharov, S., & Vondrick, C. (2023). Zero-1-to-3: Zero-shot one image to 3d object. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9298-9309). <https://arxiv.org/abs/2303.11328>
- Nataniel Ruiz and Yuanzhen Li and Varun Jampani and Yael Pritch and Michael Rubinstein and Kfir Aberman. 2023. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. <https://arxiv.org/abs/2208.12242>
- Poole, B., Jain, A., Barron, J. T., & Mildenhall, B. (2022). Dreamfusion: Text-to-3d using 2d diffusion. arXiv preprint arXiv:2209.14988. <https://arxiv.org/abs/2209.14988>
- Tang, J., Ren, J., Zhou, H., Liu, Z., & Zeng, G. (2023). Dreamgaussian: Generative gaussian splatting for efficient 3d content creation. arXiv preprint arXiv:2309.16653. <https://arxiv.org/abs/2309.16653>
- Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., & Zhao, H. (2024). Depth anything: Unleashing the power of large-scale unlabeled data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10371-10381). <https://arxiv.org/abs/2401.10891>
- Ye, H., Zhang, J., Liu, S., Han, X., & Yang, W. (2023). Ip-adapter: Text compatible image prompt adapter for text-to-image diffusion models. arXiv preprint arXiv:2308.06721. <https://arxiv.org/abs/2308.06721>

WebAR for Construction Visualization

Urs Riedlinger

Fraunhofer FIT, Sankt Augustin, Germany

Corresponding author: urs.riedlinger@fit.fraunhofer.de

Keywords: Web, Augmented Reality, WebAR, WebXR, Framework, Construction

Introduction

Augmented Reality (AR) has increasingly become a powerful tool for enhancing real-world interactions, particularly in fields requiring on-site visualization. In this context, web-based AR offers a flexible and accessible approach, eliminating the need for dedicated applications and providing cross-platform capabilities. Motivated by the demands of a project requiring on-site AR visualization of a newly constructed school building, this poster explores various web-based AR frameworks to identify the most suitable options for real-world implementation. This poster presents an exploratory comparison of several AR frameworks, with a focus on providing practical insights for developers and researchers. A small prototype was developed for each framework, allowing for an initial evaluation and comparison. Based on this overview, we conducted in-depth tests of three key frameworks: Adobe Aero, AR.js, and MindAR, and decided to create an example application with MindAR.

Related Work

Web-based Augmented Reality (WebAR) has gained significant attention due to its potential to deliver immersive experiences directly in the browser, without the need for additional software. Several studies have explored different aspects of AR frameworks and their application in mobile and web environments. (Cao et al., 2023) provide an extensive survey on *Mobile Augmented Reality*, highlighting various user interfaces, frameworks, and the role of artificial intelligence in enhancing mobile AR experiences. Their work emphasizes the importance of intuitive user interfaces and the increasing intelligence behind AR systems, which also holds relevance in the context of WebAR. Their study primarily focuses on mobile AR apps, with limited attention to web-based solutions. (Júnior and Cuperschmid, 2019) focus specifically on WebAR, evaluating the *8th Wall* framework for building visualization. Their work demonstrates the practical application of WebAR in the construction industry, showcasing how browser-based AR can be effectively used for architectural visualizations, a use case we consider in this paper as well. This study underlines the real-world potential of WebAR but does not provide a comparison between multiple frameworks, as this paper does. (Nitika et al., 2021) investigate the *performance of WebAR across different web browsers*, providing a technical evaluation of browser compatibility and rendering performance. Their findings are crucial for understanding the limitations of

WebAR, particularly in terms of hardware and software dependencies, which align with this paper's exploration of compatibility issues and performance across various frameworks. Despite these contributions, no comprehensive comparison of WebAR frameworks has been made to date. As web technologies evolve rapidly (Cao et al., 2023), the results from framework comparisons like this one risk becoming quickly outdated. Additionally, because many frameworks are built on foundational web technologies like *three.js* or *A-Frame*, there is a critical need for consistent maintenance to ensure ongoing performance and compatibility with evolving web standards. This paper aims to fill the gap by providing a detailed comparison of key WebAR frameworks, contributing to the body of knowledge on web-based AR solutions.

Framework comparison

The exploration of web-based AR frameworks reveals an interconnected ecosystem, where many solutions are built on technologies like *three.js* and *A-Frame*. This modularity offers flexibility but complicates direct comparisons, as shown in Figure P 32, which illustrates how AR frameworks layer over foundational engines. Our comparison began with a review of current WebAR solutions, assessing them based on infrastructure, tracking, and development features (summarized in Table P. 4). From this review, we selected three frameworks for in-depth testing: Adobe Aero, AR.js, and MindAR. Adobe Aero offers a no-code, visually rich experience but is limited by its proprietary nature and higher system requirements. AR.js, an open-source solution built on *three.js* and *A-Frame*, is lightweight and cost-effective but lacks some advanced features. MindAR, focused on marker-based AR, is ideal for high-performance tracking but is less versatile for complex AR applications. We evaluated these frameworks based on development ease, customization, tracking accuracy, and compatibility. While detailed tracking comparisons could be insightful, many frameworks rely on similar underlying engines, so tracking differences were not a primary factor. Though cost wasn't part of the comparison table due to varying licensing models, it is a key consideration. Open-source tools like AR.js and MindAR offer budget flexibility, while proprietary tools like Adobe Aero may justify their higher costs with additional features and support. Ultimately, we chose MindAR for its balanced performance, ease of use, and cost-effectiveness, aligning with the project's on-site AR visualization needs.

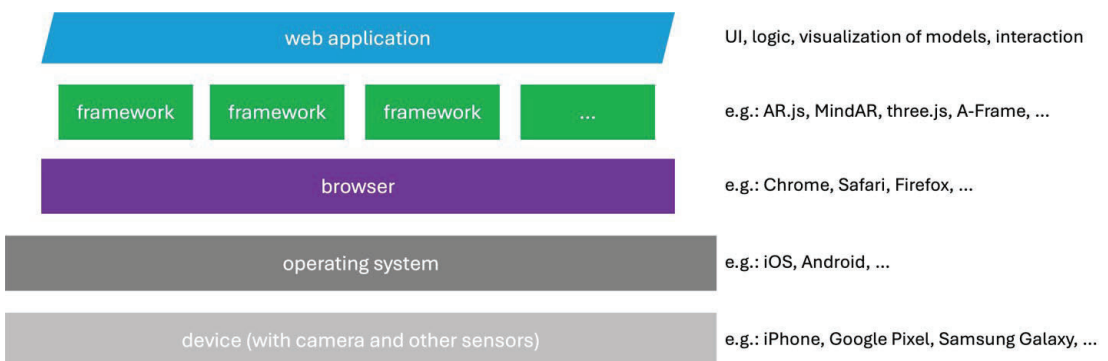


Figure P 32. Technology stack for WebAR applications.

Table P. 4. Comparison of different WebAR frameworks

Example application: Digital Construction Fence

A practical example of WebAR is the digital construction fence, which provides immersive visualizations for stakeholders in construction and urban development using MindAR and A-Frame (cf. Figure P 33). This application offers three distinct types of visualizations. First, a small AR experience can be triggered from a business card, allowing users to view a 3D model of the future construction, making it ideal for quick presentations and networking. Second, on-site visualization allows users to scan a designated area with their mobile devices to see a real-time AR overlay of the planned construction in its actual environment. Finally, a purely virtual version enables remote stakeholders to explore the construction site in detail via a web browser, offering flexibility for those who cannot visit the site in person.

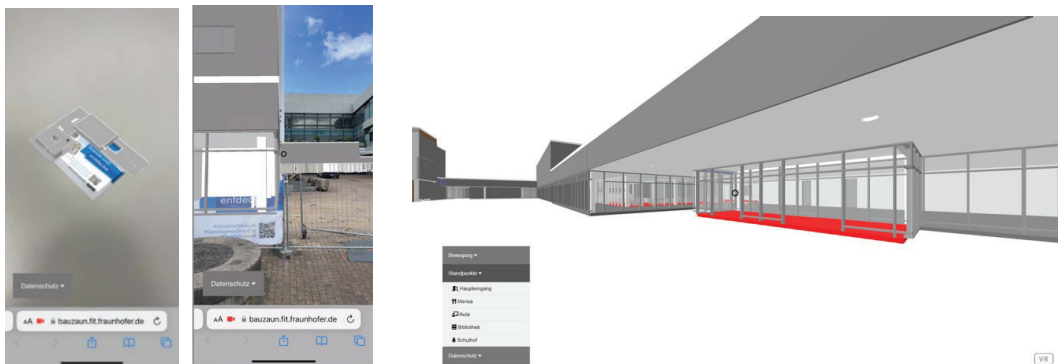


Figure P 33. Example application: Digital Construction Fence (f.l.t.r.: on business card, on-site, virtual)

Discussion

WebAR offers several key advantages that make it an appealing choice for developers and users alike. One of the most significant benefits is device compatibility. Unlike native applications, WebAR runs directly in a browser, which makes it accessible across a wide range of devices without the need for app downloads or installations. This ensures broader participation, as users can access AR experiences on smartphones, tablets, and even emerging devices like the Apple Vision Pro. As WebAR evolves, compatibility with such cutting-edge devices promises to deliver even more immersive and sophisticated AR experiences across platforms. Another advantage of WebAR is the ease of participation. Users can engage with AR content through simple links or QR codes, eliminating many of the traditional barriers to entry associated with app-based AR. This lower friction makes WebAR

particularly useful for short-term projects, marketing campaigns, or educational initiatives, where quick access to content is critical. However, WebAR is not without its limitations. One notable challenge is access rights management, particularly when it comes to handling permissions for camera access. WebAR applications require permission to use the camera, and inconsistent permission handling across browsers and operating systems can lead to user confusion or reluctance to grant necessary access, potentially limiting participation. Furthermore, hardware requirements can pose limitations on WebAR's reach. While the technology works across many devices, it struggles to offer the same level of experience on older or low performance phones. This hardware may lack the processing power or browser capabilities needed to deliver smooth and high-quality AR interactions, which creates a gap in user experience, with more advanced and newer devices capable of richer, more immersive AR, while older or weaker ones may experience delays or diminished functionality. In conclusion, while WebAR presents an accessible and widely compatible option for delivering AR experiences, considerations around device capabilities and access permissions must be managed carefully to ensure a seamless user experience across different devices and platforms.

Conclusion and Future Work

This exploration of WebAR frameworks highlights their versatility and accessibility for delivering immersive experiences across various devices. Our comparison of frameworks like Adobe Aero, AR.js, and MindAR revealed key trade-offs in ease of use, technical performance, and compatibility, leading to the selection of the most suitable solution for our project. WebAR's broad device compatibility and ease of access make it ideal for wide-reaching projects, but issues like access rights management and hardware constraints must be addressed to ensure a consistent user experience. Future research could focus on optimizing performance for older devices, refining permission management, and improving tracking technologies across varied environments. Additionally, as advanced devices like the Apple Vision Pro emerge, exploring WebAR's potential for even more immersive experiences is crucial. By tackling these areas, WebAR frameworks can evolve, remaining powerful tools for developers and users in the fast-changing AR landscape.

References

- Cao, J., Lam, K.-Y., Lee, L.-H., Liu, X., Hui, P., & Su, X. (2023). Mobile Augmented Reality: User Interfaces, Frameworks, and Intelligence. *ACM Comput. Surv.* 55, 189:1-189:36. <https://doi.org/10.1145/3557999>
- Júnior, J.N.S., & Cuperschmid, A.R.M. (2019). WebAR para visualização de modelo da edificação: avaliação da ferramenta 8th Wall Web. *Simpósio Brasileiro De Tecnologia Da Informação E Comunicação Na Construção* 2, 1–7. <https://doi.org/10.46421/sbtic.v2i00.136>
- Nitika, N., Sharma, T.K., Rajvanshi, S., & Kishore, K. (2021). A study of Augmented Reality performance in web browsers (WebAR), in: 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST). Presented at the 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST), pp. 281–286. <https://doi.org/10.1109/ICCMST54943.2021.00065>

Augmenting controls for tracked vehicles with force and tactile feedback: a work in progress

Thomas Howard, Benjamin Geslot, and Jérôme Perret

Haption S.A.S, Soulge-sur-Ouette, France

Corresponding author: thomas.howard@haption.com

Keywords: Haptics, Force Feedback, Vibrotactile Feedback, Off-Highway Vehicles

Motivation

The THEIA^{XR} project (TheiaXR, 2024) is concerned with integrating novel human-machine interaction (HMI) technologies in off-highway vehicles, in order to improve operation safety, efficiency and operators' quality of work. In the off-highway vehicle domain, many vehicles move using tracks rather than wheels, as this allows better stability and traction on soft and loose ground such as e.g., sand, mud or snow (Wong, 1997). Such tracked vehicles are operated using a pair of levers which allow the operator to independently set the velocity setpoint for each track, allowing agile maneuvering on complex terrain (see Figure P 34). Existing track controls exclusively act as input devices for the operator, and provide no form of assistance to the user or feedback on the current status of the operated machine.

The provision of haptic feedback has been extensively studied in wheeled road vehicles (Gaffary and Lécuyer, 2018) and their simulators (Petermeijer et al., 2015). Force feedback through the steering wheel has been shown to be effective for e.g., steering assistance and vehicle targeting (Dennerlein et al., 2000, lane-keeping (Prinoth, 2024), and the safe maneuvering of vehicles with attached trailers (Morales et al., 2013). Tactile feedback on the other hand has been shown to effectively assist vehicle navigation (Hwang and Ryu, 2010), obstacle avoidance [1] and speed control (Birrell et al., 2013). Force feedback has also been explored for teleoperated tracked robots and vehicles [5], but has never been applied within a tracked vehicle cabin in a drive-by-wire context.

In this work, we therefore seek to investigate potential benefits that could be obtained by turning track controls into haptic feedback devices capable of delivering dual force and tactile feedback to an operator located directly inside the vehicle cabin.

Acknowledgements

This work has been funded by the European Union within the Horizon Europe Framework under Grant Agreement No. 101092861 (Research and Innovation Action "THEIA-XR"). Views and opinions expressed are however those of the authors 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.

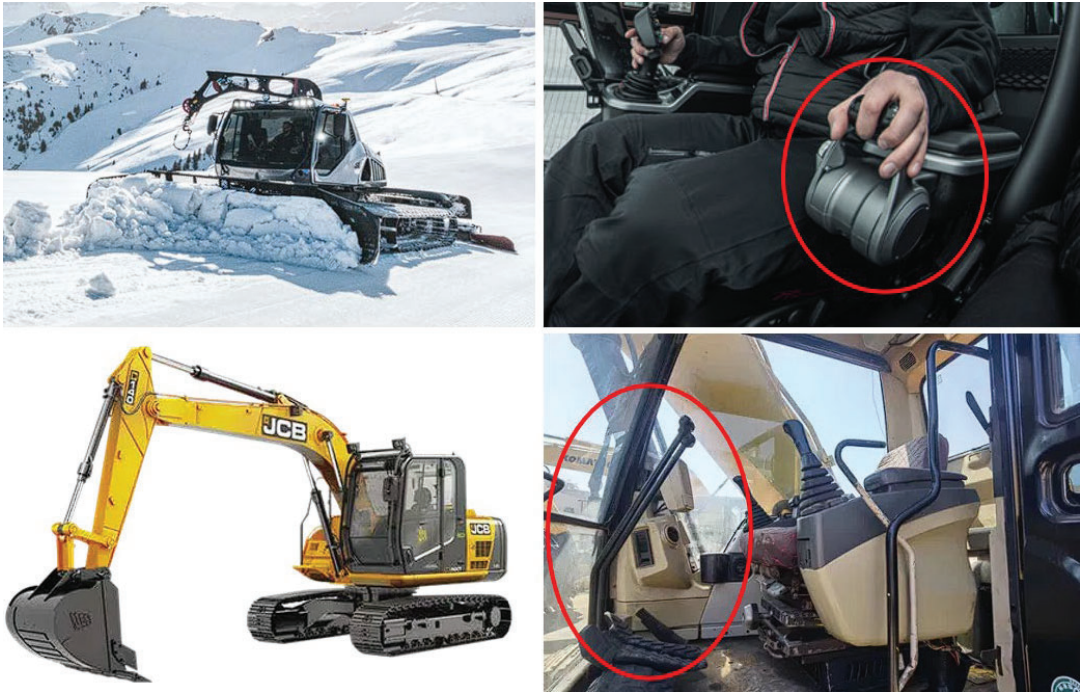


Figure P 34. Examples of tracked off-highway vehicles studied within the THEIAXR project (left) with examples of associated track control levers (circled in red, right)

Device prototype

We designed a prototype (see Figure P 35) substitute for a pair of track control levers of a Prinoth Leitwolf snowgroomer [9].

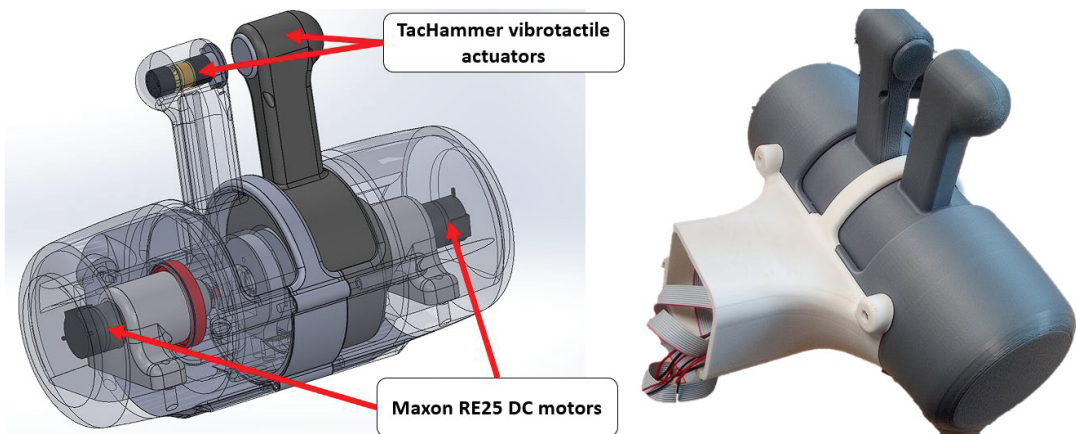


Figure P 35. (Left) CAD design of the first prototype, showing the vibrotactile actuators integrated into the handles and the DC motors providing independent torque feedback to each lever. (Right) Assembled prototype.

The prototype features a pair of symmetrically arranged levers rotating a shaft whose angular position is measured using a Maxon MR 1000 point relative encoder, while torque feedback can be

independently provided on each shaft through an attached Maxon RE25 DC motor coupled with a HFUC-2A Harmonic Drive strain wave gear reductor with a ratio of 50. We integrated Titan Haptics TacHammer Carlton linear magnetic ram actuators capable of delivering intense high-fidelity vibrotactile feedback cues within the tip of each lever. The vibrotactile actuators are driven using custom-built electronics based on an Arduino Micro and a pair of Texas Instruments DRV2605 haptic drivers, connected to the main embedded controller via SPI. Force feedback control is achieved using a proprietary embedded controller designed by Haption, communicating with a host software on a PC via EtherCAT. This host software is e.g. capable of interfacing the haptic track controls with a simulated snowgroomer and environment in Unity3D.

Envisioned haptic assistance functions and future work

In upcoming evaluation work, we will integrate the designed prototype into a realistic snowgrooming simulator environment in order to test the acceptability and effectiveness of the haptic assistance functions detailed below.

We will investigate the impact of tactile feedback on improving lane-keeping during snowgrooming, e.g., to ensure consistent overlap and parallelism between passes on groomed ski slopes (see Figure P 36 left). Velocity has a significant impact on resulting snow quality, we will also investigate the use of tactile warnings in case of excessive velocity. Also, significant discrepancies between the commanded track velocities and the vehicle ground velocity measured e.g., using GPS, may be an indicator of the vehicle treading in place, digging itself into a pit and damaging the groomed snow slope. We will therefore investigate tactile warnings relating to excessive or insufficient vehicle velocity (see Figure P 36 center). Finally, we plan to assess whether tactile warnings of impending collisions can assist the operator in quickly reacting to avoid them (see Figure P 36 right).

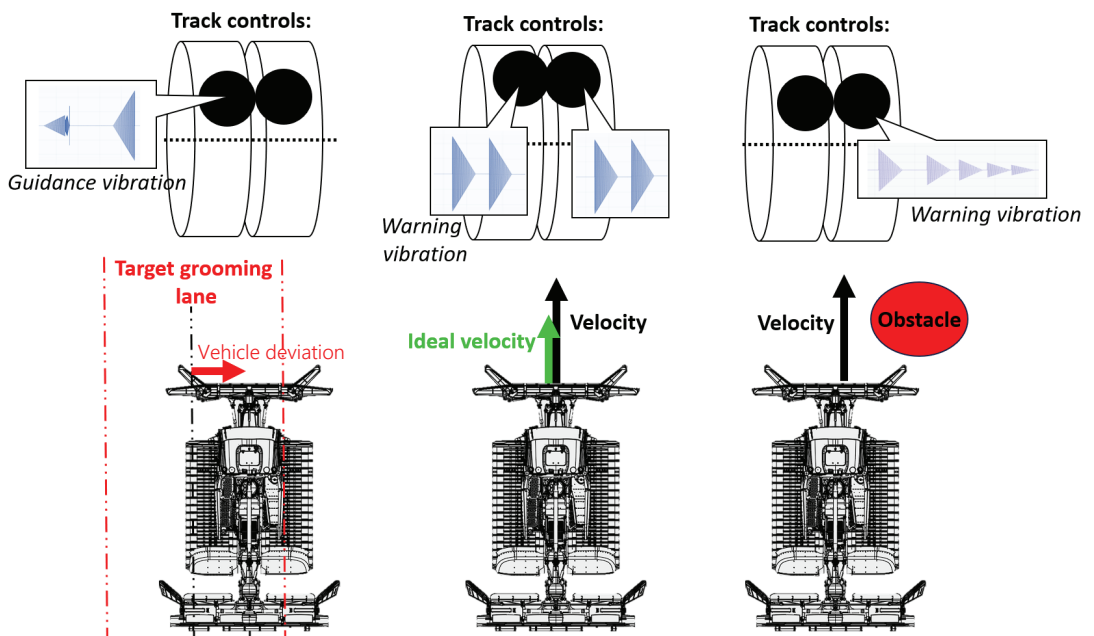


Figure P 36. Concepts for tactile feedback assistance for navigation (left), speed control (center) and obstacle avoidance (right)

Regarding the use of force-feedback capabilities, we will investigate the idea of guiding the user towards a lower velocity as a complementary form of feedback to the tactile velocity warnings mentioned above. Furthermore, we are in the process of implementing a navigation assistance controller, which gently guides the operator towards the ideal track control lever pose required to reach a defined vehicle orientation and velocity (see Figure P 37).

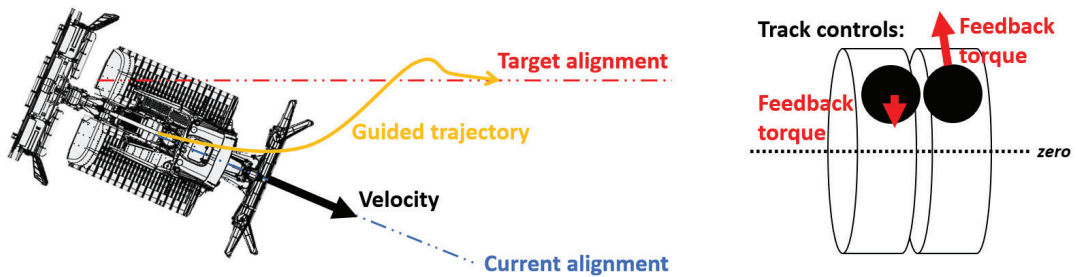


Figure P 37. Concept for torque feedback navigation assistance

References

- Ahtamad, M., Gray, R., Ho, C., Reed, N., & Spence, C. (2015). Informative collision warnings: effect of modality and driver age. in Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (Salt Lake City, Utah), 330–336. <https://doi.org/10.17077/drivingassessment.1590>
- Birrell, S. A., Young, M. S., & Weldon, A. M. (2013). Vibrotactile pedals: provision of haptic feedback to support economical driving. *Ergonomics* 56, 282–292. <https://doi.org/10.1080/00140139.2012.760750>
- Dennerlein, J. T., Martin, D. B., & Hasser, C. (2000). Force-feedback improves performance for steering and combined steering-targeting tasks. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 423–429).
- Gaffary, Y., & Lécuyer, A. (2018). The use of haptic and tactile information in the car to improve driving safety: A review of current technologies. *Frontiers in ICT*, 5, 5. <https://doi.org/10.3389/fict.2018.00005>
- Horan, B., Creighton, D., Nahavandi, S., & Jamshidi, M. (2007). Bilateral haptic teleoperation of an articulated track mobile robot. In 2007 IEEE International Conference on System of Systems Engineering (pp. 1–8).
- Hwang, S., & Ryu, J. H. (2010). The Haptic steering Wheel: Vibro-tactile based navigation for the driving environment. In 2010 8th IEEE PERCOM Workshops (pp. 660–665).
- Morales, J., Mandow, A., Martínez, J. L., Reina, A. J., & García-Cerezo, A. (2013). Driver assistance system for passive multi-trailer vehicles with haptic steering limitations on the leading unit. *Sensors*, 13(4), 4485–4498. <https://doi.org/10.3390/s130404485>
- Petermeijer, S. M., Abbink, D. A., Mulder, M., & De Winter, J. C. (2015). The effect of haptic support systems on driver performance: A literature survey. *IEEE transactions on haptics*, 8(4), 467–479.
- Prinoth A.G. (2024), <https://www.prinoth-snowgroomers.com/en/products/leitwolf>
- Switkes, J. P. (2005). Handwheel Force Feedback With Lanekeeping Assistance: Combined Dynamics, Stability and Bounding, a dissertation of doctor of philosophy.
- THEIAXR (2024). <https://www.theia-xr.eu/>
- Wong, J. Y. (1997). Dynamics of tracked vehicles. *Vehicle system dynamics*, 28(2-3), 197–219. <https://doi.org/10.1080/00423119708969354>

Ethical impact of Artificial Intelligence in Extended Reality

Irma Pöder

Trilateral Research, London or Waterford Ireland

Corresponding author: irma.poder@trilateralresearch.com

Keywords: Ethics, Impact Assessment, Extended Reality, AI In XR

Introduction

As part of regulating artificial intelligence (AI) in Europe great efforts have been put towards defining and understanding the ethical principles for AI. As extended reality (XR) is facilitated by AI there is significant overlap in the applicability of the principles for also assessing the ethical impact of XR. Existing research has made great efforts into understanding the accessibility and inclusivity requirements for users who are e.g. visually impaired or hard of hearing (Fox & Thornton, 2022). The field of XR is still relatively new and therefore a limited number of standards are available for enabling interoperability. The sustainable development of XR applications is dependent on accounting for social, economic and environmental impacts e.g. by meeting requirements for cost-effective delivery of the underlying system and updates as well as limiting the energy consumption of the tools (Carvalho & da Silva, 2021). To understand the role of AI in impacting the sustainable development of XR applications an ethical impact assessment was carried out.

This paper presents the ethical impact assessment of XR applications and the tools that enable them using a collaborative approach between developers, social scientists and external stakeholders / potential end-users. The primary findings are around trust, health and safety, technical robustness, human agency and oversight, autonomy and beneficence. The risks are presented alongside suggested actionable mitigation measures for considering the potential ethical impacts as part of the technology design.

Approach for ethical impact assessment of AI in XR

To carry out the ethical impact assessment a framework of various ethical, legal and societal guidelines, standards, and regulations was defined. These provide the basis for identifying and assessing

the potential impacts of AI in XR. The frameworks identified as most relevant to the assessment are detailed in the below table¹⁸.

Table P. 5. Frameworks for ethical impact assessment of AI in XR applications

Framework	Applicability
Ethics Guidelines and Assessment List for Trustworthy AI (High-Level Expert Group on Artificial Intelligence (HLEG), 2019)	For the sustainable development of XR and adoption of the technology in a wide variety of use cases the solutions need to be trustworthy. The AI High Level Expert Group (AI HLEG) have identified key requirements for AI systems to meet in order to be considered trustworthy.
Data ethics (Tranberg et al., 2018)	As using AI for XR processes large amounts of data, considering data ethics is necessary for the sustainable development of XR, by adhering to best practice standards such as data minimisation.
Recommendation of the Council on AI (OECD, 2024)	The OECD principles for responsible stewardship of trustworthy AI are a series of values-based practical recommendations.
Privacy-by-design (<i>Social Impact Assessment</i>, n.d.)(Cavoukian, 2009)	To propose actionable mitigation measures to the ethical risks identified the foundational principles of privacy by design were followed so that ethics may be embedded into the applications.
EU Charter of Fundamental Human Rights (CFREU)	Ethical values are codified and given expression in the form of human rights. Though multiple instruments of human rights exist also at the national level, the CFREU has a strong normative influence on national human rights laws.
General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679)	On top of potentially processing personal data for creating XR applications, GDPR will also be applicable for any third party devices that collect or process the data from the user.
EU AI Act (European Parliament, 2023)	A first of its kind, the EU AI Act will regulate AI systems put on the market on the EU using a risk-based approach. Similarly to the GDPR the regulation is applicable not just in the EU but extraterritorially as well and is expected to impact future AI regulation.
UN Sustainable Development Goals (SDGs) (United Nations, 2015)	The AI HLEG has suggested that AI can help achieve the UN SDGs and was hence chosen to help assess the social impact and sustainability of AI in XR and how the technology may support positive societal impacts.

Following the definition of the framework to be used for assessing the ethical impacts in AI used for creating XR applications, a pre-identification of potential risks was carried out by the ethics advisors. These risks were taken to the developers and discussed in dialogue sessions to be sure there was a shared understanding of the process and the implications between the technical experts, ethicists and the social scientists. The outcomes of these dialogue sessions were then taken to a validation workshop with stakeholders to be sure any assumptions made about potential concerns were valid and that any risks that were missed by the development team could be considered as part of the co-design process.

¹⁸ A limitation of this study is that the frameworks identified for this particular assessment are Western-centric as this is where the regulation on AI is currently most advanced. However, future iterations could be adapted to be inclusive of non-Western ethics and regulatory frameworks for more globally applicable findings.

Ethical risks and potential impacts

The ethics of AI in XR can be highly context specific: depending on the application and use case area the approach for addressing trustworthy AI principles, such as fairness and explainability, may need to vary. The impact of the ethical implications of AI in XR is of course also affected by the user and the level of prior exposure or experience they have had in using XR devices, especially headsets. Below are some of the risks associated with the use of AI for creating XR applications.

Equality and non-discrimination. Bias or discrimination can creep in to the design of an XR application through overlooking diversity and inclusion metrics when collating requirements, insufficient stakeholder engagement in the design process resulting in the development of a technological solution that is not useful for its target audience, or omitting just a certain type of stakeholder resulting in the reinforcement of pre-existing biases from the developers.

Technical robustness and safety. In instances where XR applications are connected to the internet a lack of technical robustness will also have inherent cybersecurity risk. These threats may be directly from targeted attacks by malicious actors or through poor data management and cybersecurity protocols leading to unintentional leaks. A lack of technical robustness will lead to an unreliable application which can cause safety concerns, e.g. in a high pressure working environment.

Trust and human agency. As technology gets smarter and virtual realities become more realistic humans may become overreliant or too trusting of the system, believing it is capable of more than what it was designed to do. Or on the opposite end of the spectrum a system can be too opaque, miscommunicated and its functions or capabilities hidden, creating distrust in users and the public. The aim of new technology in the workplace should not be to replace established practices skills and knowledge but to enhance current roles through providing insights and creating new responsibilities by supporting the completion of more tedious roles.

Privacy. End-users' privacy could be at risk through not establishing a legal basis for the sharing and processing of personal data, non-consensual collection of personal data, e.g. through third-party devices, such as Hololens requesting biometric data, excessive processing collection and further (re)processing of data.

Health and safety. A key risk to consider is that many XR applications will be headset-based and it is well documented that some users may experience physical side effects (Saredakis et al., 2020). The side effects are exacerbated as it is not always clear which preconditions in a user may cause the side effects. Though it is known that age or pre-existing eye conditions may be contributing factors users are not always aware of such ailments before using XR devices. Additionally, using headsets for longer period of time, for example in a work environment, can lead to disorientation when adjusting back to the real world.

Beneficence and do-no-harm. Sustainability and health and safety may be impacted through over-collecting irrelevant data, leading to a requirement of greater amounts of processing power and storage spaces, and greater threats to privacy. A lack of consideration for sustainability can also have an adverse environmental impact, for example using excessive amounts of computing power could lead to preventable environmental harm. Working alongside machines may reduce employees value in their work, impacting their autonomy and ultimately leading to job loss. There is also a risk of low uptake or engagement which could render the social value of the XR applications limited.

In summary, with XR having found most uptake in the gaming sector at the moment it's impact on society has been limited, however it is expected the XR industry will grow 33% by 2032 (Fortune Business

Insights, 2024) as more ideas for exploiting the technology become apparent. As such more sector and use case specific ethical risks will occur. This account of ethical concerns is by no means exhaustive as highly context specific principles of responsible AI such as bias, explainability and privacy will need to be considered in more specific detail to truly implement an ethics-by-design approach.

References

- Carvalho, R., & da Silva, A. R. (2021). Sustainability Requirements of Digital Twin-Based Systems: A Meta Systematic Literature Review. *Applied Sciences*, 11(12), Article 12. <https://doi.org/10.3390/app11125519>
- Cavoukian, A. (2009). Privacy by Design The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada, 5(12).
- European Parliament. (2023, June 8). EU AI Act: First regulation on artificial intelligence. Topics | European Parliament. <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- Fortune Business Insights. (2024). Extended Reality (XR) Market Size, Share & Industry Analysis, By Type (Virtual Reality, Augmented Reality, and Mixed Reality), By Industry (Healthcare, Education, Retail & E-commerce, Gaming, Automotive, Media & Entertainment, and Others), and Regional Forecast, 2024-2032. <https://www.fortunebusinessinsights.com/extended-reality-market-106637>
- Fox, D., & Thornton, I. G. (2022/). The IEEE Global Initiative on Ethics of Extended Reality (XR) Report–Extended Reality (XR) Ethics and Diversity, Inclusion, and Accessibility. The IEEE Global Initiative on Ethics of Extended Reality (XR) Report–Extended Reality (XR) Ethics and Diversity, Inclusion, and Accessibility, 1–25. The IEEE Global Initiative on Ethics of Extended Reality (XR) Report–Extended Reality (XR) Ethics and Diversity, Inclusion, and Accessibility. <https://ieeexplore.ieee.org/document/9727122/?arnumber=9727122>
- High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- OECD. (2024). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>
- Saredakis, D., Szpak, A., Birkhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience*, 14. <https://doi.org/10.3389/fnhum.2020.00096>
- Tranberg, P., Hasselbalch, G., Olsen, K., & Byrne, C. S. (2018). DATAETHICS – Principles and Guidelines for Companies, Authorities & Organisations. DataEthics.EU.
- United Nations. (2015). Transforming our World: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. <https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981>

A Virtual Reality Evaluation Testbed for External Human-Machine Interfaces in Autonomous Vehicles

Yiran Zhang, Guoqin Zang, and Cédric Bach

Human Design Group, Versailles, France

Corresponding authors: yiran.zhang@humandesign.group, guoqin.zang@humandesign.group,
cedric.bach@humandesign.group

Keywords: Human Factors, Virtual Reality, Autonomous Vehicles

Abstract

Autonomous vehicles (AVs) face greater challenges navigating urban areas compared to highways, particularly in communicating with pedestrians to ensure road safety and traffic flow. Previous research has explored various communication methods, including implicit and explicit ones, with virtual reality (VR) emerging as a cost-effective evaluation tool. However, many of studies are limited by single testing scenarios, overlooking some important factors that may influence pedestrian decision-making. To address these gaps, we identified four key factors from a prior survey on pedestrian behavior with both classical and autonomous vehicles. We then propose a VR-based evaluation testbed that allows experimenters to configure diverse urban scenarios. This ongoing work aims to provide a comprehensive assessment of autonomous vehicle communication systems.

Introduction and Rational

With advancements in technology, more and more autonomous functions are being integrated into vehicles. The transition to autonomous systems poses challenges for pedestrians in interpreting AVs' intentions. The absence of human drivers impairs traditional communication methods, such as eye contact, facial expressions, and hand gestures, in AV interactions (Sucha et al., 2017; Schmidt et al 2009). To address this issue, previous research has investigated both implicit and explicit communication methods for external human-machine Interface (eHMI) in AVs. While implicit cues, like vehicle behaviors allow pedestrians to deduce the vehicle's intentions (Risto et. al., 2017; Zimmermann et al., 2017), explicit signals directly convey AVs' intentions and provide guidelines to pedestrians through visual (De et al., 2019; Weber et al., 2019; Rettenmaier et al., 2019; Hensch et al., 2019; Palmeiro et al., 2018) and audio cues (Matthews et al., 2017; Florentine et al., 2016). To assess pedestrians' comprehension of these designs and evaluate their effectiveness, various methods, such as images (Fridman et al., 2017; Fuest et al., 2020), videos (De et al., 2019; Song et al., 2018; Eisma et al., 2019), Wizard of Oz (WoZ) (Hensch et al., 2019; Fuest et al., 2018; Joisten et al., 2020; Currano et al., 2018; Rothenbucher et al., 2016) and virtual reality (Schmidt et al., 2019; Dietrich et al., 2020; De et al., 2019; Chang et al., 2017; Weber et al., 2019;

Bernhard et al., 2008) are employed. Each method has its own strengths and limitations, making them suitable for different stages of evaluation of AVs communication.

Image- and video-based methods are effective for early-stage evaluations, enabling quick collection of user feedback to guide further studies. For example, participants can rate images of approaching AVs based on their understanding (Fridman et al., 2019) or willingness to cross (Fuest et al., 2020). Video studies can assess participants' crossing behavior in response to various vehicle driving behaviors and automation states at unmarked roads (Dey et al., 2019). However, both images and videos fall short in providing an immersive experience and interactive evaluation. Since AV functions are still under development, the WoZ technique, where a hidden driver mimics an AV (Fraser et al., 1991), is often employed. A study using WoZ technique has shown that a visible driver doesn't always enhance pedestrians' perceived safety, and pedestrians tend to prioritize vehicle behavior over eHMI features in safety assessments (Hensch et al., 2019). While WoZ technique offers realistic evaluations, it usually poses safety risks and is costly due to the complexity of the setup and data collection (Fuest et al., 2020).

With advancements in technology, virtual reality has emerged as a valuable tool for providing a cost-effective and immersive experience in evaluating pedestrian-vehicle interactions. For example, virtual reality is used to assess pedestrians' reactions and trust by simulating different trajectories, driving behaviors and strategies (Fuest et al., 2020; Schmidt et al., 2019; Dietrich et al., 2020). Additionally, it allows for the simulation of more interactive eHMI concepts, such as an "eye" that appears to look towards pedestrians (Chang et al., 2017). Its straightforward setup also facilitates the replication across different cultural contexts.

Understanding pedestrian behavior is complex and influenced by various factors, such as road conditions, traffic characteristics, and social dynamics, all of which significantly impact decision-making when interacting with AVs. However, certain important factors, such as group size and road design, have been underexplored in existing research, limiting the potential to fully validate designs for real-world applications (Rasouli et al., 2019). Therefore, a comprehensive testbed that integrates these diverse factors is essential for ensuring the effectiveness and practicality of AV communication systems in the industry. Therefore, in this paper, we propose a VR-based testbed for evaluating eHMI. VR technology is particularly well-suited for this purpose, as it offers immersive, cost-effective, and easily set-up simulations that can be tailored to meet various evaluation needs (Fuest et al. 2020).

Testbed design

Our testbed uses **ProuVR, a Unity based multi-user platform** (developed by IRT b<>com) **that enables ergonomists to customize virtual environments (VE) according to experimental needs and gather in-depth feedback from end users** (Devreux et al., 2023). When setting up the VE for experiments, ergonomists can configure it by adding objects from an existing library or removing them from virtual scene, as well as replacing and rescaling them from both the PC and VR sides. This flexibility allows for **precise control over the experimental setup, ensuring that it meets the specific requirements of each study.**

Based on a survey of factors influencing pedestrian decision-making in both classical pedestrian–driver interaction studies and modern AV-related research (Rasouli et al., 2019), we identified key elements that can be configured using VR methods. These factors were then categorized into four distinct classes and integrated into our testbed for comprehensive eHMI evaluation: **AV profile, road design, pedestrian characteristics, and weather and lighting conditions.**

AV profile

In our testbed, ergonomists can **choose from a diverse range of virtual vehicle models in the library**, including both aggressive and friendly designs (Dey et al., 2017) in various colors. The system allows for **detailed configuration of vehicle parameters**, including the AV's starting position, initial velocity, braking point, and deceleration rate. Such precise control over vehicle speed and distance from pedestrians allows the ergonomist to measure gap acceptance in time which pedestrians consider safe for crossing. Furthermore, it enables **simulations of different driving strategies** to evaluate how pedestrians interpret the AV's behavior and intention. For instance, it is possible to configure a more defensive driving strategy that involves early deceleration when yielding, or a more aggressive approach that simulates sharp braking just before the pedestrian. As spatial audio has been shown to enhance task performance compared to unimodal sound (Bernhard et al., 2008), our system includes the option to incorporate spatial sounds, such as engine or braking noises of the AV.

To address the debate on whether pedestrians disregard the driver's state (Dey and Terken 2017) or rely on eye contact to assess the driver's intentions (Rasouli et al., 2017a; Walker et al., 2007; Guéguen et al., 2015), our system provides the **option to include or exclude the driver's presence**. Additionally, the integrated eye-tracking system in the VR headset tracks participants' gaze patterns, offering insights into where they focus during interactions. **The size of the vehicle can be adjusted** once the model has been added to the virtual scene. Vehicle size is a crucial factor in the pedestrian decision-making process. Pedestrians generally exhibit greater caution when facing larger vehicles (Das et al., 2005), and a larger vehicle may also increase the likelihood that pedestrians underestimate its arrival time (Caird et al., 1994). Multiple vehicles can be added to the scene to simulate varying traffic volumes as higher traffic volumes generally make pedestrians less likely to cross the street (Ishaque et al., 2008).

Road design

Different modular road structures are provided in the library, including straight roads, curved roads, crossroads, roundabouts, T-junctions, and Y-junctions. Ergonomists can combine different structures to customize the road layout as needed. **The road width can be modified by adjusting the size of the road structure**, as some previous research indicates that road width has impact on pedestrian crossing behavior. For instance, pedestrians tend to exercise more caution on wider roads, while they are more likely to accept smaller traffic gaps on narrower roads (Schmidt et al., 2009; Rasouli et al., 2017b). Moreover, our testbed **offers a variety of street markings**, including traffic signals and zebra crossings, allowing ergonomists to customize their road designs. These features have been shown to significantly influence pedestrian behavior and their level of compliance with traffic laws (Moore et al., 1953; Mortimer et al., 1973).

Once the road layout is established, ergonomists can **define the vehicle's trajectory** within the VE by specifying various control points. The simulation of the vehicle's movement is driven by predefined AV parameters and follows a path generated by a Bezier curve based on these control points, utilizing the Unity engine. This setup also allows ergonomists to explore specific scenarios, such as simulating a vehicle's behavior under law violations and different right-of-way situations.

Pedestrians' characteristics

The state of pedestrians, particularly their speed and direction of movement, influences their crossing decisions. Previous studies have shown that while walking, pedestrians receive better optical flow information, enhancing their ability to estimate the speed and distance of approaching vehicles. As a result, walking pedestrians are generally less conservative in their crossing behavior compared to

those who are standing still (Oudejans et al., 1996). To address this question, our testbed enables ergonomists **to select whether participants stand still at the beginning of the experiment or start with a predefined walking speed**. In the latter scenario, the pedestrian avatar can navigate the VE at the specified velocity using joystick controls.

Whether pedestrians walk alone or in a group also influences their crossing behavior. Group size impacts pedestrian flow, which affects crossing speed. Generally, pedestrians tend to walk slower in denser groups (Wiedemann et al., 1974). Some social factors, such as imitation, also play a role in shaping pedestrian behavior. Studies suggest that individuals in a group are more likely to follow someone, whether they are adhering to or violating traffic laws (Rosenbloom et al., 2009). As group size increases, the likelihood of pedestrians imitating others decreases (Dolphin et al., 1970). To evaluate pedestrian behavior in group contexts, our testbed **provides a variety of animated humanoid walking agents in the library**. These agents can be preconfigured by setting their start and end points, as well as their walking speed. Ergonomists can position the agents within the VE to create pedestrian groups of varying sizes. Additionally, the movement of these agents can be triggered in real time using the WoZ technique to simulate extreme scenarios, such as crossing at a red light, allowing for the study of pedestrian imitation behavior.

Weather and light conditions

Pedestrian behavior is also influenced by weather and lighting conditions (Harrell et al., 1991). For instance, pedestrians typically have poorer speed estimation in bad weather conditions (Sun et al., 2015). Similarly, lower illumination levels impair pedestrians' visual function, making them more likely to take riskier decisions (Rasouli et al., 2019). To enable ergonomists to analyze AV communication forms under various weather and lighting conditions, our testbed **includes options for simulating adverse weather**, such as rainfall using Unity's particle system and wet pavement effects with diffuse light reflections. Additionally, ergonomists can **adjust illumination levels** to simulate different lighting environments.

Conclusion and future work

Ensuring pedestrians understand AVs' intentions is crucial. However, many eHMI evaluation studies overlook key factors that influence pedestrian decision-making. In this ongoing work, we identify four key factors: AV profile, road design, pedestrian characteristics, and weather & lighting conditions. Besides, we propose a VR-based testbed that integrates these elements, allowing ergonomists to configure test environments for a comprehensive assessment of AV communication systems. Future work will focus on further developing testbed functions and conducting user tests to validate its effectiveness.

References

- Bernhard, M., Grosse, K., & Wimmer, M. (2008). Bimodal task-facilitation in a virtual traffic scenario through spatialized sound rendering. *ACM Transactions on Applied Perception (TAP)*, 8(4), 1-22. <https://doi.org/10.1145/2043603.2043606>
- Caird, J. K., & Hancock, P. A. (1994). The perception of arrival time for different oncoming vehicles at an intersection. *Ecological Psychology*, 6(2), 83-109. https://doi.org/10.1207/s15326969eco0602_12
- Chang, C. M., Toda, K., Sakamoto, D., & Igarashi, T. (2017, September). Eyes on a Car: an Interface Design for Communication between an Autonomous Car and a Pedestrian. In *Proc. 9th int. Conf. automotive user interfaces and interactive vehicular applications* (pp. 65-73). <https://doi.org/10.1145/3122986.3122989>

- Currano, R., Park, S. Y., Domingo, L., Garcia-Mancilla, J., Santana-Mancilla, P. C., Gonzalez, V. M., & Ju, W. (2018, September). ¡Vamos! Observations of pedestrian interactions with driverless cars in Mexico. In Proc. 10th Int. Conf. automotive user interfaces and interactive vehicular applications (pp. 210-220). <https://doi.org/10.1145/3239060.3239086>
- Das, S., Manski, C. F., & Manuszak, M. D. (2005). Walk or wait? An empirical analysis of street crossing decisions. *Journal of applied econometrics*, 20(4), 529-548. <https://doi.org/10.1002/jae.791>
- De Clercq, K., Dietrich, A., Núñez Velasco, J. P., De Winter, J., & Happee, R. (2019). External human-machine interfaces on automated vehicles: Effects on pedestrian crossing decisions. *Human factors*, 61(8), 1353-1370. <https://doi.org/10.1177/00187208198363436>
- Devreux, G., Lahoual, D., Zhang, Y., Calvet, G., Jegou, G., & Gody, A. (2023, October). Feedback from the ProuVR Platform: Design, Layout and Assessment of a Nuclear Control Room. In Proceedings of the 18th "Ergonomie et Informatique Avancée" Conference (pp. 1-7). https://doi.org/10.1007/978-3-030-27928-8_27
- Dey, D., & Terken, J. (2017, September). Pedestrian interaction with vehicles: roles of explicit and implicit communication. In Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications (pp. 109-113). <https://doi.org/10.1145/3122986.3123009>
- Dey, D., Martens, M., Eggen, B., & Terken, J. (2019). Pedestrian road-crossing willingness as a function of vehicle automation, external appearance, and driving behaviour. *Transportation research part F: traffic psychology and behaviour*, 65, 191-205. <https://doi.org/10.1016/j.trf.2019.07.027>
- Dey, D., Martens, M., Eggen, B., & Terken, J. (2017, September). The impact of vehicle appearance and vehicle behavior on pedestrian interaction with autonomous vehicles. In Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications adjunct (pp. 158-162). <https://doi.org/10.1145/3131726.3131750>
- Dietrich, A., Maruhn, P., Schwarze, L., & Bengler, K. (2020). Implicit communication of automated vehicles in urban scenarios: Effects of pitch and deceleration on pedestrian crossing behavior. In *Human Systems Engineering and Design II: Proceedings of the 2nd International Conference on Human Systems Engineering and Design (IHSED2019): Future Trends and Applications*, September 16-18, 2019, Universität der Bundeswehr München, Munich, Germany (pp. 176-181). Springer International Publishing.
- Dolphin, J., et al. (1970). Factors influencing pedestrian violations. Unpublished manuscript, Queens University, Kingston, Ontario.
- Eisma, Y. B., van Bergen, S., Ter Brake, S. M., Hensen, M. T. T., Tempelaar, W. J., & de Winter, J. C. (2019). External human-machine interfaces: The effect of display location on crossing intentions and eye movements. *Information*, 11(1), 13. <https://doi.org/10.3390/info110100131>
- Florentine, E., Ang, M. A., Pendleton, S. D., Andersen, H., & Ang Jr, M. H. (2016, October). Pedestrian notification methods in autonomous vehicles for multi-class mobility-on-demand service. In Proceedings of the Fourth International Conference on Human Agent Interaction (pp. 387-392). <https://doi.org/10.1145/2971648.2971712>
- Fraser, N. M., & Gilbert, G. N. (1991). Simulating speech systems. *Computer Speech & Language*, 5(1), 81-99.
- Fridman, L., Mehler, B., Xia, L., Yang, Y., Facusse, L. Y., & Reimer, B. (2017). To walk or not to walk: Crowdsourced assessment of external vehicle-to-pedestrian displays. arXiv preprint arXiv:1707.02698. <https://doi.org/10.48550/arXiv.1707.02698>
- Fuest, T., Maier, A. S., Bellem, H., & Bengler, K. (2020). How should an automated vehicle communicate its intention to a pedestrian?—a virtual reality study. In *Human Systems Engineering and Design II: Proceedings of the 2nd International Conference on Human Systems Engineering and Design (IHSED2019): Future Trends and Applications*, September 16-18, 2019, Universität der Bundeswehr München, Munich, Germany (pp. 195-201). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_30

- Guéguen, N., Meineri, S., & Eyssartier, C. (2015). A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. *Safety science*, 75, 87-89. <https://doi.org/10.1016/j.ssci.2015.01.018>
- Harrell, W. A. (1991). Factors influencing pedestrian cautiousness in crossing streets. *The Journal of Social Psychology*, 131(3), 367-372. <https://doi.org/10.1080/00224545.1991.9713863>
- Hensch, A. C., Neumann, I., Beggiano, M., Halama, J., & Krems, J. F. (2019). Effects of a light-based communication approach as an external HMI for Automated Vehicles--a Wizard-of-Oz Study. *Transactions on Transport Sciences*, 10(2). <https://doi.org/10.5507/tots.2019.012>
- Ishaque, M. M., & Noland, R. B. (2008). Behavioural issues in pedestrian speed choice and street crossing behaviour: a review. *Transport Reviews*, 28(1), 61-85. <https://doi.org/10.1080/01441640701365239>
- Joisten, P., Alexandri, E., Drews, R., Klassen, L., Petersohn, P., Pick, A., ... & Abendroth, B. (2020). Displaying vehicle driving mode--Effects on pedestrian behavior and perceived safety. In *Human Systems Engineering and Design II: Proceedings of the 2nd International Conference on Human Systems Engineering and Design (IHSED2019): Future Trends and Applications*, September 16-18, 2019, Universität der Bundeswehr München, Munich, Germany (pp. 250-256). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_27
- Matthews, M., Chowdhary, G., & Kieson, E. (2017). Intent communication between autonomous vehicles and pedestrians. arXiv preprint arXiv:1708.07123. <https://doi.org/10.48550/arXiv.1708.07123>
- Moore, R. L. (1953). Pedestrian choice and judgment. *Journal of the Operational Research Society*, 4(1), 3-10. <https://doi.org/10.1057/jors.1953.2>
- Mortimer, R. G. (1973). Behavioral evaluation of pedestrian signals. *Traffic engineering*, 44(2), 22-26.
- Oudejans, R. R., Michaels, C. F., Van Dort, B., & Frissen, E. J. (1996). To cross or not to cross: The effect of locomotion on street-crossing behavior. *Ecological psychology*, 8(3), 259-267. https://doi.org/10.1207/s15326969eco0803_4
- Palmeiro, A. R., van der Kint, S., Vissers, L., Farah, H., de Winter, J. C., & Hagenzieker, M. (2018). Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. *Transportation research part F: traffic psychology and behaviour*, 58, 1005-1020. <https://doi.org/10.1016/j.trf.2018.07.020>
- Rasouli, A., & Tsotsos, J. K. (2019). Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE transactions on intelligent transportation systems*, 21(3), 900-918. <https://doi.org/10.48550/arXiv.1805.11773>
- Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2017, June). Agreeing to cross: How drivers and pedestrians communicate. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 264-269). IEEE. <https://doi.org/10.48550/arXiv.1702.03555>
- Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2017). Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles*, 3(1), 61-70. <https://doi.org/10.1109/TIV.2018.2809471>
- Rettenmaier, M., Pietsch, M., Schmidtler, J., & Bengler, K. (2019, June). Passing through the bottleneck--the potential of external human-machine interfaces. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1687-1692). IEEE. <https://doi.org/10.1109/IVS.2019.8814082>
- Risto, M., Emmenegger, C., Vinkhuyzen, E., Cefkin, M., & Hollan, J. (2017, June). Human-vehicle interfaces: The power of vehicle movement gestures in human road user coordination. In *Driving assessment conference* (Vol. 9, No. 2017). University of Iowa. <https://doi.org/10.17077/drivingassessment.1633>
- Rosenbloom, T. (2009). Crossing at a red light: Behaviour of individuals and groups. *Transportation research part F: traffic psychology and behaviour*, 12(5), 389-394. <https://doi.org/10.1016/j.trf.2009.05.002>
- Rothembücher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2016). A field study investigating the interaction between pedestrians and driverless vehicles. In *Proceedings of the International Symposium on Robot and Human Interactive Communication*, IEEE. <https://doi.org/10.1109/ROMAN.2016.7745210>

- Schmidt, H., Terwilliger, J., AlAdawy, D., & Fridman, L. (2019). Hacking nonverbal communication between pedestrians and vehicles in virtual reality. arXiv preprint arXiv:1904.01931. <https://doi.org/10.48550/arXiv.1904.01931>
- Schmidt, S., & Faerber, B. (2009). Pedestrians at the kerb—Recognising the action intentions of humans. *Transportation research part F: traffic psychology and behaviour*, 12(4), 300-310. <https://doi.org/10.1016/j.trf.2009.02.003>
- Song, Y. E., Lehsing, C., Fuest, T., & Bengler, K. (2018). External HMIs and their effect on the interaction between pedestrians and automated vehicles. In *Intelligent Human Systems Integration: Proceedings of the 1st International Conference on Intelligent Human Systems Integration (IHSI 2018): Integrating People and Intelligent Systems*, January 7-9, 2018, Dubai, United Arab Emirates (pp. 13-18). Springer International Publishing. https://doi.org/10.1007/978-3-319-73888-8_3
- Sucha, M., Dostal, D., & Risser, R. (2017). Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention*, 102, 41-50. <https://doi.org/10.1016/j.aap.2017.02.018>
- Sun, R., Zhuang, X., Wu, C., Zhao, G., & Zhang, K. (2015). The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. *Transportation research part F: traffic psychology and behaviour*, 30, 97-106. <https://doi.org/10.1016/j.trf.2015.02.002>
- Walker, I. (2007). Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. *Accident Analysis & Prevention*, 39(2), 417-425. <https://doi.org/10.1016/j.aap.2006.08.010>
- Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019). Crossing the Street Across the Globe: A Study on the Effects of eHMI on Pedestrians in the US, Germany and China. In *HCI in Mobility, Transport, and Automotive Systems: First International Conference, MobiTAS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings 21* (pp. 515-530). Springer International Publishing. https://doi.org/10.1007/978-3-030-22666-4_37
- Wiedemann, R. (1974). *Simulation des straßenverkehrsflusses*. *schriftenreihe heft 8*. Institute for Transportation Science, University of Karlsruhe, Germany.
- Zimmermann, R., & Wettach, R. (2017, September). First step into visceral interaction with autonomous vehicles. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 58-64). <https://doi.org/10.1145/3122986.3122988>

Exploring Boundary Visualization in AR for Co-located Collaborative Workspace Management

Inoussa Ouedraogo¹, Huyen Nguyen², and Patrick Bourdot¹

¹*Université Paris-Saclay, CNRS, LISN, VENISE Team, Orsay, France*

²*Université Paris-Saclay, CNRS, LISN, ARAI Team, Orsay, France*

Corresponding authors: inoussa.ouedraogo@lisn.fr, huyen.nguyen@lisn.fr, patrick.bourdot@lisn.fr

Keywords: Co-located Collaboration, Workspace Partitioning, Augmented Reality

Abstract

Recent research on co-located collaboration using Augmented Reality (AR) technologies demonstrates that physical workspace partitioning is essential to meet the privacy needs of collaborators. However, there is still a lack of research on how to display workspace partitioning in AR. This work aims to help users become aware of their personal and public workspaces in the physical space to perform collaborative tasks more effectively. In this paper, we propose and investigate the effects of different AR workspace boundary visualization methods, including *Line*, *Opaque*, and *Semi-transparent*.

Introduction

Many research works highlight the advantages of Augmented Reality (AR) in immersive collaboration. One key application of AR is workspace partitioning. For instance, Lee *et al.* 2022 explored virtual separators to reduce visual distractions in open-plan workspaces, demonstrating that virtual partitions can reduce distractions and allow users to personalize their space, leading to improved experiences in co-located shared environments. Similarly, Jackson *et al.* 2024 introduced the Workspace Guardian system to promote awareness of personal workspace boundaries among AR users. Their study found that automatically defined boundaries significantly reduced workspace encroachments and enhanced user comfort compared to user-defined ones. Although spatial partitioning in AR has been shown to improve workflow in co-located collaboration, few studies have examined its effects in industrial assembly settings where collaborative model assembly is required.

In this context, our study aims to investigate the effect of different types of virtual boundaries superimposed on the floor for workspaces partitioning. We performed a user study to compare three basic visual boundary cues, namely: *Line*, *Opaque*, and *Semi-transparent* boundary, in terms of performance, usability, and cognitive load.

Workspace Partitioning

We have implemented the space partitioning algorithm proposed in Ouedraogo *et al.* 2024, following several steps to generate the different types of workspace boundaries for the experimental study:

- **Step 1:** Scan the physical space to understand the scene, obtain the environment mesh data, and remove the ceiling and walls from the mesh.
- **Step 2:** Partition the space into $n + 1$ spaces using k-means clustering, where n is the number of collaborators.
- **Step 3:** Compute new meshes to define the boundaries of different workspaces.
- **Step 4:** Assign each user a private workspace and indicate the public workspace.

For the experimental study, we proposed three visualization modes to mark the boundary using varying opacity levels: *Line*, *Opaque*, and *Semi-transparent* mesh texture. In the case of *Opaque* (Figure P 39C) and *Semi-transparent* boundaries (Figure P 39D), the Delaunay triangulation algorithm is applied to each cluster from the k-means output to identify the triangles forming the new mesh during step 3. These meshes are rendered with appropriate materials: Unity's default opaque materials for *Opaque* boundary, and MRTK materials with near-fading for *Semi-transparent* boundary, allowing areas near the user to appear transparent and gradually transition to opacity. For *Line* boundary (Figure P 39B), the border vertices of each subspace's mesh are identified and connected to form *line* meshes.

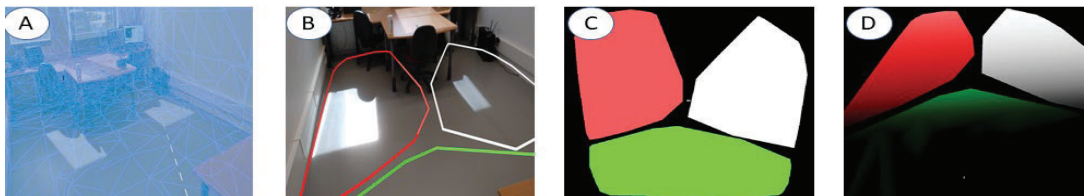


Figure P 38. A) A room's mesh is obtained by scanning using Microsoft HoloLens 2. Then, the space can be partitioned using three types of boundary visualisation generated on the floor to indicate the different workspaces (green is public area, red and white are pri

Hypotheses and Experimental Design

We conducted a user study to evaluate the effectiveness of three visualization modes for virtual boundaries in collaborative AR. The experiment followed a within-subjects design with one factor (Technique), which included three conditions (Figure P 39): *Line* boundary, *Opaque* boundary, and *Semi-transparent* boundary. The order of these techniques was counterbalanced across participants using a balanced Latin square. We formulated the following hypotheses:

- **H0:** The time required to complete tasks will be similar across all conditions (*Line*, *Opaque*, and *Semi-transparent* boundary).
- **H1:** *Line* boundary technique will be more usable, require less cognitive load, and be preferred by participants compared to the other techniques.

Six pairs of participants, aged between 18 and 32 (nine men and three women), were recruited from our Department and University to take part in the experiment.

Experiment Setup and Tasks

The system, developed in Unity3D 2021.3.15f1, includes two client hosts using Microsoft HoloLens 2 and a server running on a Windows PC with an Intel Xeon W-2135 processor, 32GB RAM, and an RTX 4000, connected through 5-GHz Wi-Fi. The experiment was conducted in an office space measuring 5m × 3.5 m. The collaborative task involved decorating an apartment, with each participant responsible for different parts of the interior. After discussing the design plan based on a 2D image (Figure P 40B), they agreed on their individual tasks. Three 3D apartment models were randomly assigned to the boundary visualization conditions, and participants had to collect six specific objects from their workspaces for each model. One participant handled decorative items (e.g., flowers, plants, lamps) (Figure P 40A), while the other managed furniture and appliances (e.g., chairs, tables, fridge) (Figure P 40C). Roles were randomly assigned at the start of each condition.



Figure P 39. Example of a task in experiment 1: A) User A's shelf contains decorative items such as lamps, flowers, plants, a TV, and a laptop. B) 3D apartment model and 2D image highlighting missing objects with red circles to indicate where they should be placed

Procedure

The 40-minute experiment began with participants receiving instructions, signing consent forms, and completing a demographic questionnaire. They were then equipped with Microsoft HoloLens 2 headsets and underwent a training session to familiarize themselves with the device and interactions. At the start of each task, they read instructions and start the task by pressing a start button. After each task, they completed the NASA-TLX and SUS questionnaires. Once the experiment ended, they ranked the three techniques based on their preferences.

Data Collection

We registered 24 trials: 3 techniques × 12 participants. We collected the following measures:

- **Quantitative data:** Task Completion Time (TCT) refers to the duration required to finish the task under various conditions.
- **Subjective data:** Technique usability was evaluated using the System Usability Scale (SUS), cognitive load was measured with the NASA-TLX, and preferences were ranked from best (1) to worst (3) using an Overall Ranking questionnaire.

Results and Conclusion

This study evaluated the effects of three boundary visualization techniques (*Line*, *Opaque*, and *Semi-transparent*) on Task Completion Time (TCT), cognitive load, and usability in collaborative AR. A one-way repeated-measures ANOVA test was used for the TCT data, which revealed no significant

differences between the techniques ($F_{2,22} = 0.36$, $p = 0.7$). This indicates that the boundary type does not significantly affect task efficiency, thus supporting the null hypothesis **H0**.

For cognitive load, the Friedman test was applied to the NASA-TLX scores, which showed no significant effect across conditions ($\chi^2(2) = 3.5$, $p = 0.17$). This suggests that the cognitive load was similarly low for all techniques. The usability scores, calculated using the SUS method, were also analyzed using the Friedman test, which again indicated no significant effect between techniques ($\chi^2(2) = 0.12$, $p = 0.94$). However, *Line* boundary achieved a slightly higher mean SUS score above the usability threshold of 68 ($\mu = 68.33$, $\sigma = 24.38$), whereas *Opaque* and *Semi-transparent* conditions scored below this threshold.

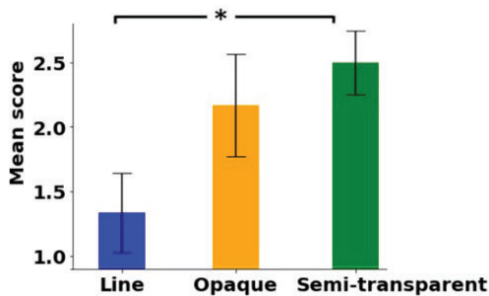


Figure P 40. Mean of overall user ranking of the three boundary visualisation techniques (1 indicating the best, 3 the worst), with error bars showing 95% confidence intervals (CI).

The overall ranking of the techniques was analyzed using the Friedman test, which revealed a significant difference in participant preferences ($\chi^2(2) = 8.66$, $p = 0.013$) (Figure P 41). Post-hoc analysis using the Wilcoxon Signed Rank test with Bonferroni correction showed a significant difference between *Line* and *Semi-transparent* conditions ($p = 0.014$), while no significant differences were found between *Line* and *Opaque* ($p = 0.27$) or *Opaque* and *Semi-transparent* ($p = 1$). Participant feedback further supported the preference for *Line* boundary. Several participants (e.g., P4 and P7) appreciated its minimal interference, while others found *Opaque* and *Semi-transparent* conditions visually overwhelming or uncomfortable (e.g., P9 and P8).

Overall, the study suggests that while task efficiency is unaffected by the type of boundary visualization, *Line* technique was subjectively preferred for its simplicity and reduced cognitive load, providing partial support for hypothesis **H1**.

ACKNOWLEDGMENTS

This work was supported by French government funding managed by the National Research Agency (ANR) under the Investments for the Future program (PIA) with the grant ANR-21-ESRE-0030 (Equipex+ CONTINUUM project).

References

- Jackson, B., Lor, L., & Heggeseth, B. C. (2024). Workspace Guardian: Investigating Awareness of Personal Workspace Between Co-Located Augmented Reality Users. *IEEE Transactions on Visualization and Computer Graphics*.
- Lee, H., Je, S., Kim, R., Verma, H., Alavi, H., & Bianchi, A. (2022). Partitioning open-plan workspaces via augmented reality. *Personal and Ubiquitous Computing*, 1-16.
- Ouedraogo, I., Nguyen, H., & Bourdot, P. (2024, October). Where to Draw the Line: Physical Space Partitioning and View Privacy in AR-based Co-located Collaboration for Immersive Analytics. In *Proceedings of the 2024 ACM Symposium on Spatial User Interaction* (pp. 1-12).

AI-Augmented Mixed Reality for Hand Motion Assistance: Bridging the Gap in Parkinson's Disease Rehabilitation

Xinjun Li, and Zhenhong Lei

Cornell University, Rhode Island School of Design, Ithaca, NY, USA

Corresponding author: lnacademic@gmail.com

Authors contributed equally and shall both be considered first authors.

Keywords: Augmented Reality, Mixed Reality, Assistive Device, Haptic Feedback, Parkinson's Disease, XR Healthcare, Hand Rehabilitation, Ergonomic Hand Motion

Introduction

Parkinson's disease (PD) affects an estimated 1 million Americans, with approximately 1.6% of individuals over the age of 65 receiving treatment annually (Bamford and Henderson, 2021). Characterized by tremors, a shuffling gait, bradykinesia, muscle rigidity, and postural instability, PD leads to significant physical impairment and increased mortality risk. Its management is further complicated by autonomic dysfunction, depression, dementia, and psychosis (Schwalbe et al., 2023). As the older population grows, innovative rehabilitative approaches are needed. Integrating mixed reality (MR) and artificial intelligence (AI) offers a promising solution, providing personalized interventions and predictive analytics to enhance hand dexterity in PD patients.

Methodology

Comparative Evaluation of Existing Therapeutic Modalities

Current therapeutic interventions for tremor disorders in older populations are typically divided between mechanical assistive devices and virtual rehabilitation strategies. Mechanical solutions, such as functional gloves and utensil grips, address the physical effects of tremors by compensating for reduced strength and dexterity. In contrast, virtual rehabilitation approaches, particularly those leveraging mixed reality (MR) technologies, focus on improving motor function and sensory feedback through immersive, real-world simulations that offer structured rehabilitation exercises (Pereira et al., 2020). While mechanical devices provide immediate support, MR-based platforms promote long-term motor skill development through personalized exercises (Mois et al., 2019; Kolsanov et al., 2020). However, a significant gap in the research is the lack of integration between these two approaches, missing the potential benefits of combining physical assistance with immersive digital rehabilitation (Dineshkumar et al., 2022; Gu et al., 2022). This paper introduces a novel MR system that bridges this gap by integrating ergonomic hand support with MR-based rehabilitation exercises. Developed iteratively with feedback from users and professionals, the system offers both immediate physical assistance and sustained functional improvements, ensuring accessibility and effectiveness.

Integrated Mixed Reality Solution

This study presents an innovative mixed-reality framework that integrates ergonomic hand motion support with MR rehabilitation, combining physical tremor stabilization and motor skill enhancement exercises (Gonsler and Lei, 2022). The system features an ergonomically adaptable hand motion assistance device, while the MR component uses AI algorithms to adjust exercise difficulty in real-time based on user progress, promoting motor skill development through interactive, gamified tasks. Telehealth capabilities enable remote monitoring and consultations, improving accessibility (Li and Lei, 2024). The system personalizes rehabilitation using real-time data, while voice-activated controls, powered by Natural Language Processing (NLP), enhance usability. Extensive validation and iterative refinements ensure the system's functionality and comfort, advancing personalized geriatric rehabilitation. (Figure P 42)

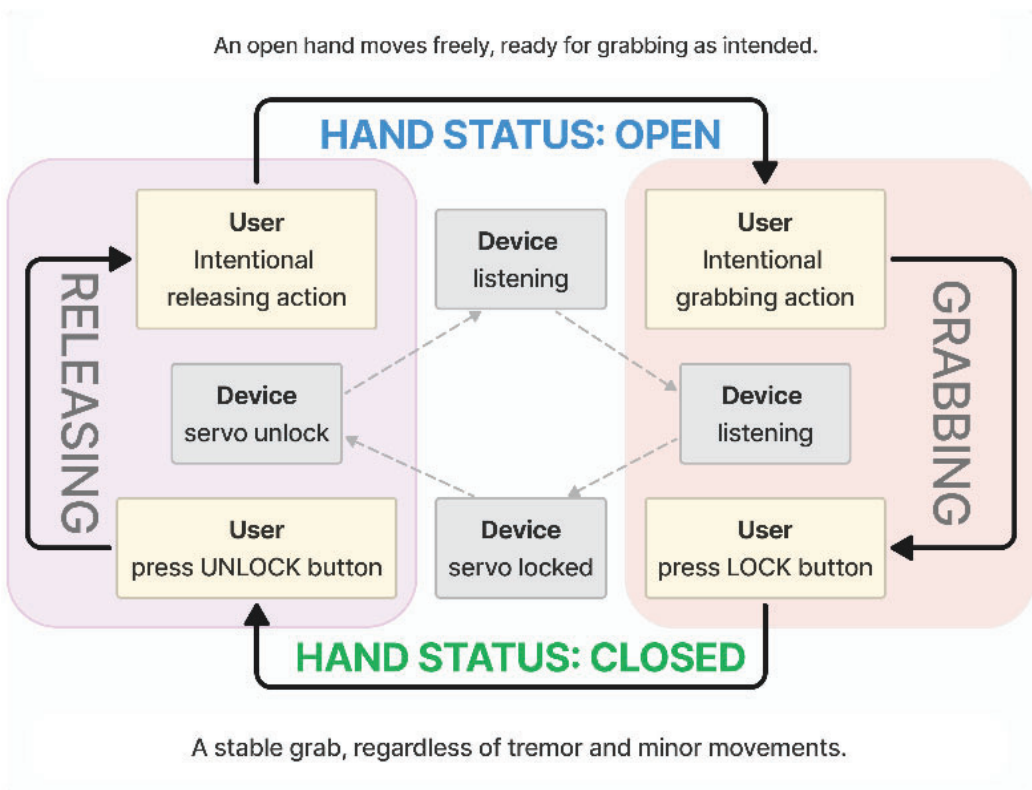


Figure P 41. Functional diagram

Prototype Design Process

Physical Prototype Design

In developing our prototype to enhance hand dexterity rehabilitation through mixed reality (MR) interventions augmented by AI, we created a 3D-printed model to examine how the gripping mechanism aligns with the natural curvature of the human palm. By integrating biomechanical precision with advanced technology, we enhanced grip stability and precision, resulting in an asymmetrical ergonomic design that facilitates smooth hand movements and active grip support. The prototype

features a dual-component system connected by a hinge, simulating the natural opposition between the thumb and fingers necessary for gripping. A servo motor embedded in the thumb component functions as both an actuator and stabilizer, enabling real-time responsiveness to hand movements. The system is powered by an Arduino Nano board with low-energy Bluetooth, and a battery housed in one of the handles powers both the motor and controller. This anatomical design ensures smooth transitions between open and closed hand positions, closely replicating natural hand movements.

In practical use, such as picking up a book, the process is divided into two cycles: grabbing and releasing. During the grabbing cycle, the device remains idle, allowing the user to intuitively grasp the object, while the hinge mechanism stabilizes hand movements, reducing tremors. A button on the user's phone switches the device to grip mode, locking the position of the handles to maintain a firm hold on the book, thus minimizing muscular strain and tremors. In the releasing cycle, the user switches the device back to idle mode, allowing the handles to follow natural hand movements and facilitating the release of the book with ease and without resistance. (Figure P 43)



Figure P 42. Physical prototype on a human hand

MR System Development

In developing our prototype to improve hand dexterity rehabilitation for geriatric patients with movement disorders, we designed a mixed reality (MR) system powered by the Unity engine and accessible through a smartphone application. This system utilizes the device's camera to provide MR-based assistive guidance, allowing users to point the camera at objects, such as books, and receive interactive instructions that facilitate effective handling. The system's core functionality is driven by AI algorithms based on machine learning that analyze objects in real time, evaluating their dimensions, shape, and optimal orientation for grasping. The system's gyroscope integration ensures synchronized motion feedback, enabling smooth user-object interactions as users follow the MR-guided recommendations. The decision to use a mobile device was informed by its accessibility, portability, and cost-efficiency, making it suitable for broad adoption in geriatric care. While we acknowledged challenges, such as the need to operate the device with one hand, these were mitigated through an ergonomic design and intuitive interface that reduces physical strain. Early formative studies confirmed the usability of this platform, demonstrating its ability to support flexible rehabilitation sessions and enhance patient engagement with therapeutic exercises. (Figure P 44)



Figure P 43. Smartphone MR interface

Result

The initial evaluation of our AI-enhanced mixed reality (MR) hand motion assistance device for older patients with essential tremors or Parkinson's disease highlighted the successful integration of ergonomic stabilization with MR-based rehabilitation exercises. Through structured interviews and observations, participants reported improved daily task performance and reduced tremor-related disruptions. The ergonomic design was well-received for its comfort and user-centered approach, showing promise as an assistive tool. The MR component promoted sustained engagement through interactive exercises, while telehealth enabled remote monitoring and guidance. However, further quantitative analysis is needed to assess the device's effectiveness fully.

Conclusion

The development and initial evaluation of our AI-augmented mixed reality (MR) hand motion assistance device represent significant progress in geriatric care, neurorehabilitation, and assistive technologies. By combining MR and AI, the system introduces a novel approach that surpasses traditional rehabilitation methods. While early results are promising, future research will focus on quantitative analysis to validate efficacy and refine the design. This ongoing work aims to maximize the potential of MR and AI in rehabilitation, setting new standards for personalized, technology-driven interventions and improving outcomes for aging populations with movement disorders.

References

- Bamford, A., & Henderson, E. J. (2021). Parkinson's disease in older people. *Medicine*, 49(1), 56-61. <https://doi.org/10.1016/j.mpmed.2020.10.008>.
- Dineshkumar, V., Dolly, D. R. J., Jagannath, D. J., & Peter, J. D. (2022). Assistive methodologies for Parkinson's disease tremor management—a health opinion. *Frontiers in Public Health*, 10, 850805. <https://doi.org/10.3389/fpubh.2022.850805>.
- Gonsler, I., & Lei, Z. (2021, October). Prototype of Force Feedback Tool for Mixed Reality Applications. In 2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (pp. 508-509). IEEE. <https://doi.org/10.1145/3514206>

- Gu, Y., Xu, Y., Shen, Y., Huang, H., Liu, T., Jin, L., ... & Wang, J. (2022). A review of hand function rehabilitation systems based on hand motion recognition devices and artificial intelligence. *Brain Sciences*, 12(8), 1079. <https://doi.org/10.3390/brainsci12081079>.
- Kolsanov, A., Chaplygin, S., Rovnov, S., & Ivaschenko, A. (2020). Augmented Reality application for hand motor skills rehabilitation. *International Journal of Advanced Computer Science and Applications*, 11(4), 51. <https://doi.org/10.14569/ijacsa.2020.0110408>.
- Li, X., & Lei, Z. (2024, March). Ergonomic Hand Motion Assistance and AR Rehabilitation: Bridging the Gap in Tremor Disorder Therapies Through Mixed Reality. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 322-325). IEEE. <https://doi.org/10.1109/VRW62533.2024.00064>.
- Mois, G., Mackin, T. E., Datta, A., Koon, L. M., Rogers, W. A., Mitzner, T. L., & Beer, J. M. (2019, November). Perceptions from people aging with a mobility impairment towards using tele-technology for exercise. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 63, No. 1, pp. 11-15). Sage CA: Los Angeles, CA: SAGE Publications.
- Pereira, M. F., Prahm, C., Kolbensschlag, J., Oliveira, E., & Rodrigues, N. F. (2020). Application of AR and VR in hand rehabilitation: A systematic review. *Journal of Biomedical Informatics*, 111, 103584. <https://doi.org/10.1016/j.jbi.2020.103584>.
- Schwalbe, M., Satz, S., Miceli, R., Hu, H., & Manelis, A. (2023). Hand dexterity is associated with the ability to resolve perceptual and cognitive interference in older adults: pilot study. *Geriatrics*, 8(2), 31. <https://doi.org/10.3390/geriatrics8020031>.

Integrating XR Technologies into Education: Bridging the Gap Between Educators & Developers

Karoline Wirbatz

University of Münster, Münster, Germany

Corresponding author: kwirbatz@uni-muenster.de

Keywords: XR Technologies, Education, Pedagogical Innovation, Digital Literacy

Introduction

The Pedagogical Alliance for XR Technology in (Teacher) Education (PAX) is a collaborative Erasmus+ project addressing the challenges and opportunities of integrating Extended Reality (XR) technologies in primary, secondary, and vocational education. In response to the European guidelines for the "Digital Decade" (European Commission, 2021) and the acceleration of digital transformation due to the COVID-19 pandemic, PAX aims to bridge the gap between educators and XR developers to foster immersive and effective learning environments.

Key initiatives of PAX include:

1. **PAX Research for Open Access Collection:** Building a comprehensive database of the latest research on XR technologies, providing an open reference source for educators, developers, and other stakeholders.
2. **PAX Framework for Teachers:** Developing a digital skills framework for teachers, with competence descriptors for the pedagogical and subject-specific integration of XR technologies into classroom practice. This framework will be integrated into the curricula of partner universities and beyond.
3. **PAX Design Guidelines XR Developers:** Creating a research-based framework to help XR designers understand the didactic affordances and challenges of implementing XR in educational settings. This will include practical guidelines for the effective design and use of XR technologies.
4. **PAX Training Space:** Providing VR training environments for both teachers and developers to facilitate hands-on engagement with XR technologies. These spaces aim to merge the digital and physical aspects of teaching and learning.

5. **PAX Refinement Hubs:** Establishing hubs where teachers and XR developers can pilot and provide feedback on XR applications to ensure continuous improvement and refinement of the technologies.

PAX promotes digital literacy and inclusion by equipping educators with the necessary XR skills and fostering cross-institutional synergies. PAX envisions a future where XR technologies significantly enhance learning outcomes and educational equity.

This project underscores the importance of a sustainable and evidence-based approach to integrating XR in education, ultimately aiming to revolutionize learning dynamics and prepare educators and students for the digital future.

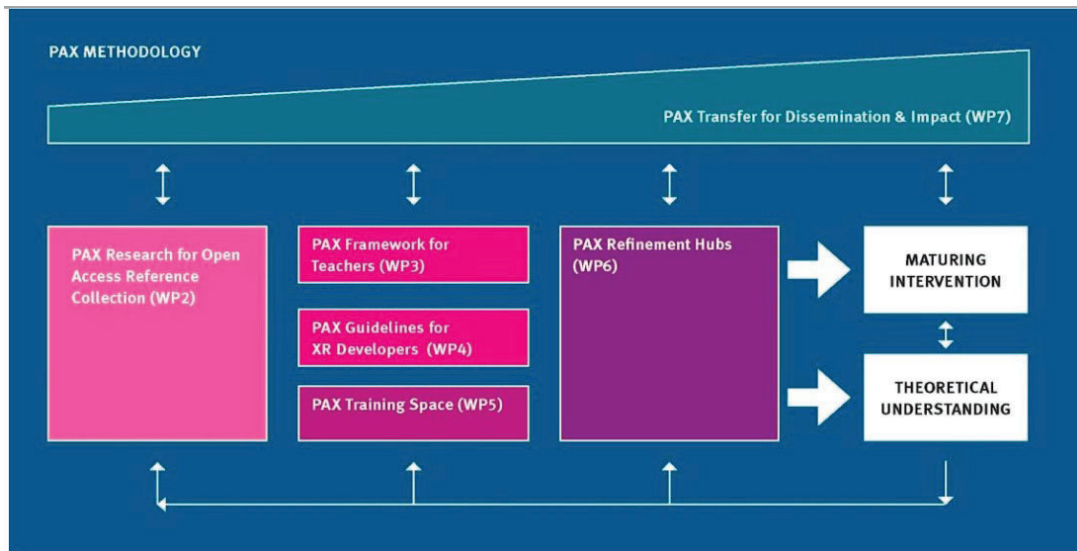


Figure P 44. PAX Methodology

References

European Commission, Directorate-General for Communications Networks, Content and Technology (2021). 2030 digital compass: the European way for the digital decade. Publications Office of the European Union. <https://data.europa.eu/doi/10.2759/425691>

More Than One Milan": Multisensory Experience with an Olfactory Map Application

Lai Pin Jia, Marco Rossoni, Marina Carulli, and Monica Bordegoni

Department of Mechanical Engineering, Politecnico di Milano, Italy

Corresponding author: monica.bordegoni@polimi.it

Keywords: Olfactory Design, Multisensory Integration, Urban Exploration, Human-Computer Interaction (HCI), User-Centered Design, Augmented Reality

Interactive and Multisensory Olfactory Map

The integration of Augmented Reality (AR) and multisensory design in urban exploration offers transformative potential for enhancing user engagement. This paper explores how the "More Than One Milan" project applied AR technology to create a richer, multisensory experience that augments traditional navigation methods with olfactory cues, presenting a novel way of interacting with urban environments. By combining AR and olfactory stimuli, users experience the city not just as a static map but as an interactive, multisensory landscape that fosters immersion and memory retention.

In recent times, designers have increasingly acknowledged the importance of smell across various design domains, especially within the context of Human-Computer Interaction (HCI). Traditionally, urban navigation tools have prioritized visual and auditory elements, often overlooking the role of smell, which is deeply connected to memory and emotion. "More Than One Milan" addresses this gap by incorporating smell into the design process as a core component of the user experience. This approach emphasizes the significance of adding smells to visual and auditory interactions, creating a more emotionally engaging multisensory experience.

The project seeks to enhance users' sensory perceptions through the use of AR, which guides them to specific locations where olfactory stimuli such as grapefruit, myrrh, cedarwood, spearmint, vanilla, and lavender are emitted via a portable Olfactory Display (OD). The OD is based on the "liquid atomizer technology". This odor generation technology used is based on piezoelectric atomizers that allow the scent to be delivered through direct injection, meaning the release of scented particles close to the user's nose. The water atomizer system was carefully designed to ensure a controlled and consistent scent release. Specifically, the device includes a liquid atomizer module that consists of an ultrasonic transducer and a circuit board powered with 5V direct current. By utilizing this technology, users can experience a reliable and synchronized delivery of fragrances associated with specific locations on the olfactory map. The atomizer mechanism is designed to disperse the scent in a fine mist, providing users with a pleasant and immersive olfactory experience. Each location is paired with a specific fragrance that deepens users' emotional connections with their surroundings. The Olfactory Display uses NFC (Near Field Communication) technology to trigger the release of scents when users tap NFC-enabled

cards collected from various locations in the city. Once triggered, the device disperses the selected scent and communicates the interaction to the app via Bluetooth.

This Bluetooth connection allows the app to sync the olfactory data with real-time AR overlays, providing users with visual and auditory cues in alignment with the scent experience. This seamless integration of NFC-triggered olfactory displays and Bluetooth communication enhances the overall multisensory journey, creating a cohesive interaction between the physical and digital worlds.

AR serves as a crucial tool in this multisensory framework by providing real-time visual overlays and audio cues that align with the olfactory experience. This interaction allows users to engage with the urban landscape in a more intuitive and immersive way. Six city landmarks act as key stops in the exploration, where users tap NFC-enabled cards to activate the OD, releasing corresponding fragrances. These olfactory cues enhance memory formation and create lasting emotional connections with the places visited.

By shifting the focus from a purely visual and auditory experience to one that incorporates smell, "More Than One Milan" showcases the potential of AR to reshape urban exploration. This multisensory approach makes the experience of navigating urban spaces more immersive and personalized, fostering stronger emotional and cognitive connections with the environment.

As AR technology evolves, it offers more possibilities for creating personalized, sensory-rich experiences in urban environments. Beyond urban exploration, the combination of AR and multisensory design holds vast potential across sectors such as tourism, education, and cultural heritage preservation, where AR could enhance the interaction with environments in a more engaging, memorable way.

In conclusion, the "More Than One Milan" project demonstrates the potential of AR to transform urban exploration through the integration of multisensory design, particularly olfactory experiences. This combination not only enhances the user experience but also paves the way for innovative, immersive interactions with cities. The project serves as a foundation for future applications in tourism, education, and cultural heritage, highlighting the broader impact of multisensory AR design.

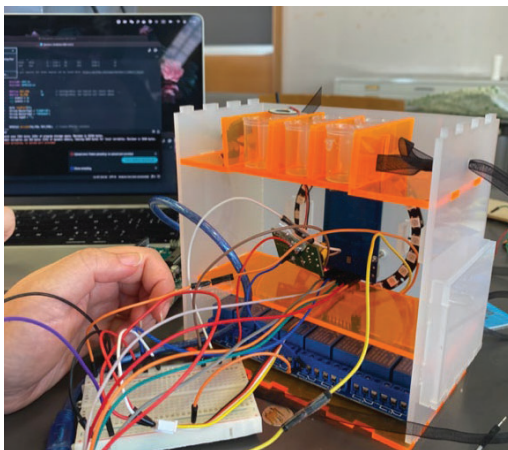


Figure P 45. The "More Than One Milan" olfactory display prototype.



Figure P 46. User is testing the "More Than One Milan" olfactory display.

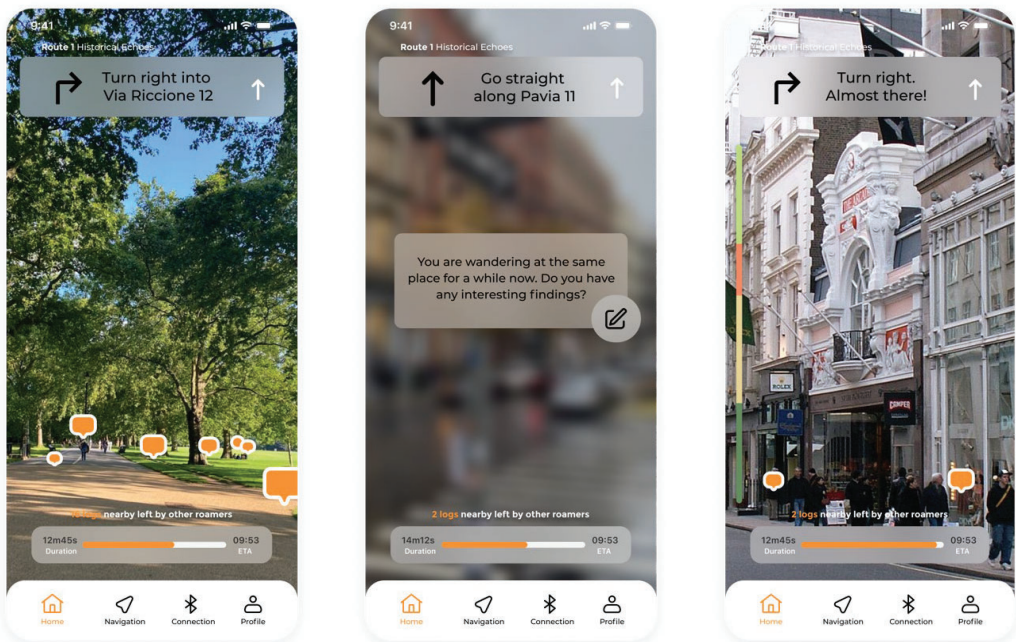


Figure P 47. "More Than One Milan" app leveraging Augmented Reality.

References

- Amores, J., Hernandez, J., Dementyev, A., Wang, X., & Maes, P. (2018, July). Bioessence: A wearable olfactory display that monitors cardio-respiratory information to support mental wellbeing. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 5131-5134). IEEE. <https://doi.org/10.1109/EMBC.2018.8513627>
- Brianza, G., Cornelio, P., Maggioni, E., & Obrist, M. (2021, August). Sniff before you act: exploration of scent-feature associations for designing future interactions. In IFIP conference on human-computer interaction (pp. 281-301). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-85616-8_17
- Dinh, H. Q., Walker, N., Hodges, L. F., Song, C., & Kobayashi, A. (1999, March). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In Proceedings IEEE Virtual Reality (Cat. No. 99CB36316) (pp. 222-228). IEEE. <https://doi.org/10.1109/VR.1999.756955>
- Maggioni, E., Cobden, R., Dmitrenko, D., Hornbæk, K., & Obrist, M. (2020). SMELL SPACE: mapping out the olfactory design space for novel interactions. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(5), 1-26. <https://doi.org/10.1145/3402449>
- Obrist, M., Velasco, C., Vi, C., Ranasinghe, N., Israr, A., Cheek, A., ... & Gopalakrishnakone, P. (2016). Sensing the future of HCI: Touch, taste, and smell user interfaces. *interactions*, 23(5), 40-49. <https://doi.org/10.1145/2978576>

Integrating Extended Reality (XR) in Teaching Sustainability: A Qualitative Study with Student Teachers across the Disciplines of STEM, Humanities, and Languages

Rachel Farrell, Eileen Bowman, Marelle Rice, and Mark Baldwin

University College Dublin, Dublin, Ireland

Corresponding author: rachel.farrell@ucd.ie

Key words: Extended Reality (XR), Sustainability Education, Teacher Education, Adaptive Expertise, Expansive Learning, Constructionism

Abstract

This study examines the integration of Extended Reality (XR) technologies—including Virtual, Augmented, and Mixed Reality—into sustainability education in STEM, Humanities, and Language subjects at the secondary education level. Using the theoretical frameworks of adaptive expertise, expansive learning, and constructionism, the research employs qualitative methods, such as a narrative literature review and focus groups with 20 student teachers from two teacher education programmes at a university in Ireland. The study investigates how XR promotes experiential learning, critical thinking, and collaborative problem-solving in sustainability education. The focus is on understanding how XR can improve teaching practices and identifying the challenges and opportunities of using XR to teach sustainability across various subjects. The findings show that student teachers are generally positive about XR's effectiveness in enhancing sustainability education, although experiences vary by discipline. Key barriers to XR integration include limited access, affordability, and usability, particularly in resource-constrained placement schools. Recommendations from the study suggest integrating specific XR training modules into teacher education programmes to equip future teachers with practical skills, as well as developing clear guidelines to support XR's integration into sustainability lessons. These guidelines should encourage expansive learning and real-world application across different subjects. The study provides valuable insights into the collaborative creation of learning artefacts, aiming to extend XR's use throughout the education continuum. Additionally, it offers practical advice for educators and policymakers looking to enhance sustainability education with XR technologies.



Introduction

Sustainability education is increasingly critical in preparing students to address global environmental and societal challenges (UNESCO, 2021). Extended Reality (XR)—encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR)—offered immersive and interactive learning experiences that deepened understanding of complex sustainability concepts (Radianti, Majchrzak, Fromm, & Wohlgenannt, 2020). Integrating XR into education has the potential to transform teaching methodologies, promoting experiential learning and critical thinking across various disciplines (Jisc, 2024).

Theoretical Framework

This study is grounded in the theories of adaptive expertise, expansive learning, and constructionism. Adaptive expertise refers to the ability to apply knowledge flexibly and innovatively in new situations (Hatano & Inagaki, 1986; Timperley, Ell, & Le Favre, 2017). Expansive learning emphasizes collaborative problem-solving that transforms activity systems, supporting the collective integration of

XR technologies in teaching sustainability (Engeström, 1987, 2016). Constructionism focuses on learning through creating, emphasizing the importance of students constructing their own knowledge (Papert & Harel, 1991). XR provides student teachers with a platform to actively engage in sustainability concepts by collaboratively designing immersive learning experiences tailored to their subject areas and linked to relevant sustainability issues.

This study is grounded in the theories of adaptive expertise, expansive learning, and constructivism.

Literature Review Summary

- Biology: Simulating Ecosystems and Biodiversity
 - VR enables students to explore ecosystems and understand biodiversity impacts from unsustainable practices (Merchant et al., 2014).
- Physics: Renewable Energy and Climate Simulations
 - XR-based simulations engage students with renewable energy technologies, providing practical insights e.g. manipulation of wind energy variables (Dunleavy & Dede, 2014).
- Mathematics: Visualizing Environmental Challenges
 - AR helps students analyse sustainability data like energy consumption and carbon emissions (Arici et al., 2019).
 - VR enables the application of mathematical models to design sustainable infrastructure (Guo et al., 2021).
- History: Exploring Environmental Histories
 - VR reconstructions illustrate how human actions, such as deforestation, have shaped the environment and enhance the understanding of long-term environmental impacts (Gates-Stuart & Hagan, 2020).
- Geography: Simulating Climate Change and Urban Planning
 - XR visualises landscape changes and models climate impacts, aiding urban planning (Guo et al., 2021).
 - AR facilitates real-time environmental analysis, improving the understanding of sustainability across regions (Arici et al., 2019).
- Language Education: Exploring Global Sustainability
 - XR immerses language learners in sustainability issues, fostering global discussions on environmental challenges (Yilmaz & Arici, 2019).
 - AR links language learning to indigenous sustainability practices, enriching cultural understanding (Krajovi et al., 2021).

Methodology

In order to answer the research question “How do student teachers perceive and utilise Extended Reality (XR) technologies in teaching sustainability across STEM, Humanities, and Language subjects in post-primary education?”, a qualitative research design is employed, utilising a narrative literature review and semi-structured focus groups to gather in-depth insights (Creswell & Poth, 2018). A purposive sample of 20 student teachers from the BEd Languages and Professional Master of Education programmes were identified, representing three disciplines: **STEM**: Focusing on applying XR to teach sustainability concepts in science and mathematics; **Humanities**: Emphasising social, ethical, and cultural aspects of sustainability and **Languages**: Incorporating sustainability themes into language education.

Thematic analysis identified patterns and themes across the data (Braun & Clarke, 2020). NVivo software enabled efficient coding and organisation of qualitative data (Zamawe, 2015).

Summary of Findings

Student teachers generally view XR as an effective tool for enhancing sustainability education, though experiences vary across disciplines. XR fosters adaptive expertise and promotes immersive, collaborative learning. However, key challenges include limited access, affordability, and usability, particularly in resource-constrained schools. For successful integration, XR must align with curriculum objectives and complement traditional teaching methods. Additionally, institutional support, training, and ongoing professional development are essential for embedding XR in sustainability education.

Recommendations

Teacher education programmes should include XR training modules to equip student teachers with practical skills in using VR, AR, and MR technologies. These modules should emphasise fostering adaptive expertise and applying constructivist learning principles to create interactive, student-centred experiences. Schools should also adopt clear guidelines for integrating XR into sustainability lessons across STEM, Humanities, and Languages, with specific examples promoting expansive learning through collaboration, problem-solving, and critical thinking, such as virtual labs or real-time data analysis for environmental studies.

References

- Arici, F., Yilmaz, R. M., Caliklar, Ş., & Yilmaz, B. (2019). Research trends in the use of augmented reality in science education: Content and bibliometric mapping analysis. *Computers & Education*, 142, 103647. <https://doi.org/10.1016/j.compedu.2019.103647>
- Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328-352. <https://doi.org/10.1080/14780887.2020.1769238>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.
- Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 735-745). Springer. https://doi.org/10.1007/978-1-4614-3185-5_59
- Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Orienta-Konsultit.
- Engeström, Y. (2016). *Studies in expansive learning: Learning what is not yet there*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316225363>
- Gates-Stuart, E., & Hagan, S. (2020). Virtual reality and visualisation in environmental heritage. *Digital Creativity*, 31(4), 291-310. <https://doi.org/10.1080/14626268.2020.1793198>
- Guo, J., Song, B., Ye, L., & Li, P. (2021). Immersive virtual reality for scientific visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 27(10), 3663-3683. <https://doi.org/10.1109/TVCG.2021.3059402>
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262-272). W. H. Freeman.
- Jisc. (2024). *XR in education: Transforming learning experiences*. Retrieved from <https://www.jisc.ac.uk/reports/xr-in-education>

- Krajovi, M., Soukup, D., & Kovařík, P. (2021). Augmented reality in education: Teaching indigenous knowledge and sustainability. *Journal of Indigenous Studies*, 7(1), 92-107. <https://doi.org/10.1177/09726527211051729>
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29-40. <https://doi.org/10.1016/j.compedu.2013.07.033>
- Papert, S., & Harel, I. (1991). *Constructionism*. Ablex Publishing.
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. <https://doi.org/10.1016/j.compedu.2019.103778>
- Timperley, H., Eil, F., & Le Fevre, D. (2017). *Developing adaptive expertise through professional learning communities*. Springer.
- UNESCO. (2021). *Education for sustainable development: A roadmap*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000374802>
- Yilmaz, R. M., & Arici, F. (2019). The impact of virtual reality-based learning environments on students' attitudes and learning outcomes in sustainability education. *Journal of Educational Technology & Society*, 22(4), 67-80.
- Zamawe, F. C. (2015). The use of NVivo software in qualitative data analysis: Advantages and challenges. *Malawi Medical Journal*, 27(1), 13-15. <https://doi.org/10.4314/mmj.v27i1.4>

In- and outside Hololens 2 AR application for visualization of architecture 3D project data

Lukas Koch, Christian Carlhoff, Gerhard Wittfeld, and Ingrid Scholl

*Mascor Institute, University of Applied Sciences FH Aachen, Germany
Kadawittfeldarchitektur Aachen, Germany*

Corresponding author: Ingrid Scholl

Keywords: Augmented Reality, Hololens 2, architectural visualization

Abstract

This paper introduces an augmented reality (AR) application for the Microsoft Hololens 2, designed to visualize the new architectural design for the Eupener Straße campus of FH Aachen. The app combines a digital twin with a 1:1 AR representation of the new buildings, allowing for interactive exploration of 3D models, lighting simulation, and mobility concepts. Unlike traditional Virtual Reality (VR) solutions, which require large datasets and complex setups, this AR application enables effective visualization with minimal data, making it particularly useful in early project phases.

Introduction

Architectural designs not only shape the site but also impact the surrounding community. Traditional methods like haptic 3D models and 2D planning documents often require a high level of spatial imagination, which can lead to misunderstandings and reduced project acceptance. Moreover, transporting fragile models poses logistical challenges (Rankohi & Waugh, 2013).

To address these limitations, we developed an augmented reality (AR) application for the Microsoft Hololens 2 that offers an immersive, intuitive way to visualize the new building concept for the Eupener Straße campus of our University of Applied Sciences FH Aachen. The winning design, "Landschaftsterassen. submitted by kadawittfeldarchitektur, was selected from a competition.

Our AR solution combines a digital twin of the campus with a 1:1 representation of the new buildings on-site, creating an interactive and informative presentation (Böhm et al., 2023). This approach aims to

facilitate communication among residents and project stakeholders, enhancing understanding and support for the project.¹⁹

Similar applications for architectural visualization, either for mobile phones²⁰ or AR devices²¹, commonly use real world markers (e.g. QR-Codes) as a base for positioning the 3D Models in the real world. Marker-based solutions, while common, are limited by the need for fixed starting points and can be cumbersome to implement on a larger scale.

Our solution deploys a markerless approach, by incorporating the Hololens 2's internal spatial mapping, as well as receiving external GPS-Data. This allows the user a dynamic indoor positioning of the digital twin, as well as experiencing the AR enhanced campus plan on site or from any onsite location.

Application

The AR application developed for the Microsoft Hololens 2 was built with Unity using the MRTK and incorporates a 3D Model of the campus grounds, based on site, floor and section plans.

In order to fully exploit the advantages of augmented reality design visualization, a hybrid approach was chosen, offering both a digital twin of the entire campus and a 1:1 AR representation of the new buildings on-site.²² Due to the different requirements within the Unity project, this hybrid approach is divided into two different scenes within the Unity project: Table model and GPS model.

This section outlines the core structure, functionality and use cases of the different scenes within the application.

Main Menu

Upon starting the app, the user is placed in the Main Menu (Figure P 49).

Overview of functionalities: It corresponds to a hub for the user where information about the contents of the two main scenes is provided, as well as a tutorial area, giving the user the opportunity to get acquainted with the interaction with virtual objects.

Scene selection: Through the use of buttons, the user can choose between the two visualization methods: the Table model and the GPS model.

¹⁹ <https://medium.com/@dani.schenker/augmented-reality-in-der-architektur-so-sieht-die-zukunft-aus-c4769da0b116>

²⁰ <https://www.darfdesign.com/arki.html>

²¹ <https://www.lakeflato.com/content/mixed-reality-lakeflato-%E2%80%93-full-scale-holograms-microsoft-hololens-2>

²² <https://www.autodesk.com/de/design-make/articles/was-ist-augmented-reality>



Figure P 48. Main Menu including scene selection and tutorial area.

Table model

The Table model presents the user with a scaled 3D model of the campus and its immediate surrounding area (Figure P 50 Left). This scene focuses on exploring and manipulating the 3D model through hand tracking (move, rotate, scale), as well as four integrated functions:

Lighting Simulation: A lighting function allows the user to rotate a sun object around the 3D model in order to visualize the changes in lighting caused by construction of new buildings.

Scene element toggle: This feature enables users to turn different elements of the 3D model on or off, such as construction phases, landscaping, roads or trees. This allows the user to highlight specific aspects of the design or follow the construction process over time.

Building selection: The user can choose certain buildings from the 3D model to manipulate individually and examine in more detail. When a building is selected, additional information like floor plans, usage concepts and technical details are provided in a pop-up window (Figure P 50 Right).

Mobility concept: The construction plan also includes changes to the way people enter and traverse the campus area via different transport options. Cars, buses and bicycles are enabled as animated 3D models to visualize different means of arrival and to provide an easy-to-understand overview of the campus accessibility. Further information regarding the mobility concept can be accessed via a pop-up window.

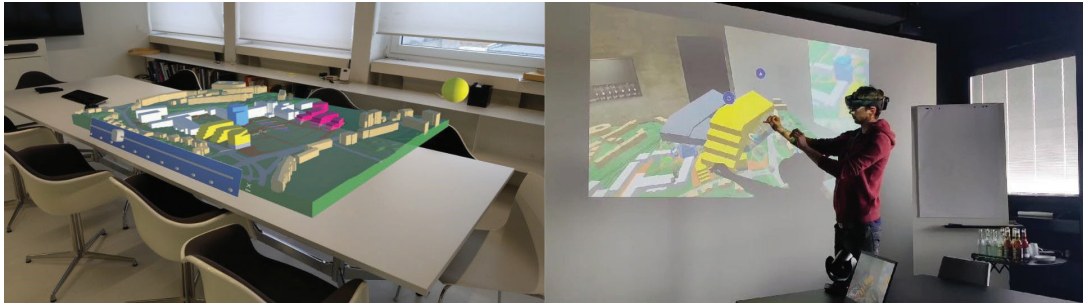


Figure P 49. Left: Table model with lighting adjustment, Right: Hololens 2 screen sharing, building selection.

GPS model

This scene is designed to offer a 1:1 scale, immersive visualization of the new structures within their actual surroundings (Figure P 51). The main features include:

On site positioning: The user can choose between two types of positioning in order to line up the 3D models with the real-world environment: GPS (smartphone or internet connection) and a manual marker placement. Due to the large scale of the campus area, markerless placement options were chosen to not bind the user to a starting location.

Augmented site walk: The on-site visualization gives users, such as university officials, architects, or even potential investors, the opportunity to experience the project as if it were already built.

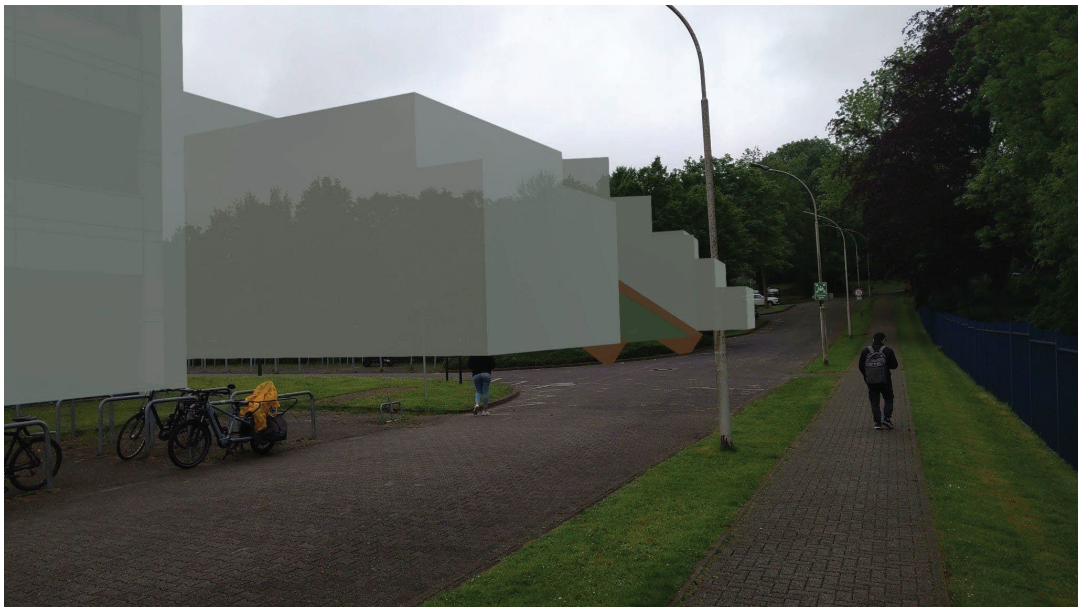


Figure P 50. GPS model spectating new Buildings on site.

Use Cases

The application is designed for a range of use cases, offering value to different stakeholders:

Architectural presentation: Architects can use the application to present the new campus design to clients, university administrators, or the public, providing a more engaging and realistic visualization compared to traditional models or drawings.

Public engagement: The 1:1 AR visualization on-site allows local residents to better understand how the new buildings will fit into their community, addressing concerns about the project's impact on the local environment.

Project collaboration: The interactive nature of the application fosters better communication between project participants, from planners and engineers to decision-makers. It provides a platform for discussing design changes, resolving issues, and aligning on project goals.

Evaluation

To evaluate its effectiveness and stability, the application was presented during an open house event at the university campus to students and campus personnel, as well as an office presentation at the architectural firm kadawittfeldarchitektur. A group of 30 Participants were introduced to augmented reality, guided through the controls and rules of hand tracking as well as the functions of the application and asked to fill out a survey after the test. The questionnaire evaluated users' opinions on the implementation of the actual application, the intuitiveness of the controls and the general effectiveness of AR in communicating architectural concepts. A total of 11 participants engaged in the survey.

One of the key findings from the written and vocal feedback was that participants initially needed time to adjust to manipulating holograms using hand gestures. Augmented Reality hardware like the Microsoft Hololens 2 is a completely new concept for a lot of people and can be overwhelming initially. To facilitate this learning curve, streaming the AR view to a monitor allows external observers to provide real time assistance and guidance in navigating through the application.

Despite the initial adaptation period, the overall reception of the application was overwhelmingly positive. The immersive nature of the AR experience allowed participants to engage with the project in ways that traditional models and drawings could not achieve. In particular, the ability to visualize the design at both a scaled level and in its real-world context was seen as a valuable tool for conveying complex architectural ideas.

Conclusion

The AR application developed for the Hololens 2 has proven to be an effective and innovative tool for visualizing complex architectural designs. Through the combination of a digital twin and on-site 1:1 AR representations, the application successfully bridges the gap between abstract planning documents and real-world understanding. The immersive and interactive nature of the application allowed participants, ranging from students to professional architects, to gain deeper insights into the new building concept for the Eupener Straße campus.

The added value of AR lies in the possibility to visualize architectural design at a very early stage and with minimal data required. In conclusion, the application enhances architectural presentations and fosters better communication, making complex designs more accessible and engaging for a wide range of stakeholders.

Due to the fact that the Hololens 2 was developed for usage in a closed environment, using the application under direct sunlight can be overwhelming for the devices display capabilities.

Future works on this application will focus on testing a different AR capable device like the Oculus Quest 3 for improved outside capabilities and implementing a multiplayer feature to allow multiple users to engage with the architectural design in a shared virtual environment.

References

Rankohi, S., & Waugh, L. (2013). Review and analysis of augmented reality literature for construction industry. *Vis. in Eng.* 1, 9. <https://doi.org/10.1186/2213-7459-1-9>

Böhm, F., Dietz, M., Preindl, T., & Pernul, G. (2021). Augmented Reality and the Digital Twin: State-of-the-Art and Perspectives for Cybersecurity. *J. Cybersecur. Priv.* 1, 519-538. <https://doi.org/10.3390/jcp1030026>

Appendix

A demonstration of the application can be found on Youtube under the following [Link](#)²³. The video contains screen recordings of the Hololens 2 showcasing the GPS and table model from a user perspective.

²³ (<https://www.youtube.com/watch?v=YTs9gSy4y20>)

Demos

"VoirVR": A Comfortable, (A-) Synchronous Collaborative VR Tool for Product Engineering

Philipp Stiens¹, Samuel Korb¹, Annegret Melzer¹, Franziska Klimant¹,
and Martin Dix^{1;2}

¹*Chemnitz University of Technology, Institute for Machine Tools and Production Processes, Germany*

²*Fraunhofer Institute for Machine Tools and Forming Technology IWU, Chemnitz, Germany*

Corresponding author: philipp.stiens@mb.tu-chemnitz.de

Keywords: Human-Centered Computing, Human Computer Interaction (HCI), Interaction Paradigms, Virtual Reality, Collaborative Interaction

Abstract

This paper describes the functionality – environment, avatar, navigation, and interaction - of a collaborative VR application for product engineering whose objective is to reduce the negative effects of using VR, such as cognitive load, through UX and environment design.

Introduction

The ability to work in distributed teams, resource savings and the shortening of development processes are drivers for the use of virtual reality (VR) in product development (Chow et al., 2019). There are already VR applications that focus only on individual aspects of the design review or on the creative design of virtual worlds ([CAD2VR](#), [Softspace](#), [Tilt Brush](#)). However, the use of VR can be associated with negative effects such as cybersickness and cognitive overload (Kazemi et al., 2023). Thereby, cognitive load describes the relative demand of a task in relation to the required mental resources (Souchet et al., 2023). If the load is too high or too low, this can lead to errors, longer processing times and lower user experience (UX). Based on the Cognitive Load Theory (Paas, 2010), the extraneous cognitive load, i.e. the choice and design of the teaching material and the additional use of multimedia objects, is of particular interest to us. With the help of usability principles (Hassenzahl, 2006) and user-centered design (ISO 9241-220:2019), it is possible to avoid negative experiences and problems during use. The challenge of establishing VR technologies in the work process therefore lies in the design process and the design of the UX. Thereby, a special feature in the creation of VR applications is the design of the VR environment as a virtual environment and, above all, its influence on the UX.

Development of the application

The demonstrator presented here is the result of the third iteration of a user-centered design process with special focus on a comfortable and load-reduced UX. The application supports the collaborative configuration of a product, using the example of a robot (see Figure D. 1). Thereby, the application can be used by up to eight users at the same time. The design was based in particular on the UX and cognitive load requirements and was evaluated and adapted several times over the course of the two-year development period.



Figure D. 1. Top left: Basic structure of the environment with Creation Space and Launch Room; top right: Configuration table with robot model on top; bottom left: Customized avatars during the configuration of the robot; bottom right: Inventory of the user with th

Software Development

The technical implementation utilized Unity version 2021.3.1f1. Therefore, the application employs the Normcore framework for network integration, which provides tailored solutions for mobile and VR applications and facilitates integrated voice chat. Moreover, it exhibits the scalability to accommodate a large client base. Performance enhancement strategies included simplifying geometries, utilizing the glTF file format with texture compression, and optimizing the use of textures and materials.

The development focus of the VR application was on current VR hardware, such as the XR head-mounted display Meta Quest 3. Beyond existing hardware compatibility, the software provides developers the flexibility to port the application to various platforms and customize it for future hardware requirements through open interfaces and adaptive architecture. Additionally, the software is engineered to support the Physically-Based Rendering model, which enables lifelike visual representations that mirror surface properties accurately. Consequently, a consistent frame rate exceeding 60 fps was achieved for the Meta Quest 3 HMD.

Features

Environment

The environment consists of a Launch Room for arriving in the virtual world and a Creation Space where users can interact and configure the product together. The Creation Space functions as the primary hub for collaborative efforts and engagement with products. Users first enter the application via the Launch Room, which acts as a secure environment for customizing avatars and interactive tutorials. The central element in this room is an interactive mirror. The overall design combines skeuomorphism and contemporary design to focus on an organic, accessible and natural aesthetic that is intuitive and emotionally appealing. In order to provide a harmonic and positive environment, the Launch Room and Creation Space offer a futuristic symbiosis of natural elements, organic shapes and clear structures.

Configuration

The main element in the Creation Space is an interactive table with a model of the product – a robot by default – on it to create a central basis for discussion and to examine and configure the product. For this purpose, interactive stations are located around the table. Users can select different parts, like color, wheels, engine, etc., to inspect and compare basic information. The selected objects can be activated, and their appearance and function can be applied directly to the miniature product on the table. When the user teleports onto the table, the avatar shrinks for a 1:1 miniature view.

Avatars

The application combines reduced avatars from [Ready Player Me](#), consisting of head, neck and upper body, with the controllers of the VR system and an abstract arm representation. The hand gestures are based on the direct representation of the VR controllers to ensure intuitive operation and mutual demonstration, as well as assistance for the users. Abstract, colored lines are used to visually link controllers and avatars. These move synchronously with the hand movements of the VR controllers and thus ensure a clear and comprehensible representation of the gestures.

Locomotion

Using the joystick of the controller, the user can navigate through the VR environment. They aim at the desired position by tilting the joystick forward and holding it. Virtual feet visualize the target position. Additionally, they can use the controller's stick to determine the landing orientation. To trigger the movement, the user must release the joystick. Furthermore, there is the option of using a snap-turn mechanic to change the viewing direction while standing still.

Inventory and Tools

Each user has a personal inventory of tools, such as a camera, microphone, pen and notepad, with which they can interact in a natural and intuitive way. These tools make it possible to create photos and audio recordings, draw 3D sketches and markers, and leave text comments. For entering text, a virtual keyboard with virtual mallets is available to the user. In addition, a cutting-edge feature has been integrated in the form of a head-space menu, which becomes visible upon upward gaze, providing access to meta-information and functionalities akin to those found in a video conferencing interface.

User Feedback

As mentioned earlier, the application was evaluated several times by dozens of users. It showed that the environment was the most exciting and, significantly, most pleasant environment compared to all its iteration levels. In addition, it had the strongest effect of encouraging collaboration while being accessible and having the lowest extraneous load. Finally, with regard to usability, the application was rated overwhelmingly positive, especially in terms of hedonic quality, i.e., aspects of stimulation and novelty.

Conclusion

In this paper, a [prototype](#) for an interactive, collaborative visualization tool is presented. It allows users to engage live as well as asynchronously to exchange ideas, discuss and go through an entire product development process. With this, the prototype lays the foundation for the design of future platforms.

Acknowledgements

This work was funded by the Central Innovation Programme for small and medium-sized enterprises (SMEs) of the Federal Ministry for Economic Affairs and Climate Action. Also, many thanks to our project partners – A.MUSE Vision and VRENDEX - for the successful cooperation.

References

- Chow, K., Coyiuto, C., Nguyen, C., & Yoon, D. (2019). Challenges and design considerations for multimodal asynchronous collaboration in VR. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1-24. <https://doi.org/10.1145/3359142>
- ISO 9241-220:2019 (2019). *Ergonomics of human-system interaction - Part 220: Processes for enabling, executing and assessing human-centred design within organizations*
- Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, 25(2), 91-97. <https://doi.org/10.1080/01449290500330331>
- Kazemi, R. & Lee, S. (2023). Human Factors/Ergonomics (HFE) Evaluation in the Virtual Reality Environment: A Systematic Review. *International Journal of Human-computer Interaction*, 1-17. <https://doi.org/10.1080/10447318.2023.2227835>
- Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational psychology review*, 22, 115-121. <https://doi.org/10.1007/s10648-010-9133-8>
- Souchet, A. D., Lourdeaux, D., Pagani, A., & Rebenitsch, L. (2023). A narrative review of immersive virtual reality's ergonomics and risks at the workplace: cybersickness, visual fatigue, muscular fatigue, acute stress, and mental overload. *Virtual Reality*, 27(1), 19-50. <https://doi.org/10.1007/s10055-022-00672-0>

Pneumatics-Case: Enhancing Learning through Augmented Reality and Digital Twin Technology

Rostyslav Boychuk, Ivan Symotiuk, Daniil Rõbnikov, Vladimir Kuts, Kashif Mahmood, Simone Luca Pizzagalli, and Tauno Otto

Tallinn University of Technology, Department of Mechanical and Industrial Engineering, Tallinn, Estonia

Corresponding author: rostyslav.boychuk@taltech.ee

Keywords: Digital Twin, Augmented Reality, XR for Education, Learning Activities

Introduction

The advancement of technology in education has opened new avenues for interactive learning. Augmented Reality (AR) has become a prominent tool in enhancing the learning experience, offering immersive and intuitive ways to understand complex concepts (Kuts et al., 2018). The Pneumatics-Case system is a headset-based AR application, designed specifically for the educational sector to facilitate the teaching and learning of pneumatic circuit behavior.

Traditional approaches to teaching pneumatic systems frequently rely on static diagrams and theoretical explanations (Shevtshenko et al., 2017), which are often abstract and difficult for beginners to grasp. Furthermore, physical pneumatic components can be unreliable or worn out, posing additional challenges to practical learning. Pneumatics-Case seeks to overcome these limitations by offering a cloud-based digital twin that facilitates real-time simulation and monitoring of pneumatic circuits. This paper outlines the design and methodology of the Pneumatics-Case system, aimed at enhancing both the learning curve and practical understanding of pneumatic systems.

Learning Workflow & Interaction Mechanics

Designed with a user-friendly interface (Figure D. 2), the Pneumatics-Case system accommodates both students and educators. Teachers can set up lessons and experiments that students can explore in real time. Key features of the system are the cloud-based digital twin that ensures for any modifications made to the simulated pneumatic circuit are immediately reflected in the AR environment. The system synchronizes the state of each component on the server, enabling collaborative learning among multiple users.

The tubes within the AR environment are designed to snap into place when components are connected, enhancing the user experience by simplifying the construction of pneumatic circuits. This feature reduces the learning curve for new users and allows for a more intuitive assembly process.

The workflow is designed to provide students with exercises, which can include pneumatic circuit diagrams, circuit description or desired outcome. The task for the student is to configure respective mechanism with the given components. Upon finalizing the circuit, the student should see changes to the circuit, like movement of the spring retractable cylinder upon push of a button valve. The system is enhanced with symbolic representations for each component individually, which makes it easier for beginners to comprehend the logic behind the circuit.

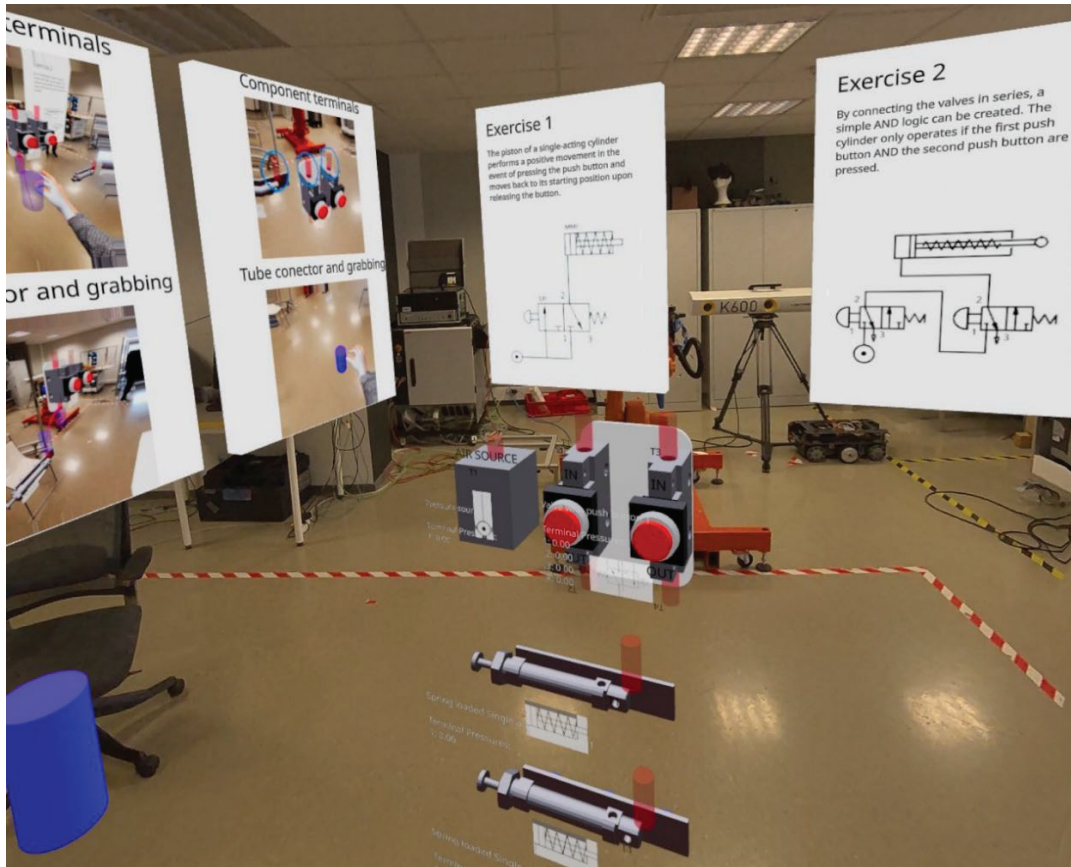


Figure D. 2. User interface of the Pneumatics-Case system showcasing the AR environment.

System Architecture

The Pneumatics-Case system is built as a web application using the Next.js²⁴ framework with TypeScript (Next.js, n.d.). It utilizes the react-three/fiber library²⁵ (Poimandres, n.d.), a React renderer for Three.js^{26, 27} (React, n.d.; Three.js, n.d.), which facilitates the rendering of 3D graphics and provides access

²⁴ <https://nextjs.org/docs>

²⁵ <https://r3f.docs.pmnd.rs/>

²⁶ <https://threejs.org/docs/index.html#manual/en/introduction/Installation>

²⁷ <https://react.dev/>

to the device's native WebXR Device API²⁸ (Mozilla Developer Network, n.d.). This technological stack was selected for its cross-platform capabilities, allowing the application to run on any headset or mobile device that supports WebXR.

The system maintains a list of predefined pneumatic components, each characterized by a JSON-based description, a corresponding behavior function, and a Graphics Library Transmission Format (GLTF) 3D model. The initial set of components includes:

- Pressure Source
- T-Shaped Splitter
- Spring-Retractable Single-Acting Cylinder
- Button Valve

This modular architecture allows for easy extension of the component library by adding custom descriptions and models, facilitating scalability to various educational needs.

Methodology

Simulator Design

The simulator within the Pneumatics-Case application is inspired by the Simulation Program with Integrated Circuit Emphasis (SPICE) electrical simulator (Nagel, 1975) and adapted for pneumatic systems, focusing on pressures and air masses instead of voltages and charges. On each iteration, the simulator loops over each component and invokes its behavior function. This function takes the pressures at the component's terminals and the current state of the component, then returns the mass of air to be added or subtracted from the mass residual of the connected tube and updates the component's state.

Mathematically, the behavior function is defined as:

$$beh \left(\begin{pmatrix} P_1 \\ P_2 \\ \dots \\ P_n \end{pmatrix}, S_t \right) = \left(\begin{pmatrix} \Delta m_1 \\ \Delta m_2 \\ \dots \\ \Delta m_n \end{pmatrix}, S_{t+1} \right)$$

Where P_i represents the pressure at terminal i , Δm_i represents the change in mass at terminal i , S_t is the current state of the component, and S_{t+1} is the updated state.

Pressure Calculation

To calculate the pressure within the tubes, the simulator employs the ideal gas law, utilizing the relationship $p_{g,i} = c * m_i$, where $p_{g,i}$ is the gauge pressure at point i , and m_i is the mass residual. By iteratively performing these calculations at a high rate, the simulator achieves and maintains dynamic pressure equilibrium with a specified level of precision.

²⁸ https://developer.mozilla.org/en-US/docs/Web/API/WebXR_Device_API

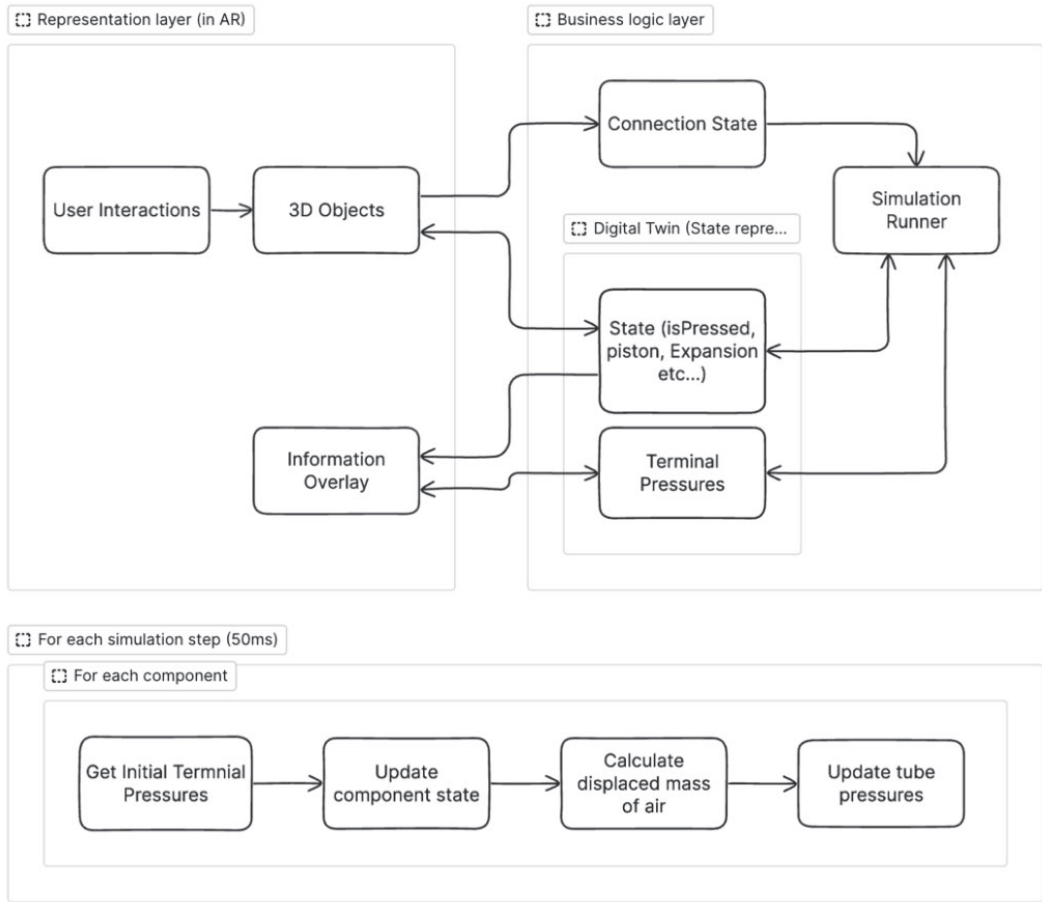


Figure D. 3. Block diagram of system architecture for the Pneumatics-Case.

Conclusion & Future works

The Pneumatics-Case system represents a significant advancement in the pedagogical tools available for teaching pneumatic systems. By integrating Augmented Reality with a cloud-based digital twin, the system offers an interactive and immersive learning environment that enhances students' understanding of pneumatic circuits. Preliminary evaluations indicate that students using the Pneumatics-Case system exhibit improved comprehension of pneumatic concepts and enhanced problem-solving abilities.

This work underscores the potential of AR technology in educational settings, particularly within mechanical and industrial engineering disciplines. Future research will focus on expanding the system's capabilities to simulate more complex mechanical systems and exploring its application across other technical fields.

Acknowledgments

The current work is supported by Erasmus+ Programme under the action of cooperation partnerships in school education (KA220-HED), VERT23070 (XREN).

References

- Kuts, V., Otto, T., Caldarola, E. G., Modoni, G. E., & Sacco, M. (2018). Enabling the teaching factory leveraging a virtual reality system based on the Digital Twin. Proc. 15th Ann. EuroVR Conf.
- Mozilla Developer Network. (n.d.). *WebXR Device API*. Retrieved from https://developer.mozilla.org/en-US/docs/Web/API/WebXR_Device_API
- Nagel, L.W. (1975). SPICE2: A Computer Program to Simulate Semiconductor Circuits, Memorandum No. ERL-M520, University of California, Berkeley
- Next.js. (n.d.). 2024. Next.js Documentation. Retrieved Sep. 13 from <https://nextjs.org/docs>
- Poimandres. (n.d.). 2024. Introduction - React Three Fiber. Retrieved Sep. 13 from
- React. (n.d.). 2024. The React reference documentation. Retrieved Sep. 13 from <https://react.dev/>
- Shevtshenko, E., Karaulova, T., Igavens, M., Strods, G., Tandzegolskiene, I., Tutlys, V., Seyed, T., & Kuts, V. (2017). Dissemination of engineering education at schools and its adjustment to the needs of enterprises. Mat. Sci. 18(1).
- Three.js. (n.d.). (2024). Three.js Documentation. Retrieved Sep. 13 from <https://threejs.org/docs/index.html#manual/en/introduction/Installation>

Extended Reality simulator to evaluate remote control interactions for reach stackers

Andrea Alesani, Kaj Helin, Vladimir Goriachev, and Timo Kuula

VTT, Tampere, Finland

Corresponding author: andrea.alesani@vtt.fi

Keywords: Situational Awareness, Augmented Reality, Mixed Reality, Remote Control

Introduction

Remote control of off-highway machinery systems can bring many benefits to companies and operators. Managers don't need to handle personnel relocations, and operations time is optimized as an operator doesn't have to move from one machine to another, as they can seamlessly switch to control another machine from the same workstation. Operators could also benefit from better ergonomics, less noisy environments, and the risk for injuries would also be significantly reduced. Remote control would also allow to delegate some easier tasks, such as driving towards a specified location, to an automated system, and operators can intervene only for more delicate processes.

That said, remote control does not come without some caveats. An operator's situational awareness risks being considerably reduced, especially when they can only rely on traditional camera feeds with a limited field of view and that can only capture bidimensional information. It is also hard to perceive other important environmental things such as wind strength and direction and outdoor temperature (to know, for example if the ground might be slippery).

The proposed simulator allows for a user-centric approach (Helin et al., 2024) to the interaction and system design, and it was used to reach an acceptable usability level in terms of features and functionality before starting the implementation of the real remote-control prototype (see Figure D. 4). This abstract describes the final simulator and features, and it outlines the key results that were taken and considered for the actual prototype.

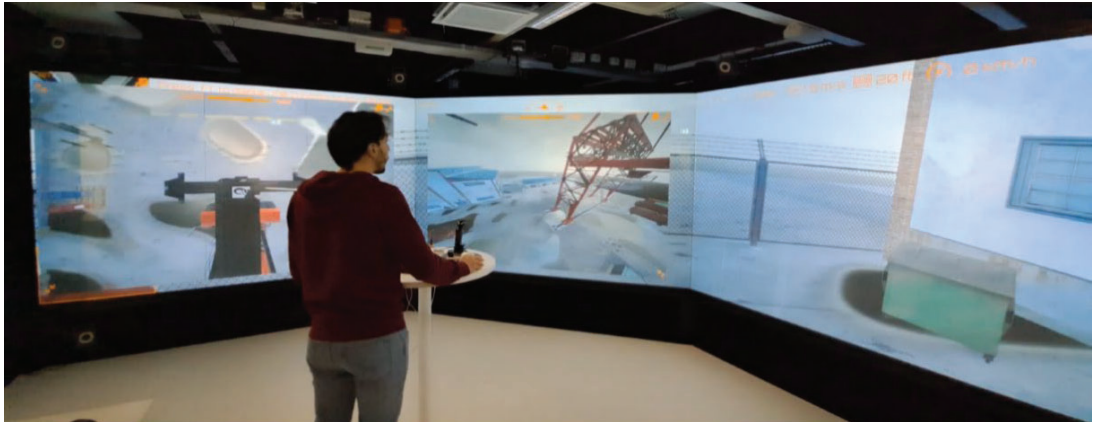


Figure D. 4. The simulator application running on VTT's Powerwall.

Remote reach stacker simulator for User-Centric Design

To simulate a realistic scenario and obtain usability results from the users, they were asked to use the simulator to (i) pick up a container on the harbor's ground by aligning the reach stacker's spreader to it and then locking the twist locks into the container's corner castings; (ii) release the container on a truck's flatbed. They had no time limits, and they didn't have to use all the functionalities but only the ones that were deemed useful (an initial tutorial would still showcase all the possibilities offered by the simulator).

Hardware description

The application was developed to run on VTT's wall-sized Powerwall display, as well as on a traditional FullHD monitor. The UI was adapted for these two cases, but apart from that the application was the same one. Unity 2022.3 LTS was chosen as for the game engine, while a Logitech Extreme 3D joystick and an Ultraleap IR 170 hand-tracking sensors were used as input devices (together with mouse and keyboard). The application was tested on Windows 11 operating system.

Visualization

The application is based on four 360° camera streams. The cameras are placed in key locations that were determined in previous studies (Alesani, 2023), namely one front camera placed in what is now the operator cabin in real reach stackers; a second camera is placed on top of the spreader, to allow for visibility behind a grabbed container and to have a "top view" when trying to align the spreader's twist locks to a container's corner castings; a third camera is placed in the back of the reach stacker, mostly for safety reasons; the fourth camera shows a "drone view" that allows for a third-person perspective on the reach stacker and its surroundings (see Figure D. 5). Camera views are highly customizable: users can dock them to the bottom of the screen, choose which one to have as main camera, and place them around the Powerwall display as "picture-in-picture" views that can be resized and customized with, for example, different fields of view.

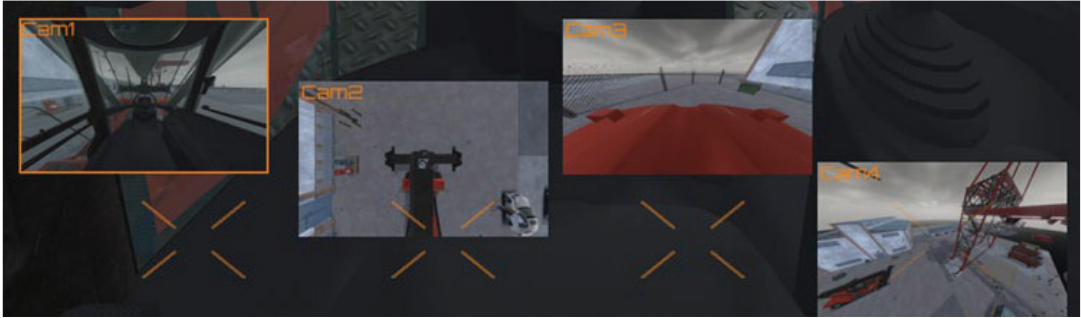


Figure D. 5. Camera system's preview images and their docking positions.

Even with multiple views, it might still be hard to have a precise sense of perspective when trying to lock a container or releasing it on a truck's flatbed. UI indicators and AR hints were developed to enhance the user's situational awareness and performance for this type of task.

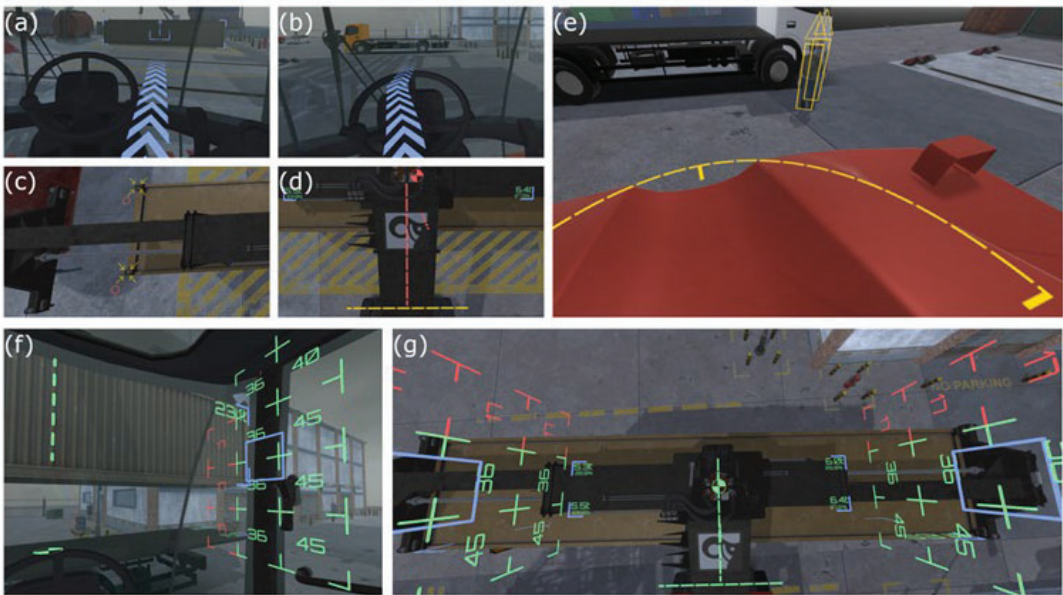


Figure D. 6. AR guidance system elements: (a), (b) - vehicle path and pick up/placement locations; (c), (d) - twist lock/corner casting positions, center line, and tipping axis; (e) - counterweight safety area and obstacle tracking; (f) - load chart, centre of mass

While many of the AR visuals are always available (like the ones seen above), some others are context-dependent and are only visible during a specific part of the task the user is performing, for example when locking the container (see Figure D. 7).



Figure D. 7. When the user has to lock a container, color-coded AR lines appear to indicate the proper alignment between twist locks and corner castings (on the left). Another experimental feature is a real-time LiDAR scan of the container that shows a 3D rotatable

Interactions

The Logitech Extreme 3D joystick was used to simulate a real reach stacker joystick, and the mapping of the controls was kept as close as possible to the one in the real vehicles (see Figure 5). Tilting the stick forward and backward makes the reach stacker's boom rotate respectively downward and upward, while tilting it left and right would makes the telescopic part of the boom respectively retract and extend. Buttons on top of the joystick are used to control the reach stacker's spreader, which can rotate clockwise or counterclockwise parallel to the ground, or shift left and right along the turntable the connects it to the boom (a real spreader has some additional controls that were not considered for the simulator).



Figure D. 8. The user aligns the spreader's twist locks to the container's corner castings by controlling the joystick.

Ultraleap's IR 170 hand-tracking sensor is used to track a few gestures that were made available to the user. Hand gestures only affect visualization aspects of the application, and have no effect in the

simulated real-world, remote-controlled environment. They are only tracked for the left hand, as the right one is supposed to handle the joystick and might produce undesired gestures when controlling the reach stacker. There are two types of gestures. When a hand swipe is recognized, the main view switches camera feed either to the next one (if swiping from right to left) or to the previous one (if swiping from left to right). When the user performs a pinch with their thumb and index finger, the swipe gesture recognition is disabled, and they become able to “pan” the view on the screen around the 360° camera footage by moving the hand up, down, left or right (see Figure D. 9).

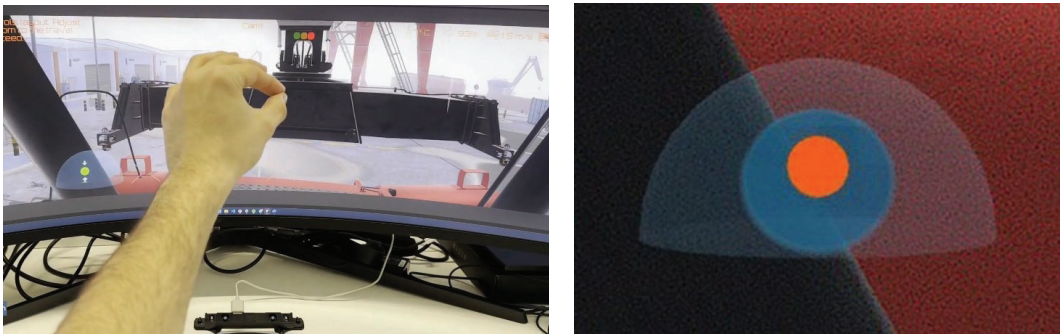


Figure D. 9. The user can “pan” through the 360° view by pinching the fingers and moving the hand up, down, left or right. An AR hint is shown on the bottom of the screen which visualizes the current tracking status (orange for no hands tracked, yellow for hand recognized but out of bounds, green hand available for gestures). Vertical arrows appear when the pinch gesture is active, and horizontal arrows show when the user can activate the swipe gesture.

Acknowledgements



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

- Alesani, A. (2023). XR application for remote operations. Politecnico di Milano. <https://hdl.handle.net/10589/210680>
- 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>

Appendix

Video of the final prototype: <https://youtu.be/gfgrw0SgZeM?si=H-YwZ16ZCszbddXC>

Industrial remote cooperation

Bastian Kraye, Narek Minaskan, Alain Pagani, and Didier Stricker

German Research Center for Artificial Intelligence, Germany

Corresponding author: bastian.kraye@dfki.de

Keywords: Collaborative and Distributed XR, Augmented Reality and Mobile Devices, Industrial Applications

Introduction

In recent years, the market for mixed-reality glasses has steadily been growing, with many VR capable headsets gaining AR functionality as new models are developed. Together with this progress, standardization of common features is being worked on with APIs such as OpenXR and WebXR, making it more feasible to develop applications compatible with a wide variety of devices. In addition, there have been many initiatives in remote work opportunities, connecting people in different places.

Our demonstrator aims to combine these fields in the context of industrial remote cooperation. This cooperation takes the form of an augmented video call. In our scenario, a technician working in an industrial setting, like a factory, might not yet be familiar with a specific machine or needs to do some uncommon maintenance work.

They then call a remote expert familiar with the topic. This expert can communicate like in a normal video call but also sees a virtual view of the machine and the technician's pose. They can now augment the technician's view by placing annotations inside of the virtual scene to supplement any explanations. One goal of our demonstrator is that different devices can communicate with each other, such as desktop PCs and headset devices.

Architecture description

Our demonstrator is developed using Unity²⁹ and makes use of the Rainbow™ platform³⁰ created by the French company Alcatel-Lucent Enterprise. The platform provides common communication features such as video and audio as well as an experimental low-level data channel. The data channel is used to exchange information between clients in a call via a messaging layer. Using these messages, all clients maintain a consistent view of the system's state. This includes data such as the current pose

²⁹ <https://unity.com>

³⁰ <https://developers.openrainbow.com/home>

of the technician in 3D space, roles of the participants and created annotations. This communication layer is independent of the execution platform and allows for cross-platform interoperability.

Desktop PC version

The desktop application is available for any Windows computer and contains two different modes. One mode is for the technician role. This mode will be used on-site, and its main tasks are determining the technician's camera's position and orientation with regards to the machine to be worked on and displaying external 3D data, such as annotations in the camera's view. As most laptop or PC systems are or can be easily equipped with a webcam, we use a model-based tracking by Rambach et al. (2017) to robustly determine the camera pose from just the camera image. The available 3D model is also used to project it into the camera image, making it possible to simulate effects such as occlusion and shadowing when adding virtual objects, thus aiding the user in 3D perception.

The other mode is for the expert role. In this mode, the expert sees a purely virtual representation of the scene in which they can freely move around. The model as well as a representation of the technician's pose is visible to aid the expert in knowing where the technician's attention is. A list of annotations for several common shapes is available. They can be created in different colors and placed in the scene, where they can be moved, rotated and scaled to match what the expert is explaining. The annotations and their movements are synchronized with the technician. Figure D. 10 shows such a shared scene across different clients and roles.

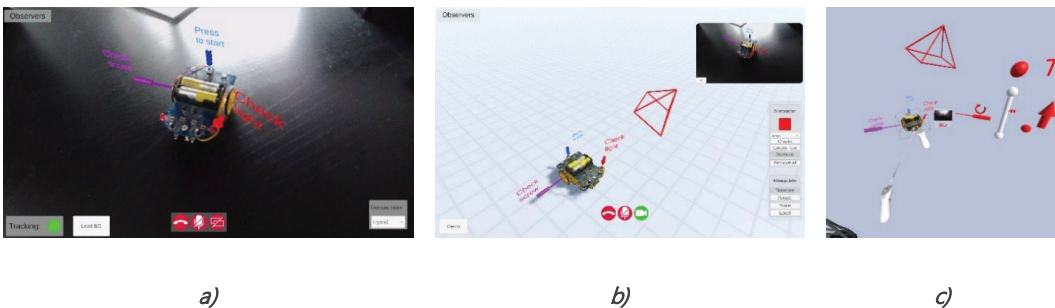


Figure D. 10. Different views of the scene. a) Shows the technician on a desktop. Annotations are displayed in the camera view. The same scene is shown on a desktop expert in b) and a VR headset in c). In both, the technician's camera stream is shown in addition to

XR headset version

The XR headset application utilizes OpenXR³¹ via the Unity OpenXR plugin. This makes it possible to reuse code across various headsets. Interaction with virtual objects is implemented using the XR interaction toolkit, which provides necessary building blocks. Like the desktop version, the headset application has two modes, for the technician and expert. In contrast to the desktop client, which uses a common mouse and keyboard interaction scheme, the XR headset version utilizes controllers and hands, between which the user can dynamically switch. This speeds up positioning the annotations considerably, as these controls directly map to the 3D poses.

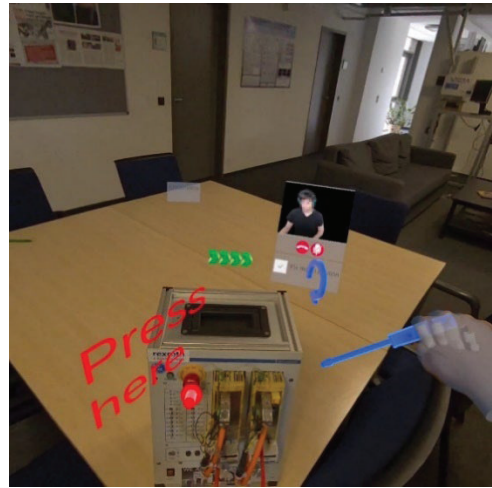
³¹ <https://www.khronos.org/openxr>

While the desktop technician uses a camera, this is generally not possible in XR headsets due to privacy concerns. To cope with this limitation, the technician starts by choosing the model of the machine they are working with and places it on the real object. Placement is done by intuitively moving, rotating and scaling the virtual object either with the user's hands or controllers. Afterwards, the object is made invisible, just providing depth information, such that occlusion can still occur on the camera pass-through. Annotations are always defined relative to the model and will be displayed accordingly, thus anchoring them in the world. Figure D. 11 shows this process.

The expert similarly sees a virtual representation of the machine. They can move around and scale the model as they wish. Additionally, available annotations are placed on an adjustable ring from which they can be grabbed. When this happens, the annotation gets added to the scene and will be visible to the technician.



a)



b)

Figure D. 11. Example of the XR technician view. In a) the technician places the virtual object on top of the real object. In b) the virtual object is hidden, and annotations are placed in the 3D scene by the expert, who is also sharing a webcam video.

References

Rambach J, Pagani A, & Stricker D. (2017). Augmented Things: Enhancing AR Applications leveraging the Internet of Things and Universal 3D Object Tracking. Poster presentation at 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), p. 103–8. <https://doi.org/10/ggr6rz>

The XR4ED Intelligent XR Tutor

Fotis Liarokapis

CYENS – Centre of Excellence, Nicosia, Cyprus

Corresponding author: f.liarokapis@cyens.org.cy

Keywords: Extended Reality, Education, Avatars, Artificial Intelligence

Introduction

The European EdTech sector is pivotal in driving digital transformation across education and training, seeing a 15% increase in investments in 2021. However, the sector remains fragmented with few globally prominent companies. Meanwhile, Europe's XR industry is a leader in software and content production, with XR technologies providing immersive, skills-based learning opportunities. Developing XR applications requires specialized skills, time, and financial resources.

FiorMarkets predicts that between the period 2018-2025 the global VR market is expected to reach USD 56.25 Billion by 2025 at a CAGR of 34.2%³². However, VR technology is still not ready for representing ultra-realistic virtual avatars that perfectly mimic a person's movements and expressions. Getting these avatars into production, though, will require further breakthroughs in graphics, computer vision, machine learning, and audio processing (Gent 2020).

Existing AR systems that make use of video conferencing are fragmented and provide limited capabilities. A typical example is ZugSTAR³³ which is a video conferencing system like Skype, Teams, etc. but with the added functionality of being able to see one another's "augmented" experience in real-time. The ideal AR conferencing system should offer features like 2D/3D recording, voice calls, collaboration tools, and the ability to modify virtual objects and information³⁴.

XR4ED (Liarokapis et al, 2024) is an Innovation Action EU project aiming at the design, deployment, piloting, and market entry for a novel product along with its value-added services. In this direction, the overall methodology of the XR4ED project implementation is targeted towards achieving the commercialisation of the XR4ED platform. The XR4ED project aims to unify the EdTech and XR communities, overcome fragmentation, and accelerate innovation in personalized, inclusive learning

³² <https://www.fiormarkets.com/report/vr-market-by-technology-nonimmersive-technology-semi-immersive-fully-386072.html>

³³ <http://zugara.com/augmented-reality-and-virtual-reality-technology/augmented-reality-video-conferencing>

³⁴ <https://www.devteam.space/blog/how-to-build-an-augmented-reality-conferencing-app/>

through XR. Key initiatives include creating a central marketplace for XR education applications, facilitating knowledge sharing, and enabling startups and SMEs to develop market-ready XR solutions.

XR4ED Components

XR4ED is based on open standards (OpenXR/WebXR), providing a single access point for XR content, tools, and solutions. It promotes a sustainable ecosystem, supporting SMEs, start-ups, and developers, while ensuring alignment with European values of privacy, ethics, and inclusiveness. XR4ED involves several technical components aimed at enhancing XR for educational purposes:

- **XR DevOps Platform:** A web-based DevOps platform integrates third-party tools via REST APIs, with PHP backend and cloud services for scalability. Users can log in with GitHub for seamless onboarding and collaboration. The platform will offer pedagogical add-ons and DevOps services, integrating various third-party tools.
- **XR Avatars:** The project focuses on developing avatars that improve interaction and non-verbal communication in collaborative VR education settings. Different avatar styles (cartoonish or fully animated) will be developed depending on the lesson's goals, with attention to scalability, accessibility, and gender representation.
- **Marketplace:** A web application where users can browse XR4ED services, preview public offerings, and test APIs after authentication. The Marketplace will evolve with open calls for new components, and users can filter results based on categories like content triggers, input/output devices, and third-party plugin requirements. Different user groups (developers, educators, students) will have tailored access to ensure privacy and appropriate interaction.
- **XR Authoring Tool:** This tool uses XR to enhance real-world teaching environments by superimposing digital information that can be manipulated through tangible interfaces. It supports collaborative learning and allows students to experiment with virtual teaching material in real-time, offering an intuitive learning experience. The focus will be on resolving perceptual discontinuities in learning and ensuring that the system is usable across diverse educational settings.
- **XR UI/UX components design:** Different devices (Meta, Apple, Vuzix) have varying screen and input methods (gestures, head movements, system triggers). UI elements must be tailored for each device while ensuring intuitive interaction. Designers need to account for diverse user groups and maintain consistency across platforms. XR4ED aims to create accessible, user-friendly XR interfaces suitable for all ages and backgrounds.

These components aim to create a comprehensive, user-friendly ecosystem for XR-driven education, ensuring scalability, collaboration, and privacy.

XR Avatar Methodology

The creation of human digital twins for VR has many practical applications today and consists of three interrelated parts: capturing, modeling, and simulation. However, for human digital twins in AR there are several differences. An interesting application is the case of real-time AR teleportation which is currently under development in XR4ED. A simplified model of the process of real-time AR teleportation of a human digital twin into another environment consists of five steps which are listed below:

- **Capturing:** Motion capturing of humans including skeleton representation and avatar reconstruction. The skeleton will be used to determine the body of the avatar. In addition, the skeleton is going to be the basis for the animations that will be performed later.
- **XR Tool:** It is important to register (in terms of position and orientation) the realistic representation of the virtual human into the desired position. Having an effective interaction with the superimposed information is also a crucial part of the user experience. The XR application can be used in portrait and landscape modes, and both consist of showcase and learning modes.
- **Eye Gaze and Head Motion:** A model for generating head and eye movements during gaze shifts of virtual characters, including eyelid and eyebrow motion (Krejsa2018). The goal is for the model to match the communicative accuracy of the real human gaze as closely as possible, and for its generated movements to be subjectively perceived as nearly natural. The proposed model consists of two fully independent components: (a) the gaze controller and (b) the face controller.
- **Lip Synchronization:** Lip-sync animation is also important to increase realism when speech is involved. In this work, we propose to replace each viseme shape with a set of constraints on certain parameters of the face which define a wide range of allowed poses (Krejsa2021).
- **Animation and Rendering:** The final step concerns the realistic rendering of the avatars into the AR environment (i.e., calculating light, shadows, etc.) in such a way that is fully blended in the environment. Multi-layer rendering will be performed allowing for different forms of audio-visual information to be presented to the participants. Real-time animation will require the individual to be adapted to the movements of the host user. However, pre-recorded animation will be also included in this stage to account for intelligent agents.

AR teleportation will allow the teachers to deliver their courses in physical classrooms while being at a remote location. However, there are several issues that need to be solved and one of them is the communication between non-player-characters (NPCs). In the following section, the initial results of an XR tutor that communicates between non-player characters and learners are presented. Furthermore, the methodology will be also extended to VR, offering shared spaces for collaborative and blended learning. This raises challenges in terms of non-verbal communication in shared VR spaces and preliminary results are provided.

Intelligent XR Tutor

While AR technology has proven to be an engaging and effective tool, it lacked intelligent avatar communication, which is essential for creating immersive social XR experiences. To address this, XR4ED leverages intelligent AR avatars to help students explore environments and access information. These avatars, supported by Azure Cognitive Services for speech interaction and ChatGPT for real-time question answering, offer a more interactive and engaging learning experience. The avatars are superimposed into real environments using QR codes or SLAM algorithms and can be customized with gender-specific avatars.

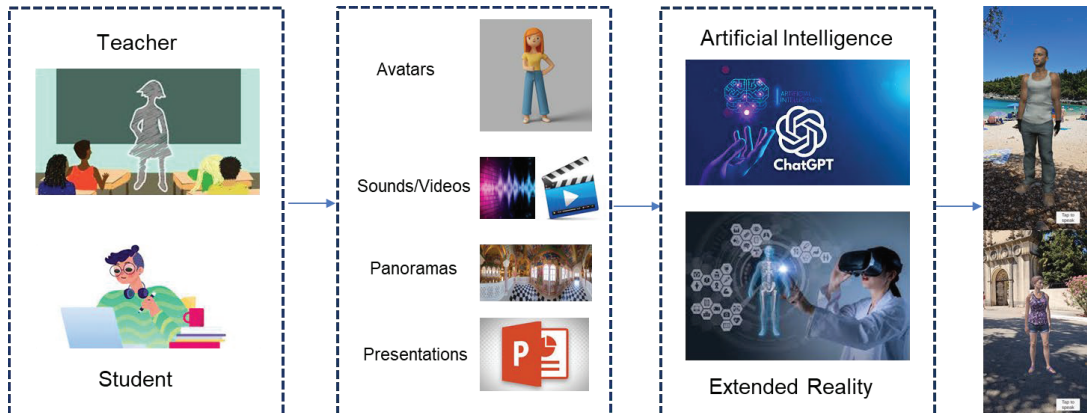


Figure D. 12. Overview of the XR4ED Intelligent XR Tutor

The implementation, built using Vuforia and Unity, integrates gaze and face controllers for realistic head, eye, and lip-syncing movements, enhancing avatar realism. The AR system also supports multimedia formats like videos, panoramic images, presentations, and text, sourced from teachers or the marketplace. Users can search and insert web-based images and metadata into the AR environment using text prompts. Initial tests with Erasmus students demonstrated that the intelligent AR tutor was well-received, with students impressed by its real-time responsiveness and potential to enhance learning outside traditional classrooms.

Acknowledgement

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101093159. It has also been partially supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy under grant agreement No 101061303.

References

- Gent, E. (2020). Forget Video Conferencing—Host Your Next Meeting in VR, IEEE Spectrum. <https://doi.org/10.1109/MSPEC.2020.2988535>
- Liarokapis, F., Milata, V., Ponton, J.L., Pelechano, N., & Zacharatos, H. (2024). XR4ED: An Extended Reality Platform for Education, IEEE Computer Graphics and Applications, 44(4): 79-88. <https://doi.org/10.1109/MCG.2024.3041234>
- Krejsa, J., Kerouš, B., & Liarokapis, F. (2018). A Model for Eye and Head Motion for Virtual Agents, Proc. of the 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games 2018), IEEE Computer Society, Würzburg, Germany, 5-7 September, 1-4. <https://doi.org/10.1109/VS-Games.2018.8493406>
- Krejsa, J., Liarokapis, F. (2021). A Novel Lip Synchronization Approach for Games and Virtual Environments, Proc. of the IEEE Conference on Games (CoG), Copenhagen, Denmark, 17-20 August, 1-9, 2021. <https://doi.org/10.1109/CoG.2021.9619128>

Haptic Directional Awareness in Virtual Reality

Jean Botev, Emanuel Dias da Silva, and Ningyuan Sun

VR/AR Lab, Department of Computer Science, University of Luxembourg, The Netherlands

Corresponding author: jean.botev@uni.lu

Keywords: Virtual Reality, Haptics, Vibration, Navigation.

Introduction

Haptic feedback is an essential component of the human sensory system, constituting an important channel that can provide extensive complementary information without significant cognitive interference with other senses (Prewett et al., 2006). As such, haptic technology holds immense potential for virtual reality (VR), which currently mainly combines audiovisual information with limited tactile or haptic information. Advanced haptic feedback can improve immersion and interaction by giving users a sense of touch or offering meaningful information without further cluttering the visual and auditory sensory channels (Monica and Aleotti, 2023; Wang et al., 2024). To explore how users can be provided additional directional information concerning their environment through real-time sensory augmentation and replacement, we devised a VR study employing a wearable vibrotactile vest with minimal tracking effort. The vest's haptic feedback assists users in solving a set of location tasks with varying directional semantics, guiding users toward specific targets or warning of imminent hazards in a virtual environment that otherwise would be difficult to complete due to a lack of salience or visibility.

Study Design

Hardware. The VR hardware employed in this study included an HTC Vive Pro headset and controllers selected for the precise outside-in motion tracking with a set of base stations. We further opted for the bHaptics X40 TactSuit, a commercially available wearable haptic vest with 40 vibrotactile motors (20 on the front and 20 on the back). Each motor can be precisely controlled, enabling the creation of specific tactile sensations on the wearer's upper body.

Environment. We designed a dedicated environment with different zones, as shown in Figure D. 13. One of the environment's central characteristics is its minimalist appearance, with little to no task-related salience, i.e., visual cues. Participants begin in the first zone (X) at the bottom center, where they go through a brief tutorial to familiarize themselves with the controls, general locomotion approach, and the vest's vibrotactile sensation. Following this tutorial, they are asked to complete a series of challenges involving different tasks in subsequent zones A (Floating Orbs), B (Fog Labyrinth), and C (Trapdoor Floor).

- In the *Floating Orbs* zone (A), participants enter a room full of randomly distributed, identical spheres (cf. Figure D. 14a). Participants have no way of visually identifying the two target spheres they are tasked to find, and –in the worst case– would have to try and return each sphere to the designated placers.
- In the *Fog Labyrinth* zone (B), participants navigate a maze to locate a target object and return it to its placer at the entrance. The environment is filled with dense smoke (cf. Figure D. 14b), severely obstructing visibility and further increasing disorientation.
- The *Trapdoor Floor* zone (C) confronts participants with an empty room where they must reach the opposite side without stepping on collapsing floor elements. If they approach one of these trapdoor elements, participants receive a warning to move in a different direction. If they still enter this area, the floor collapses underneath (cf. Figure D. 14c), with participants falling and respawning back where they entered the zone.

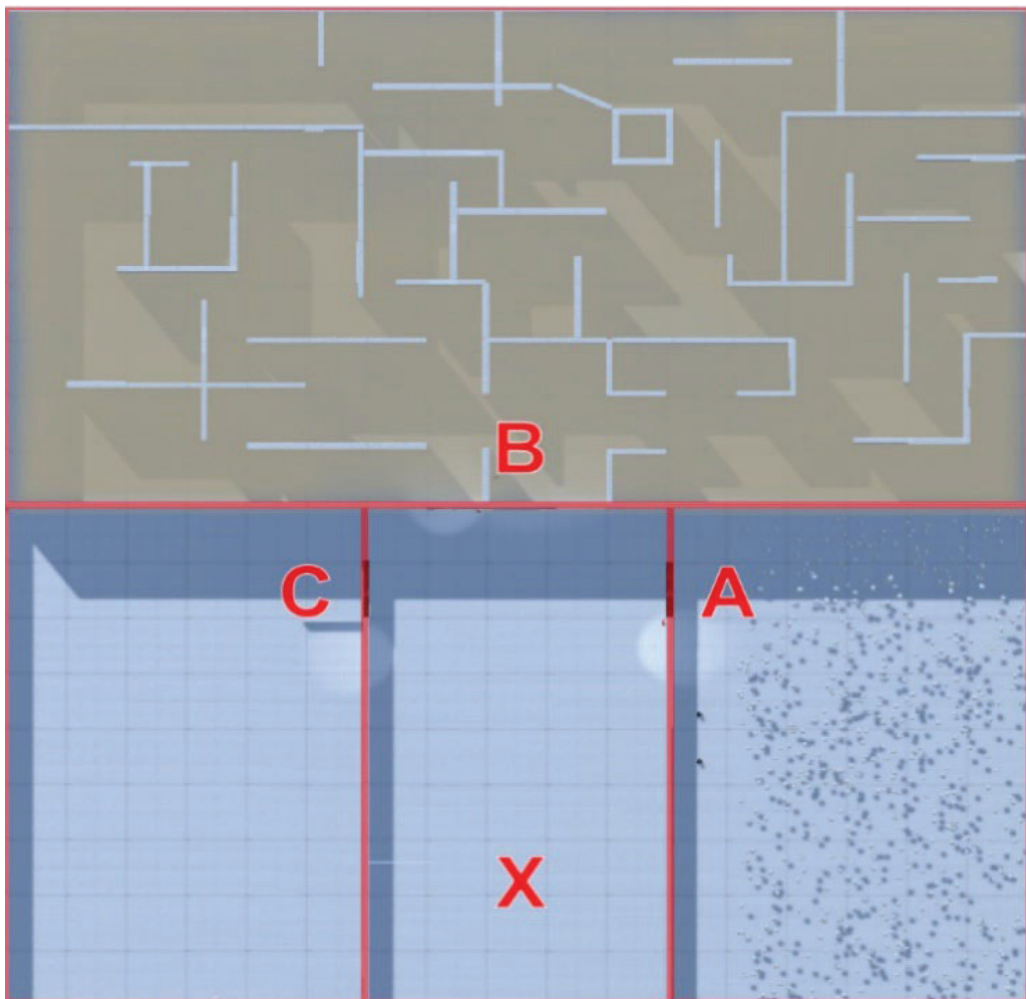


Figure D. 13. Top-down view of the entire virtual environment and different task zones: Start/Tutorial (X), Floating Orbs (A), Fog Labyrinth (B), Trapdoor Floor (C).

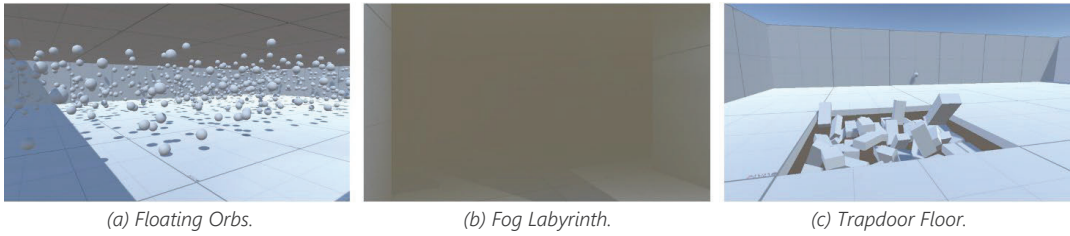


Figure D. 14. First-person perspective screenshots of each challenge.

Haptic output. As illustrated in Figure D. 15, participants receive haptic feedback on the vest about a target location in relation to their position in the virtual environment. The motors to engage are determined by the intersection of the line connecting the participant's face with the target and a 4×5 matrix of spherical phantom objects placed in front of and behind the user in the virtual world, mapping the vest's motors. A small pilot study indicated that gaze deviating from overall body posture has no perceptible effect on the relative direction information, and the haptic information does not need to be transformed to account for this difference. However, we found that body size is a significant factor in resolving directional information from the motor arrays on the front and back of the vest. Therefore, the user's height is factored into the vibrotactile output, as shown in Figure D. 16. The shorter a person, the more punctual the haptic feedback; the taller a person, the more support motors are used to enlarge the feedback area.

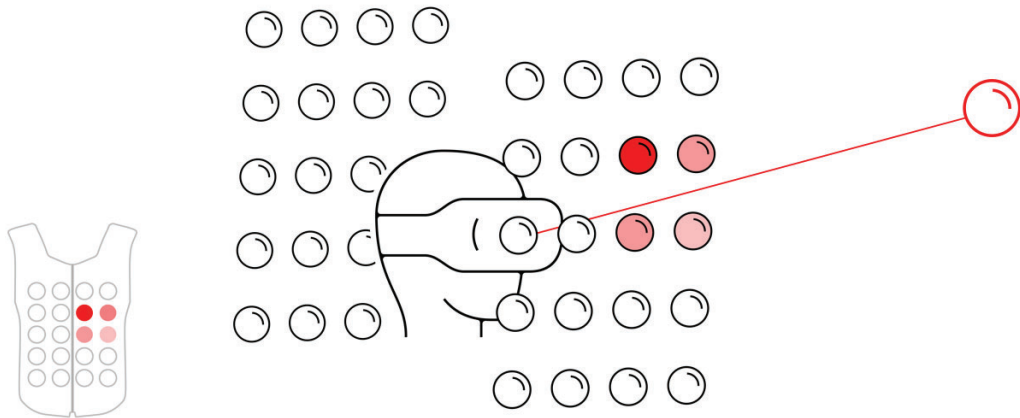


Figure D. 15. Target location mapping on motor array related to virtual position.

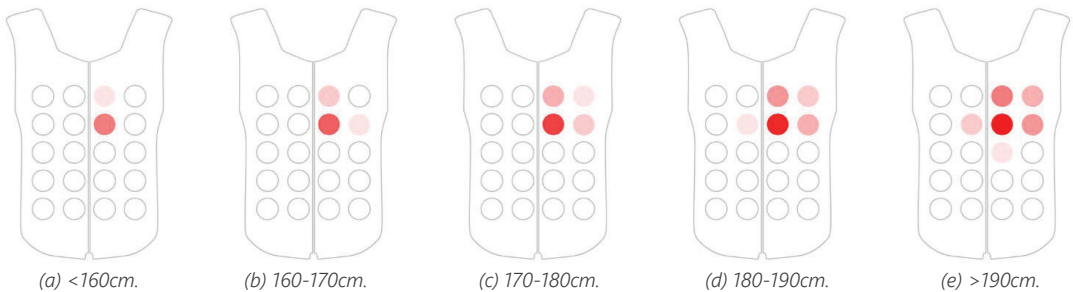


Figure D. 16. Example haptic feedback intensities and ranges relative to user height.

Data collection. Various performance-related data were collected for objective evaluation, such as task completion times or the user's distance to the targets during each motor activation. Further quantitative data stem from the continuous tracking of the users' pose, orientation, and movement. In addition to quantitative data, qualitative data were collected before and after the experiment using pre- and post-trial questionnaires, which mainly collect information related to user experience and haptic feedback, including confidence, usability, immersion and presence. The data collected before the experiment included demographic information as well as previous experience with VR, games, controller use, and haptics.

Results

All participants (2 female, 10 male, mean age = 26.7) successfully completed the different challenges. Figure D. 17 shows the completion times also by proficiency levels regarding VR and haptics. Interestingly, VR experts, on average, exhibit slightly higher completion times than VR novices. This may be because users familiar with VR want to explore the virtual environment more, while novice-level users focus more on the task at hand. However, we see only a slight but insignificant advantage for participants familiar with haptics compared to novice-level participants. There were also only minor differences between participants with different gaming or controller use expertise. This similarity in performance indicates that haptic feedback can effectively supply directional cues to users independent of background or expertise.

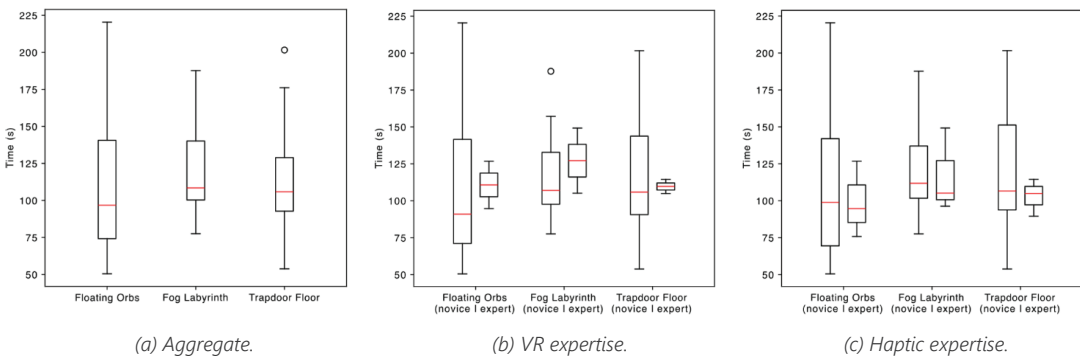


Figure D. 17. Mean completion times per challenge.

The questionnaire revealed that participants were confident in completing the tasks. The VR environment was reported to be immersive, and the navigation controls were easy to master. Most participants also subjectively assessed the haptic feedback as clear, intuitive, and useful.

Summary and Outlook

This work presents a novel approach to haptic directional awareness in VR, demonstrating how it can be achieved with minimal tracking effort using a wearable vibrotactile vest. While preliminary due to the limited group size, the results of this study demonstrate the effectiveness and potential of haptic directional feedback to guide users in difficult or otherwise impossible-to-navigate spaces through sensory augmentation or replacement.

The current study opens up many interesting paths to explore. For instance, while our motor activation algorithm works well, different intensities and activation ranges might be even more effective. Physiological measurements might help further assess the user response and explore the parameter

space comprehensively to fully understand and validate the generalizability of haptic directional awareness approaches beyond VR applications.

References

- Monica, R., & Aleotti, J. (2023) Improving Virtual Reality Navigation Tasks Using a Haptic Vest and Upper Body Tracking. *Displays* 78, 102417. <https://doi.org/10.1016/j.displa.2023.102417>
- Prewett, M.S., Yang, L., Stilson, F.R.B., Gray, A.A., Covert, M.D., Burke, J., Redden, E., & Elliot, L.R. (2006) The Benefits of Multimodal Information: A Meta-Analysis Comparing Visual and Visual-Tactile Feedback. In: *Proceedings of the 8th International Conference on Multimodal Interfaces (ICMI '06)*. pp. 333–338. <https://doi.org/10.1145/1180995.118105>
- Wang, G., Wang, H.H., & Ren, G. (2024) Visual and Haptic Guidance for Enhancing Target Search Performance in Dual-Task Settings. *Applied Sciences* 14(11), 4650. <https://doi.org/10.3390/app14114650>

An Open-Source Authoring Solution for Cultural Heritage Augmented Reality Applications

Mathis Bodin, Jean-Philippe Farrugia, Fabrice Jaillet, and Bruno Tellez

Université Claude Bernard Lyon 1, CNRS, INSA Lyon, LIRIS, Bourg-en-Bresse, France

Corresponding author : Jean-Philippe.Farrugia@univ-lyon1.fr

Keywords: Augmented Reality, Digital Model, Cultural Heritage

Introduction

Augmented Reality (AR) technology has revolutionized the way we interact with our surroundings, offering immersive experiences that blend the physical and digital worlds seamlessly. In the cultural heritage and scientific mediation domain, the ability to create AR applications may significantly enhance visitor engagement with historical sites and artifacts. However, the technical complexities of AR app development often pose a barrier for curators and museum managers who lack coding knowledge. We present an open-source solution that is set to democratize AR application authoring by eliminating the need for coding expertise. Our tool is specifically designed to empower museum managers to effortlessly design and deploy AR experiences. By open-sourcing this solution, we hope to pave the way for widespread diffusion and adoption.

Here are the key features of our framework:

- Compatible with a wide range of Android devices.
- Web-based and therefore easily deployable on any smartphone without installation.
- No need for any coding abilities, users may freely create scenes with 3D meshes, annotations, images, and sounds.
- Open-source: the code is available for anyone to access, modify, and contribute to.

This paper is organized as follows: following this introduction, we will overview related work on augmented reality authoring frameworks in section 2. Section 3 will be dedicated to presenting the design and architecture of our framework. Section 4 will focus on the practical applications and case studies demonstrating the effectiveness of our framework. Finally, section 5 will debrief and conclude with suggestions on potential enhancements.

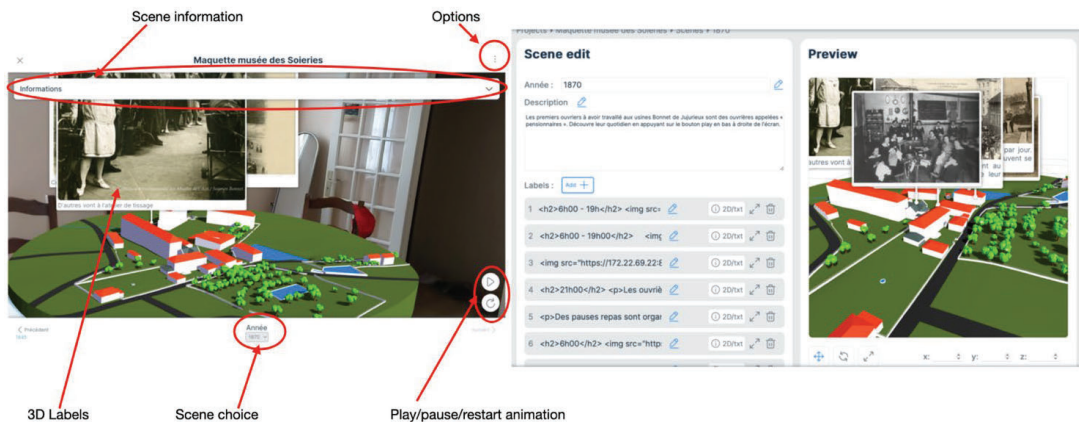


Figure D. 18. Our augmented reality framework: viewer (left) and editor (right).

Related work

Many augmented reality (AR) frameworks currently exist in the market (Adams 2024(Adams, 2024)), each offering different features and capabilities to address various needs and challenges in AR application development. As presented in (Hampshire et al., 2006), AR authoring frameworks mainly divide into two categories: those focused on advanced coding capabilities for experienced developers, and those which aim to simplify the authoring process for users with limited technical knowledge. Our contribution belongs to this second category. An extensive exposition of all augmented reality authoring frameworks would exceed the scope of this paper. The closest of our contribution are ARIS (Holden et al., 2014), authAR (Whitlock et al., 2020), ARComposer (Shekhar et al., 2019) or TrainAR (Blattgerste et al., 2023) but none of them entirely matches the objectives of our solution.

Therefore, to our knowledge, our framework is the only one which combines no-code authoring tools (for user-friendly application creation) with open-source architecture and code (for customizing and enhancing to specific needs).

Framework architecture

As seen on Figure D. 19, the architecture is segmented into three levels - Applications, Services and Data, each of these having different access privileges.

1. The **Data level** holds multimedia content: 3D models, annotations, and any other resources that creators may want to integrate into their AR experiences. It is implemented with a **SQLite** database. For security concerns, this level is private and is only accessible via the API (detailed below).
2. The **Service level** is a restricted level containing services needed to access data and deploy the applications. It may only be accessed by the technical administrator of the system. It contains a *Web Server* and an *access API*:
 - A *Web Server* is needed to deploy the two user applications (detailed below). The server needs to be able to handle **node.js**. We chose **Apache** in our own implementation.

- The *API* acts as a bridge that allows the applications to access, add or remove elements in the database. It has been implemented using **Express.js**.
3. The **Applications level** is public and offers access to the two user applications:
- The *Visualization Web Site* is the interface through which users can experience the AR applications created using the framework. It has been implemented using **Vue.js**, **Three.js** and **WebXR**.
 - The *Authoring Web Site* offers a friendly interface that allows users to add assets, create interactive elements, design scenes, and deploy their AR applications. It has been implemented using **Vue.js** and **Three.js**.

Figure D. 19. Architecture of the framework

All technologies used in our framework are proven, robust, stable and well-known among web developers.

Framework usage

Using the framework is done through two web applications: authoring and viewing. The authoring application allows creators to organize digital assets into scenes, while the viewing application enables consumers to access and interact with the augmented reality experiences created by the authors.

Authoring application: running on desktop or mobile, it allows to create projects composed of an ordered collection of scenes. A scene itself is a localized collection of assets: it is created by placing assets (3D meshes and sounds) and annotations (textual or pictural) into the environment with integrated 3D editor in the application. Each scene can be customized with specific triggers for information display. With this scripting capability, one may develop personalized animations to exhibit specific features of the exposed artifacts. Figure D. 18 shows a screen capture of the authoring application.

Viewing application: visualizing the project is done through the web mobile viewing application (Android only at present time). After selecting the desired project, a target appears to place the virtual elements into the environment. The viewing application relies on ARCore and therefore does not need any physical marker for registration, but additional instructions may be given to the user if a specific placement is needed to align the virtual elements with the environment. Navigating through the scenes is done with the buttons at the bottom of the main screen, as shown on Figure D. 18.

Two example projects are shown in the accompanying video: the evolution over time of the architectural heritage of the [Soieries Bonnet](#) in Jujurieux, France, and a short presentation of the [Khotyn Fortress](#) in Ukraine.

Conclusion and perspectives

Our framework is an intuitive open-source solution for designing augmented reality experiences. It offers accessible tools for designing immersive and engaging AR experiences, particularly suited for cultural heritage. It is user-friendly and does not require any coding abilities, making it accessible to curators, museum managers and educators, and is open-source, promoting collaboration, customization, and innovation within the AR development community. A practical case usage has been demonstrated. The source code and a test implementation are available at hera.univ-lyon1.fr.

Future Extensions are numerous and in the hands of the developers and creators. We may mention, among many others, iOS compatibility (to this date, iOS does not support WebXR), integration of machine learning algorithms to enhance object recognition and interaction capabilities, as well as enhanced interactivity, gamification, social integration, real-time collaboration, and VR compatibility. We hope that these advancements will further enhance the way we experience and interact with cultural heritage artifacts.

References

- Adams, D. (2024). Awesome WebXR. Available at: <https://github.com/msub2/awesome-webxr>.
- Blattgerste, J., Behrends, J., & Pfeiffer, T. (2023). TrainAR: An Open-Source Visual Scripting-Based Authoring Tool for Procedural Mobile Augmented Reality Trainings. *Information* 14, 219. <https://doi.org/10.3390/info14040219>
- Hampshire, A., Seichter, H., Grasset, R., & Billingham, M. (2006). Augmented reality authoring: Generic context from programmer to designer. <https://doi.org/10.1145/1228175.1228259>
- Holden, C.L., Gagnon, D.J., Litts, B.K., & Smith, G. (2014). ARIS: An Open-Source Platform for Widespread Mobile Augmented Reality Experimentation, in: *Technology Platform Innovations and Forthcoming Trends in Ubiquitous Learning*. IGI Global, pp. 19–34. <https://doi.org/10.4018/978-1-4666-4542-4.ch002>
- Shekhar, S., Maheshwari, P., J, M., Singhal, A., Singh, K.K., & Krishna, K. (2019). ARComposer: Authoring Augmented Reality Experiences through Text, in: *Adjunct Proc. 32nd Ann. ACM Symp. on User Interface Software and Technology, UIST '19 Adjunct*. Association for Computing Machinery, New York, NY, USA, pp. 63–65. <https://doi.org/10.1145/3332167.3357116>
- Whitlock, M., Fitzmaurice, G., Grossman, T., & Matejka, J., (2020). AuthAR: Concurrent Authoring of Tutorials for AR Assembly Guidance. *Graph. Interface Conf.*

Exploring the use of experiential marketing and extended reality technology to support the consumer learning process of sustainability policies and principles.

Rumbidzai Gandiwa¹, Rose Leahy¹, and Pio Fenton²

¹*Munster Technological University, Department of Marketing and International Business, Cork, Ireland.*

²*Munster Technological University, Teaching and Learning Unit, Cork, Ireland.*

Corresponding Author: rumbidzai.gandiwa@mycit.ie

Keywords: Experiential Marketing, Extended Reality Technology, Virtual Reality, Consumer Learning, Sustainability Messaging, Immersive Experiences

Abstract

Purpose: This study will contribute to existing knowledge by providing insight into how marketers can harness the potential of extended reality technologies, by creating experiences that may lead to modified behaviour regarding sustainability issues. **Design/methodology/approach:** This is an exploratory research study, which will adopt a mixed methodology research design. **Research implications:** The research will contribute to original literature on the following aspects; the incorporation of fun/games in the context of experiential marketing and extended reality technology campaigns, the impact of utilising extended reality technology in promoting/ encouraging sustainability through experiential marketing campaigns, and the influence of combining experiential marketing and extended reality technology on promoting customer learning of sustainability principles. **Practical/Social implications:** The results of the research will contribute to the global efforts to change attitudes while prompting people towards adopting a more sustainable and environmentally friendly lifestyle through the use of impactful sustainable messaging (through effective integration of XR technology in experiential marketing.) **Originality/value:** The research addresses several fundamental gaps found in the existing literature including the incorporation of fun in the context of experiential marketing combined with extended reality technology campaigns, the outcome of utilising extended reality technology in promoting sustainability through experiential marketing campaigns, and the influence of combining experiential marketing and extended reality technology in promoting customer learning of sustainability principles.

Recyclage VR application



Figure D. 20. Images showing the Recyclage VR application

Introduction and Contribution of the proposed research

Sustainability concerns are increasingly significant as consumers demand greater environmental responsibility and authenticity from brands. Dettori (2019) suggests that employing experiential tactics can effectively educate customers about sustainability in an engaging manner. Extended Reality (XR) technologies, such as Virtual Reality (VR), offer multi-sensory experiences that allow marketers to simulate real-life scenarios, thereby influencing customer perceptions, judgments, and behaviours (Laukkanen et al., 2021).

This study aims to promote a shift towards more sustainable waste management practices, aligning with the United Nations Sustainable Development Goal 12, which advocates for sustainable consumption and production. Ineffective household waste management poses serious public health risks, including physical injuries, non-communicable diseases from prolonged exposure, and rodent infestations (Fadhullah et al., 2022). These issues also extend to environmental degradation, contributing to air, land, and water pollution, which adversely impacts biodiversity and local ecosystems.

Globally, nations are grappling with the adverse effects of the triple planetary crisis, encompassing climate change, biodiversity loss, and pollution. This situation has amplified the demand for sustainable lifestyles and a shift in public mindset. Effective customer education is pivotal in this transition, as studies reveal that informed individuals are more likely to engage in waste separation, whereas uninformed

individuals tend to mix waste, leading to contamination and unsustainable landfill disposal (Hasan, 2004; Amasuomo & Baird, 2016). Hasan emphasizes that public and private sector organizations can mitigate these issues through ongoing awareness and education campaigns. Evison and Read (2001) further suggest that these campaigns should be free, simple to use, and easily comprehensible.

Given that marketing managers can reach broad audiences, they have the potential to educate consumers on responsible consumption and production practices. The goal of the proposed research is to develop a framework that enables marketing practitioners to enhance their educational campaigns by leveraging impactful technologies such as Extended Reality (XR). These campaigns could be executed through experiential activations or pop-up booths at high-traffic events like festivals, conferences, and trade shows, thereby boosting brand authenticity, value, customer engagement, and image.

The Environmental Protection Agency (2022) has emphasized the need for enhanced support for households in waste sorting, recommending targeted awareness campaigns as a solution. Public policymakers could utilize the proposed framework to educate communities about waste management at popular events, such as the Cork on a Fork Fest and Cork Carnival of Science. The innovative nature of XR technologies, coupled with strategic social media promotion, is expected to generate significant interest and participation through word-of-mouth, enhancing the campaign's effectiveness.

Research Methodology

The research will adopt the research philosophy of Pragmatism which is the paradigm that supports use of mixed methods and is used by researchers who want to produce knowledge that is socially useful. In keeping with the theme of sustainability, this research will focus on the Waste Management Industry and how marketing practitioners can potentially use experiential marketing and extended reality technology to transform people's waste management culture into a more sustainable one. This is in line with SDG 12 which focuses on ensuring sustainable consumption and production patterns. For the purposes of this study the [RecyclageVR application](https://recyclagevr.com/) will be used (See Figure D. 20. This application was developed by VRAI Studios (France), and has been licensed for the purposes of the study. To use this application, participants will wear a VR headset and engage in a virtual process of sorting waste.

The researcher will use purposive sampling where participants for the study are selected based on certain characteristics (for example; over 18 and with self-reported partial knowledge of waste characterisation). Participants who fit this description will be provided with a baseline test to determine their knowledge levels and a further selection of participants will be done. The participants will be split into 2 groups by random assignment (control and experimental). The control group will be asked to watch on a computer/tablet screen, an animated 2D explainer and demonstration video in the RecyclageVR application with instructions on how waste characterisation works. They will take a test in the same application at the end to test their applied knowledge. The experimental group on the other hand will take part in the lesson via a Head Mounted Device using the VR application RecyclageVR³⁵. After the lesson demonstration exercise, the participants will also take a test to check their levels of applied knowledge. See Figure D. 21 below for a summary outline of methodological plan.

³⁵ <https://recyclagevr.com/>

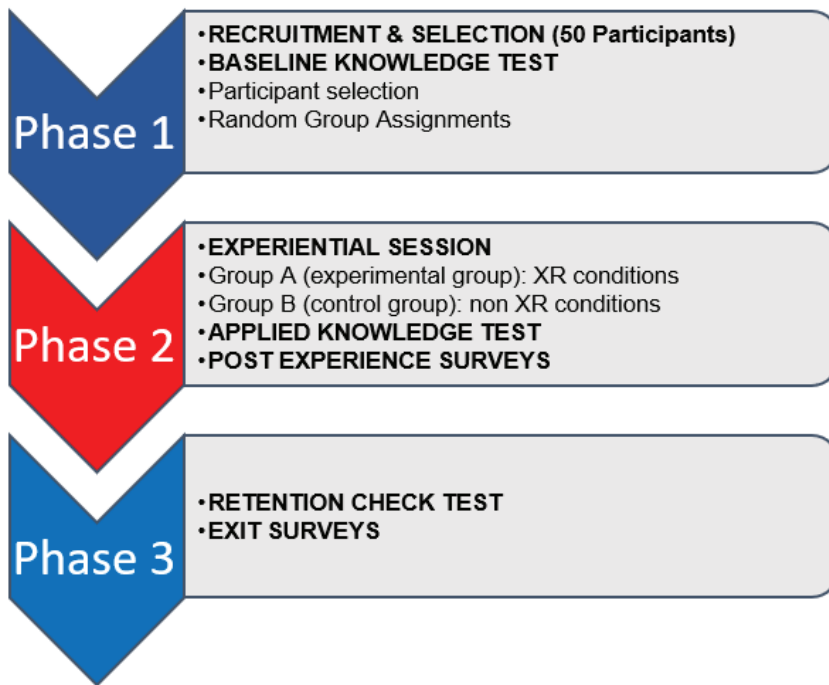


Figure D. 21. Summary of proposed methodology

References

- Amasuomo, E., & Baird, J. (2016) The Concept of Waste and Waste Management. *Journal of Management and Sustainability*, 6(4), p. 88. Available at: <https://doi.org/10.5539/jms.v6n4p88>.
- Dettori, A. (2019). Sustainability as a Matrix of Experiential Marketing, *International Journal of Marketing Studies*, 11(2), p. 29. Available at: <https://doi.org/10.5539/ijms.v11n2p29>.
- Environmental Protection Agency (no date) 2022 Household Municipal Waste Characterisation. Available at: <https://www.epa.ie/our-services/monitoring--assessment/waste/national-waste-statistics/waste-characterisation/>
- Evison, T. & Read, A.D. (2001) Local Authority recycling and waste — awareness publicity/promotion. *Resources, Conservation and Recycling*, 32(3–4), pp. 275–291. Available at: [https://doi.org/10.1016/S0921-3449\(01\)00066-0](https://doi.org/10.1016/S0921-3449(01)00066-0).
- Fadhullah, W., Imran, N. I. N., Ismail, S. N. S., Jaafar, M. H., & Abdullah, H. (2022). Household solid waste management practices and perceptions among residents in the East Coast of Malaysia. *BMC public health*, 22, 1-20. Available at: <https://doi.org/10.1186/s12889-021-12274-7>.
- Hasan, S.E. (2004) Public Awareness Is Key to Successful Waste Management. *Journal of Environmental Science and Health, Part A*, 39(2), pp. 483–492. Available at: <https://doi.org/10.1081/ESE-120027539>.
- Laukkanen, T., Xi, N., Hallikainen, H., Ruusunen, N., & Hamari, J. (2022). Virtual technologies in supporting sustainable consumption: From a single-sensory stimulus to a multi-sensory experience. *International Journal of Information Management*, 63, 102455. Available at: <https://doi.org/10.1016/j.ijinfomgt.2021.102455>

5D/3D Virtual Reality for an effective Cognitive Rehabilitation: the VESPA 2.0 Project

Emanuele Maria Merlo¹, and Salvatore Marco Pappalardo²

¹ *Department of Biomedical, Dental Science and Morphological and Functional Images, University of Messina, Messina, Italy*

² *Software Engineering Italia S.r.l., Piazza A. Lincoln 7, Catania, Italy*

Corresponding author: emerlo@unime.it

Keywords: VESPA 2.0 Project, Virtual Reality, Virtual Reality Interventions, Neurocognitive Disorders, Developmental Disorders, Cognitive Rehabilitation

Topic of interest: Medicine and Rehabilitation

Evidence referred to the efficacy of Virtual Reality interventions on diagnosis and treatment of cognitive deficits linked to developmental disorders and diseases, psychiatric conditions and neurocognitive disorders is growing attention (Cammisuli et al., 2022; Fernández Montenegro et al., 2020; Goharinejad et al., 2022; Hundert et al., 2022; Magrini et al., 2022; Pappalardo et al., 2020; van den Bergh et al., 2021; Vicario & Martino, 2022).

VESPA 2.0 (Virtual Environment for a Superior neuro-PsychiAtry, 2nd generation³⁶ is a project funded and supported by the 2014-2020 PO FESR program (extended to 2024 due to CoVID-19 issues) of the Sicilian Region for Research and Innovation in Enterprises and is structured through the development, research and validation of applications based on Virtual Reality (VR) for the diagnosis and treatment of neurocognitive conditions.

Its characteristics will be presented depending on the first (2013-2015) and second (2021-ongoing) generation of VESPA, with particular reference to the first mentioned literature on the application and development of VR technology, VR treatment of neurocognitive conditions and previous versions of this intervention (Latella et al., 2024; Merlo & Pappalardo, 2022).

The reference is linked to a set of technologies directed to the interaction of people with computer applications in 3D Virtual Reality (5D) highly immersive space, involving in real-time the senses, cognition and human skills. Virtual Reality can in fact provide credible/realistic experiences in a specific virtual environment, through interactions involving subjects and objects properly related to the built

³⁶ <https://www.progettovespa.it>

environment leading to neuronal re-organization in people suffering for neurological diseases fostering training of fundamental cognitive functions (e.g. visuo-spatial memory, eye-hand coordination, etc.). The VESPA 2.0 project is also a network that shares diagnostic and performance data in an internationally distributed environment whose services allow access to the data produced by clinical practices according to the laws on archiving and managing sensible/personal data.

Validation is on-going and must be continued in accordance with the cross-validation definition, thus in line with the accepted test methods considering the compliant objectives of the included tools. This process involves previously validated methods and tools, along with innovative solutions. In these terms VESPA is subjected to experimentation and validation together with classically used (and validated) solutions, in order to compare the results and consider the significance represented by an acceptable error threshold.

The first generation of VESPA started in 2013 (ended in December 2015), financed by the 2007-2014 PO FESR funds of the Sicilian Region for Research and Development in Enterprises. In particular, since the first generation, VESPA was referred to both the domain of development and adult. The diagnostic fields involved were related to Alzheimer's disease for adults, Intellectual Disability and Language Disorders for developmental domains respectively. The second-generation implementation level concerns the treatment of Multiple Sclerosis (Relapsing-Remitting Multiple Sclerosis - RRMS), Parkinson's Disease, consider in clinical terms and consequently in terms of evaluation and treatment.

With reference to the structure of the project tools to organize rehabilitation paths through batteries of tasks, to be organized in series. The specific tasks include parameters and metrics closely related to patient performance. Considering the parameters, they are established in terms of inputs/settings useful for the performance of tasks. The metrics are related to the performance of functions, with direct reference to cognitive functions.

The course of each task has been structured with direct reference to deficient functions, detected through psycho-diagnostic tools.

The rehabilitation activity runs through a set of standardized protocols whose instances are bound to a specific patient and highly customizable. The protocols, both for adults and for developing age groups and for clinically acceptable (inclusion criteria), are structured as follows:

- Administration of neuropsychological tests (A - baseline, ex ante evaluation);
- Administration of VESPA tasks (B treatment, organised in sessions)
- Administration of neuropsychological tests (ex post evaluation)

Adult batteries have been set up to provide tasks involving patients in interactions with objects, then subject to a series of repetitions.

The rehabilitation tasks of the VESPA 2.0 (2nd Generation) project consist of a set of serious games, in line with VESPA (1st Generation) (Merlo et al., 2023).

The level of implementation concerns a revision of the 1st Generation task, with porting, adaptation, adaptation and reassuming referred to the latest technological adaptations, with in addition new implementations of batteries of new generation. The division into specific age domains has to do, in continuity, with Adults and Minors. The level of technological expression continues to involve 3D Virtual Reality for totally immersive 3D Virtual Reality system "CAVE" (which reaches up to 5D for the adult domain) and simultaneously, in the 2.5D version, for Tablet PC running Android/iOS and Windows PC as well as Meta VR Head Mounted Display (3D). The diagnostic domains treated (see adult and minor distinctions and related state of the art) are Alzheimer's disease, Multiple Sclerosis, and Parkinson's disease for adults, Intellectual disability and Communication disorders in children.

In Europe, National Health Services promote an innovative approach to cognitive disabilities aimed at building a support network that makes cognitive rehabilitation easier and produces social integration, seeking a more effective deployment of resources. The adoption of this pioneering approach moves through the implementation of diagnostic and rehabilitation services available not only in hospitals, but also in schools (e.g. ID and CD), specialist structures and healthcare facilities (AD, PD, MS).

Ideally, to reduce costs, cognitive-motor rehabilitation provided by hospitals should be done by video conferencing or tele-supervision. A fully immersive Virtual Reality computerized system such as VESPA, remotely tele-supervised by clinical personnel located in control centers, fully respects the approach described above. This system provides a quantitative and qualitative evaluation of motor-cognitive functions, allowing a quick and easy diagnosis or treatment (Latella et al., 2024).

VESPA 1st and 2nd Gen have developed a highly integrated ICT system that includes:

- A cognitive software (The VESPA Cognitive Training Software), developed in different formats (VR 3D/5D, VR 2.5D, etc.), that combines specific Tasks/Exercises and Cognitive and Executive functions for training; rehabilitation paths (Protocols) are expressed in terms of training sessions that combine several tasks in a time slot on a specific device.
- An electronic infrastructure built on a cloud computing solution and hosting the necessary services for the operation of the system such as:
 - A management Web Application where clinicians and healthcare professionals can define treatment paths and monitor patient performance (the VESPA Portal);
 - A clinical database that stores the metrics related to the cognitive training activities of the patient;
 - Other web services necessary for the -real-time management of the information produced by the system and the training activities.
 - Support of Big Data Analysis and automatic progression of rehabilitation paths depending on patient's performance. Designed for AI integration (on-going).
- A set of dedicated display devices such as:
 - CAVE Virtual Environment, a VR environment where projectors are directed to the walls of 2 to 5 cubes the size of a room
 - Touch devices such as Touch Tables, TVs and Tablet PCs;
 - Display Meta (Oculus) Rift, Quest e Quest2;
 - Personal computer microsoft Windows 10+ VESPA systems combine these elements in order to provide:
- A fully immersive 3D VR cognitive rehabilitation software that harnesses the power of CAVE Virtual Environment;
- A Meta (Oculus) 3D VR rehabilitation app;
- VR 2.5 mobile app for cognitive rehabilitation for Tablet, Touch table and TV for iOS and AndroidOS;
- VR 2.5 mobile AndroidOS native app for cognitive rehabilitation for Tablet and Smartphones;
- A MicroSoft Windows rehabilitation app for personal computers;
- A multi-user web portal for the management of rehabilitation pathways (for doctors) and performance monitoring (also for healthcare professionals and patients/caregivers).

A demonstrator of the VESPA 2.0 system and its e-infrastructure is easily set-up everywhere necessary. The VESPA web portal is available 24 h a day through the Internet. A person willing to test the rehabilitation features can easily register as a patient in the example Clinical Organisation, where a clinical environment is simulated. The patient/caregiver/doctor's operation on the Web Portal and consequences on the planning for devices and rehabilitation can be shown. In the meanwhile, the Clinical App can be demonstrated as simple as one-two-three through the several devices already integrated in the VESPA 2.0 platform. We propose to demonstrate the Clinical Software by mobile devices as Tablet and Meta VR HMD. A video recording or pictures of the operation in a CAVE virtual environment can also be simply shown and commented. HMI devices used in a CAVE are easily transportable so that Tracking system cameras or Flystick for object manipulation could be presented as well.

Pictures

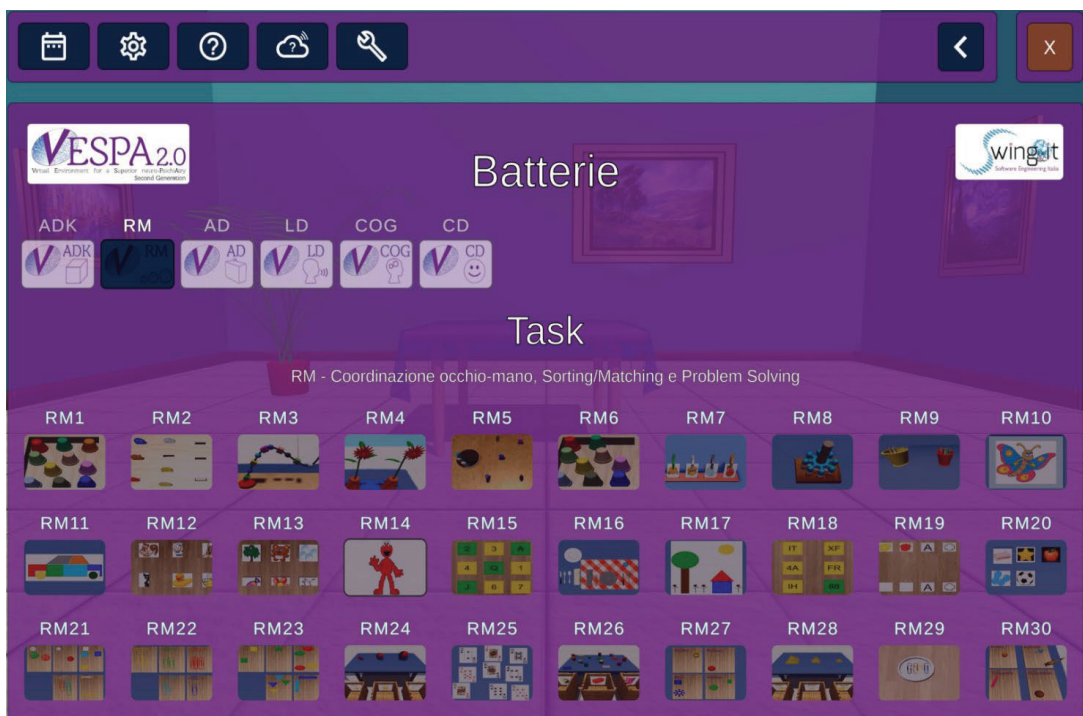


Figure D. 22. Menu screen

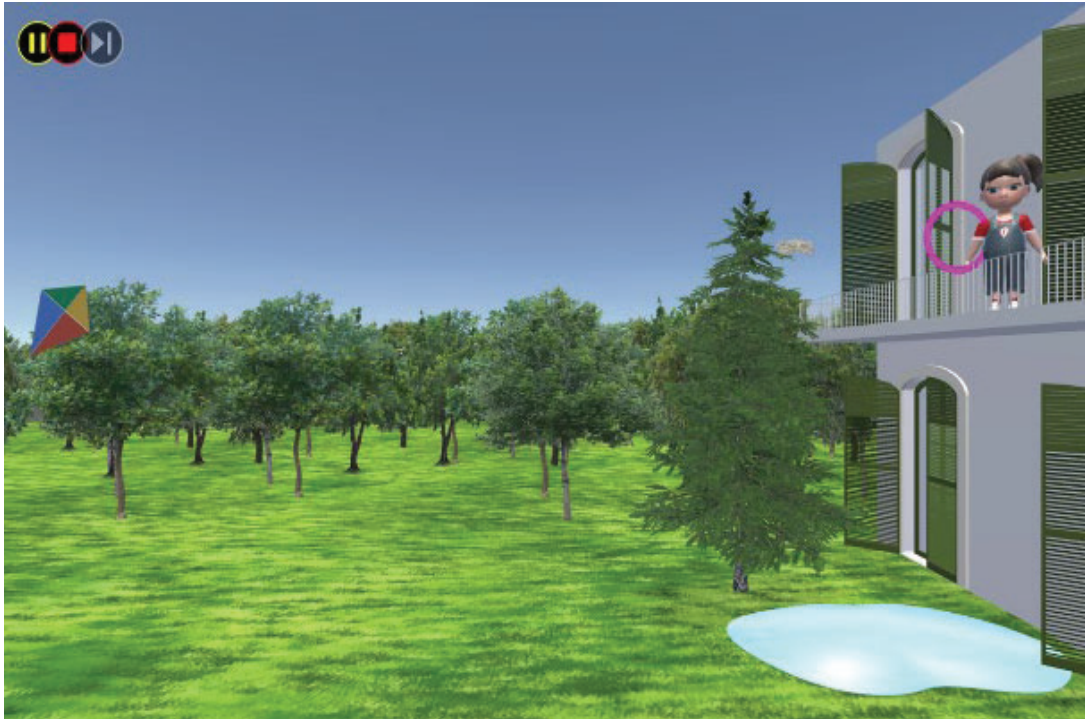


Figure D. 23. CD2 Task performance

References

- Cammisuli, D. M., Cipriani, G., & Castelnuovo, G. (2022). Technological solutions for diagnosis, management and treatment of Alzheimer's disease-related symptoms: A structured review of the recent scientific literature. *International Journal of Environmental Research and Public Health*, 19(5), 3122. <https://doi.org/10.3390/ijerph19053122>
- Fernández Montenegro, J. M., Villarini, B., Angelopoulou, A., Kapetanios, E., Garcia-Rodriguez, J., & Argyriou, V. (2020). A survey of alzheimer's disease early diagnosis methods for cognitive assessment. *Sensors*, 20(24), 7292. <https://doi.org/10.3390/s20247292>
- Goharnejad, S., Goharnejad, S., Hajesmaeel-Gohari, S., & Bahaadinbeigy, K. (2022). The usefulness of virtual, augmented, and mixed reality technologies in the diagnosis and treatment of attention deficit hyperactivity disorder in children: an overview of relevant studies. *BMC psychiatry*, 22(1), 4. <https://doi.org/10.1186/s12888-021-03632-1>
- Hundert, A. S., Birnie, K. A., Abla, O., Positano, K., Cassiani, C., Lloyd, S., ... & Stinson, J. (2022). A pilot randomized controlled trial of virtual reality distraction to reduce procedural pain during subcutaneous port access in children and adolescents with cancer. *The Clinical journal of pain*, 38(3), 189-196. <https://doi.org/10.1097/AJP.0000000000001017>
- Latella, D., Formica, C., Ielo, A., Grioli, P., Marra, A., Costanzo, D., Merlo, E. M., Pappalardo, S. M., Corallo, F., Marino, S., Quartarone, A., Calabrò, R. S., Maresca, G. (2024). A feasibility and usability study of virtual reality tool (VESPA 2.0) for cognitive rehabilitation in patients with Mild Cognitive Impairment: an ecological approach. *Frontiers in Psychology*, 15. (accepted). <https://doi.org/10.3389/fpsyg.2024.1402894>
- Magrini, M., Curzio, O., Tampucci, M., Donzelli, G., Cori, L., Imiotti, M. C., ... & Moroni, D. (2022). Anorexia nervosa, body image perception and virtual reality therapeutic applications: State of the art and operational

- proposal. *International Journal of Environmental Research and Public Health*, 19(5), 2533. <https://doi.org/10.3390/ijerph19052533>
- Merlo, E. M., & Pappalardo, S.M. (2022). Virtual Reality Application for Neurocognitive and Developmental Disorders: The Vespa 2.0 Project. In *Book of Abstract, 30° Congresso dell'Associazione Italiana di Psicologia* (pp. 519-519).
- Merlo, E. M., Myles, L. A. M., & Pappalardo, S. M. (2022). The VESPA project: Virtual reality interventions for neurocognitive and developmental disorders. *Journal of Mind and Medical Sciences*, 9(1), 16-27. <https://doi.org/10.22543/7674.91.P1627>
- Merlo, E.M., Pappalardo, S. M. (2023). Subjective experience of Virtual Reality neurorehabilitation: qualitative data from Vespa 2.0 project. *Proceedings: XXIII National Congress Italian Psychological Association Clinical and Dynamic Section Florence, 15th – 17th September 2023* (p. 578). *Mediterranean Journal of Clinical Psychology*, 11(2-Supplementum).
- Pappalardo, S. M. (2020). *Vlad-Virtual Reality Application for Treatment of Psychosomatic Conditions: A report at final stage of software validation process*. *Mediterranean Journal of Clinical Psychology*, 8(3). <https://doi.org/10.6092/2282-1619/mjcp-2868>
- Vicario, C. M., & Martino, G. (2022). Psychology and technology: how Virtual Reality can boost psychotherapy and neurorehabilitation. *AIMS neuroscience*, 9(4), 454. <https://doi.org/10.3934/Neuroscience.2022025>
- van den Bergh, R., Bloem, B. R., Meinders, M. J., & Evers, L. J. (2021). The state of telemedicine for persons with Parkinson's disease. *Current opinion in neurology*, 34(4), 589-597. <https://doi.org/10.1097/WCO.0000000000000952>

UnityMol-Lite: Cross-platform Collaboration and Interactive Methods for Molecular Modeling in XR

Mariano Spivak¹, Bastien Vincke², Hubert Santuz³, Jérôme Henin³,
Antoine Taly³, Marc Baaden³, and Nicolas Férey¹

¹*Laboratoire interdisciplinaire des sciences du numérique, Université Paris-Saclay, Paris, France*

²*Laboratoire SATIE, CNRS, Paris, France*

³*Laboratoire de Biochimie Théorique, CNRS, Paris, France*

Corresponding author: mariano.spivak@universite-paris-saclay.fr

Keywords: Extended Reality, Tangible Interface, Molecular Dynamics

Introduction

Molecular modeling consists of a battery of theoretical and computational methods to model the behavior of molecules. With the growth of computing power, molecular modeling has become a cornerstone of material, pharmaceutical and biomedical research. Among the computational methods, molecular dynamics simulations have shown tremendous predicting power and have become a perfect complement to experimental methods. Interactive Molecular Dynamics (IMD) is an advanced simulation technique that provides real-time interaction with the molecules, through picking and pulling of individual pieces (e.g. atoms, residues). However, this interface mode makes the technique slow and imprecise. As such, IMD has become a great tool for educational and dissemination purposes, but not for advanced research (Lanzerac, Férey et al., 2022).

On the other hand, the usage of extended reality (XR) technologies in molecular modeling have shown great promise, because it provides a framework for natural 3D-visualization and intuitive interaction. (Kuták et al., 2023) Examples of interactive visualization and IMD in XR include (Deeks et al., 2020) and (Lindow et al., 2019). Finally, XR collaboration frameworks improve the exchange of ideas and discussion, so ever present in research.

In this work, we use state-of-the-art technologies to create a collaborative experience of an interactive molecular dynamics simulation (Figure D. 24). We propose an efficient interaction technique based on the concept of a tangible interface, which could greatly improve IMD capabilities. Finally, we present a cross-platform solution improving ease of use and dissemination.

UnityMol-Lite

UnityMol-Lite is an optimized version of the molecular visualization application UnityMol, (Laureanti et al., 2020) that was built to support new XR technologies and hardware. In particular, the application offers support for both PC-based and standalone HMDs, with a focus on cross-platform functionality, collaborative sessions and advanced interactive methods. In order to target standalone HMDs (e.g. Meta Quest 2 and 3) and their new technologies (video see through and hand tracking), a complete rework of the XR/interaction components (using OpenXR, XR Interaction Toolkit) was done, compared to the original UnityMol application. In contrast, many of the advanced visualization features of UnityMol, e.g. raytracing, were removed to support compatibility and performance.

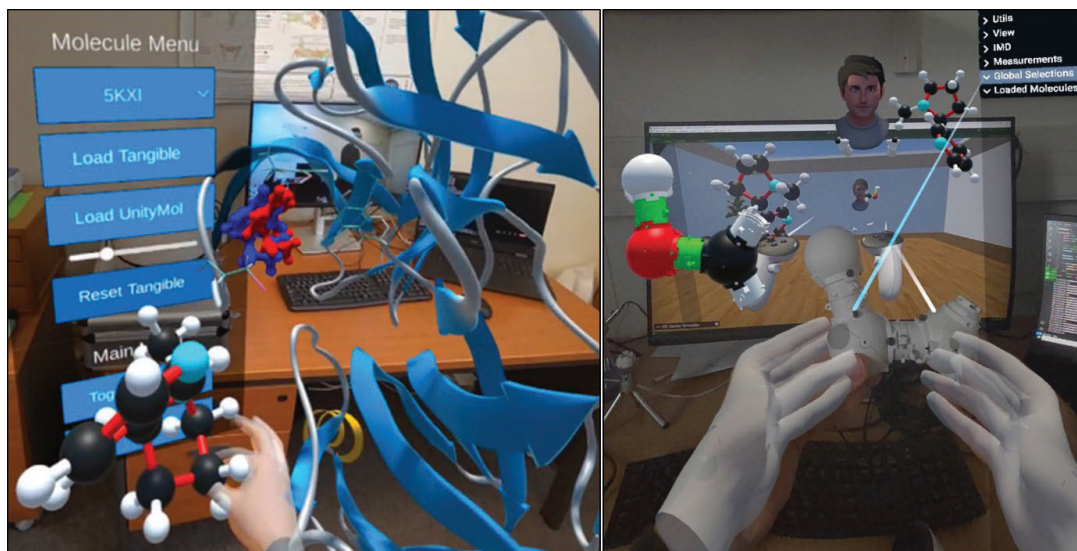


Figure D. 24. Snapshot of an IMD experience on UnityMol-Lite (left). The user can interact with a virtual twin of a small molecule (on the front in black, red and white), defining a target conformation for the running simulation. Multiplayer session showing player avatar and tangible interface (right)

In addition to the traditional banner user interface (UI), (Figure D. 25, center) we designed a new XR UI with accessibility at its core, providing an intuitive and immersive experience for users. This UI is integrated directly into the XR environment, by being anchored to the controller (Figure 2, left) or hand (Figure 2, right), allowing for natural interactions. It can be toggled and it ensures that frequent controls are always within reach, reducing the need for users to break their immersion and possible sight issues. This design allows users of all skill levels, including those new to XR, to engage with the interface in a more organic and user-friendly manner.

Collaborative sessions are supported through Photon Unity Networking. Avatars are used to display players' positions. The UI allows for alternate control of the scene elements.

The interactive molecular dynamics (IMD) method is achieved by the MDDriver library (Lanzerac, Laurent et al., 2022). Next section presents the complementary development done on the MDDriver library to support the new interface for IMD.

New Interface for Interactive Molecular Dynamics

The novel interface method is inspired by the concept of the tangible interface, whereby a physical object is used to model and interact with its virtual twin. This is particularly useful in molecular modeling because many drug discovery studies involve a molecular system characterized by a small component (e.g. ligand) embedded in a large one (e.g. protein).

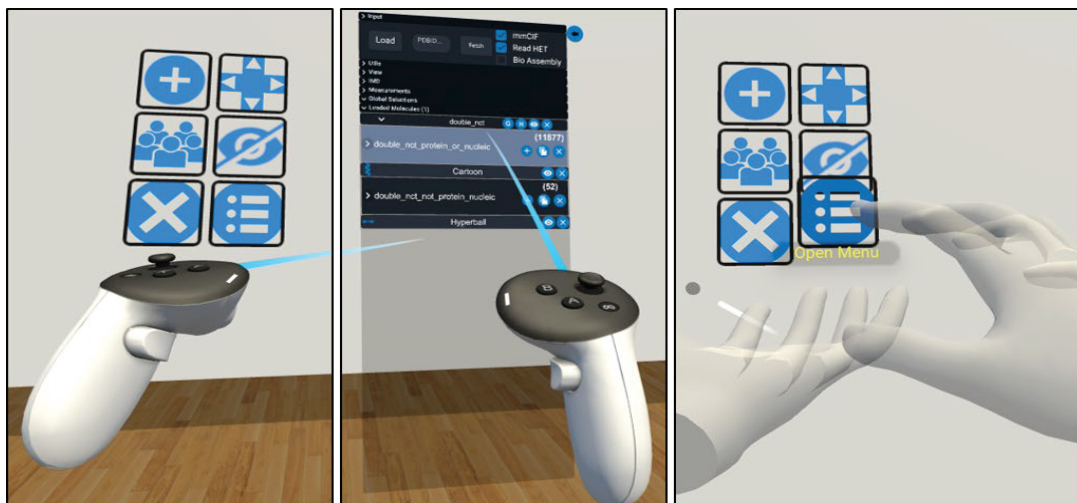


Figure D. 25. Composite of the user interfaces (UI) for UnityMol-Lite. A traditional banner UI, center, provides full functionality and information of the system. The new XR UI allows easy access to frequently used controls, and it is anchored to the controller (left) or the hands (right)

Therefore, in addition to the traditional interface method of pushing and pulling the individual components, we define a virtual “tangible” interface of the small component in the scene, as a way to interact with the simulation.

Figure D. 24 (left) shows a snapshot of the IMD experience where the user manipulates the virtual interface, by changing the position, orientation and local conformation (e.g. angles, dihedrals) in real time. MDDriver compares the virtual interface coordinates with the simulation twin, and utilizes the collective variables library (COLVARS, Fiorin et al., 2013) to compute the forces required by the running simulation. Thus, the virtual interface is used as the target for the running molecular dynamics simulation, resulting in a smooth and simple experience. This combination of COLVARS and MDDriver can also be used to interact with larger components: by defining appropriate collective variables the user can impose large conformational changes to a protein, akin to those obtained using other biasing methodologies (e.g. Steered MD, Targeted MD).

Finally, this development sets the algorithmic foundations for the usage of real physical objects as tangible interfaces for interactive molecular dynamics. Next section discusses the current progress on supporting a tangible molecular interface.

Support for Tangible Molecular Interface

Our group is currently developing a physical, modular and articulated interface to use in molecular modeling. (Vincke et al., 2020) The latest prototype features sensors to monitor adding, deleting and

connecting objects (atoms). In addition, the interface tracks rotation along connections, e.g. defining bond angles and dihedrals of the molecular model.

UnityMol-Lite supports these features, making the physical interface work as a molecular builder. In addition, the physical model can be used together with the interactive molecular dynamics method to define the target conformations. Figure D. 24 (right) shows a snapshot of the combined user experience with the physical interface and the AR mode (video see through). Support for tracking of absolute position and orientation will be revisited on the next prototype of the physical interface.

Conclusion and Outlook

A novel interface method for interactive molecular dynamics (IMD) based on the tangible interface concept was developed. To support the use of a physical model in extended reality, we redesigned the molecular visualization application UnityMol to support standalone HMD and video see through. UnityMol-Lite is optimized for accessible android devices (e.g. Meta Quest 2 and 3) and cross-platform collaboration using Photon Unity Networking. Next, we plan to integrate our work with tutorial and demo examples from Molplay (Baaden 2024) and distribute the application through our public repository. In addition, we will continue adding support for the tangible interface model.

Acknowledgements

This project has received funding from the Agence Nationale de la Recherche (ANR) under the grant ANR-21-CE45-0014.

References

- Baaden, M. (2024). MolPlay: Democratizing Interactive Molecular Simulations and Analyses with a Portable, Turnkey Platform. *The Journal of Physical Chemistry B*. <https://doi.org/10.1021/acs.jpcc.4c04712>
- Deeks, H. M., Walters, R. K., Hare, S. R., O'Connor, M. B., Mulholland, A. J., & Glowacki, D. R. (2020). Interactive molecular dynamics in virtual reality for accurate flexible protein-ligand docking. *Plos one*, 15(3), e0228461. <https://doi.org/10.1371/journal.pone.0228461>
- Fiorin, G., Klein, M. L., & Héning, J. (2013). Using collective variables to drive molecular dynamics simulations. *Molecular Physics* 111(22–23), 3345–3362. <https://doi.org/10.1080/00268976.2013.813594>
- Kuták, D., Vázquez, P.-P., Isenberg, T., Krone, M., Baaden, M., Byška, J., Kozlíková, B. & Miao, H. (2023). State of the Art of Molecular Visualization in Immersive Virtual Environments. *Computer Graphics Forum* 42, e14738. <https://doi.org/10.1111/cgf.14738>
- Lanrezac, A., Férey, N. & Baaden, M. (2022). Wielding the power of interactive molecular simulations. *WIREs Comput Mol Sci*. 12, e1594. <https://doi.org/10.1002/wcms.1594>
- Lanrezac, A., Laurent, B., Santuz, H., Férey, N. & Baaden, M. (2022). Fast and Interactive Positioning of Proteins within Membranes. *Algorithms* 15, 415. <https://doi.org/10.3390/a15110415>
- Laurenti, J., Brandi, J., Offor, E., Engel, D., Rallo, R., Ginovska, B., ... & Baker, N. A. (2020). Visualizing biomolecular electrostatics in virtual reality with UnityMol-APBS. *Protein Science*, 29(1), 237-246. <https://doi.org/10.1002/pro.3773>
- Lindow, N., Baum, D., Leborgne, M. & Hege, H.-C. 2019. Interactive Visualization of RNA and DNA Structures. *IEEE Transactions on Visualization and Computer Graphics* 25(1), 967-976. <https://doi.org/10.1109/TVCG.2018.2864507>
- Vincke, B., Ghaoui, M. A., Férey, N. & Martinez, X. (2020). Physical, Modular and Articulated Interface for Interactive Molecular Manipulation, *Sensors* 20(18), 5415. <https://doi.org/10.3390/s20185415>

Author Index and Tables of Figures & Lists of Tables

Author Index

Achilleopoulos		Böhme	
Nick	73	Andrea	143, 179
Ahsan		Bolierakis	
Moonisa	73	Spyridon Nektarios.....	105
Alesani		Bordegoni	
Andrea.....	35, 267	Monica.....	239
Amato		Borotis	
Giuseppe	201	Spiros	73
Ayache		Botev	
Julia.....	175	Jean	281
Baaden		Bourdot	
Marc.....	301	Patrick.....	87, 91, 227
Bach		Bowman	
Cédric	219	Eileen.....	243
Baldwin		Boychuk	
Mark.....	243	Rostyslav	261
Bardy		Briesemeister	
Benoît	175	Frank	143, 179
Baudry		Brodbeck	
David.....	193	Tanja.....	113
Beddiar		Bryson	
Karim.....	193	Kathleen.....	131
Belfiore		Buchholz	
Angelo.....	109	Florian.....	41
Benazzouz		Burmester	
Yazid	117	Michael.....	113
Bernardos		Campion	
Ana M.....	77	Simon.....	55
Biswas		Caridakis	
Suhana	41	George.....	163
Bodin		Carlhoff	
Mathis.....	287	Christian.....	249

- Carrara
Fabio 201
- Carulli
Marina 239
- Casar
José R. 77
- Casarosa
Gianluca 31
- Castillo López
Galo 117
- Chalmers
Andrew 67
- Chatzifoti
Olga 73, 171
- Cheong
DaEun 67
- Cholopoulou
Eirini 159
- Cuevas-Rodriguez
Maria 137
- Dang
Son 91
- de Chalendar
Gaël 117
- Di Benedetto
Marco 201
- Diamantis
Dimitrios 159
- Dias da Silva
Emanuel 281
- Dix
Martin 49, 257
- Döhler
Torsten 143, 179
- Donnarumma
Gian Paolo 109
- Dorociak
Radoslaw 55
- Falciglia
Marco Salvatore Carmelo 109
- Farrell
Rachel 243
- Farrugia
Jean-Philippe 287
- Federico
Giulio 201
- Fenton
Pio 291
- Ferey
Nicolas 301
- Foitzik
Andreas 143, 179
- Gandiwa
Rumbidzai 291
- Ganias
Giorgos 45
- Gao
Yuze 125
- Generali
Francesco 109
- Georgiou
Giannis 27
- Geslot
Benjamin 211
- Gonzalez-Toledo
Daniel 137
- Goriachev
Vladimir 35, 267
- Grund
Fabian 143, 179
- Gutierrez-Parera
Pablo 137
- Hakanen
Taru 23
- Han
JungHyun 67
- Häring
Simon 61
- Havard
Vincent 193

- Hebborn
 Anna Katharina.....175
- Heinrich
 Moritz143, 179
- Helin
 Kaj..... 23, 31, 35, 267
- Henin
 Jérôme..... 301
- Hong
 Leyi 183
- Hou
 Wenjun.....125, 183
- Howard
 Thomas211
- Hu
 Tianrui..... 125
- Ioannou
 Lefteris..... 27
- Islam
 Tariqul.....167
- Jaillet
 Fabrice.....287
- Jean
 Pierre.....175
- Jia
 Lai Pin239
- Kanellopoulos
 Asimakis..... 159
- Karaseitanidis
 Ioannis..... 105
- Karjalainen
 Jaakko.....23, 31, 35
- Katika
 Tina..... 97, 105
- Kavouras
 Panagiotis..... 123
- Khan
 Rehan..... 151
- Kiernan
 Paul.....31
- Kim
 Myong Gon.....67
- Klimant
 Franziska 49, 257
- Koch
 Lukas 249
- Kokkalis
 Konstantinos..... 105
- Konstantinidis
 Fotios 105
- Koppusch
 Klarissa..... 143, 179
- Korb
 Samuel..... 49, 257
- Korozi
 Maria 147
- Kougioumtzian
 Lori.....45
- Koukoudis
 Konstantinos.....97
- Kracht
 Christopher 143, 179
- Krauß
 Stephan.....61
- Krayner
 Bastian83, 273
- Krenz-Baath
 Rene 143
- Kulzer
 Manuel 113
- Kuts
 Vladimir.....261
- Kuula
 Timo267
- Kyrlitsias
 Christos27
- Lanitis.....27
- Lazou
 Chrysoula..... 159
- Leahy

- Rose 291
- Lei
Zhenhong..... 231
- Leonidis
Asterios 147
- Li
Qian..... 183
Xinjun..... 231
- Liarokapis
Fotis..... 277
- Lombard
Alexis..... 117
- López
Paula 77
- Lorré
Jean Pirre..... 117
- Lougiakis
Christos 45
- Lv
Meiyu..... 183
- Mahmood
Kashif 261
- Mangina
Eleni..... 155
- Martinez
Christian..... 151
- Melzer
Annegret..... 49, 257
- Merlo
Emanuele Maria 109, 295
- Meyer
Georg..... 55
- Michael-Grigoriou
Despina 27
- Minaskan
Narek 83, 273
- Mulay
Hrishikesh..... 155
- Naskou
Kostas..... 105
- Nguyen
Hong Son 67
Huyen 227
- Niarchos
Anastasios 45
- Oikonomou
Eleni 73
- Ojeleye
Jamiu..... 61
- Östman
Anders 87, 91
- Otto
Tauno..... 261
- Ouedraogo
Inoussa 227
- Pagani
Alain 83, 273
- Pappalardo
Salvatore Marco 109, 295
- Peltola
Johannes..... 23
- Perret
Jérôme..... 211
- Petrou
Vasilis 91
- Pistre
Camille..... 175
- Pizzagalli
Simone Luca 261
- Plexousaki
Alexandra..... 147
- Pöder
Irma 215
- Podguzova
Mariia 61
- Pudota
Jyothi..... 167

Redfern		Stavridi	
Sam	155	Vana	123
Reyes-Lecuona		Stavroulia	
Arcadio	137	Kalliopi-Evangelia.....	27
Rhee		Stefanouli	
Taehyun	67	Vasiliki.....	159
Rice		Stephanidis	
Marelle	243	Constantine.....	147
Richardson		Stiens	
Kathleen.....	131	Philipp.....	49, 257
Riedlinger		Stricker	
Urs	207	Didier	61, 83, 175, 273
Röbnikov		Strimpakos	
Daniil.....	261	Nikolaos	159
Rossoni		Sun.....	281
Marco	239	Ningyuan.....	281
Roussou		Sunkara	
Maria.....	171	Mohith.....	151
Runde		Sutherland	
Christoph Paul	13	Jessica	131
Sacco		Symotiuk	
Marco	87	Ivan.....	261
Santuz		Taly	
Hubert	301	Antoine.....	301
Schäfer		Tantaroudas	
Wolfgang.....	187	Nikolaos D.....	97
Schiavi		Tellez	
Barbara.....	193	Bruno	287
Schiettekatte		Tikka	
Andoni Retagui.....	41	Petri	23, 35
Scholl		Tomar	
Ingrid	249	Rahul.....	151, 167
Smagas		Tsaknaki	
Konstantinos.....	87	Electra	87, 91
Soto-Romero		Tsimiklis	
Carmen.....	137	Georgios	105
Spivak		Tymchenko	
Mariano.....	301	Maryna	199
Stamou		Uzun	
Aikaterini.....	91	Yücel.....	41

Vincke			
Bastien.....	301		
Volkmar			
Gioia.....	187		
Wang			
Hongrun.....	125		
Wirbatz			
Karoline.....	237		
Wittfeld			
Gerhard.....	249		
Wolf			
Eckart.....	143, 179		
Zang			
Guoqin.....	219		
Zhang			
Yiran.....	219		
Zoi			
Stavroula.....	163		

Tables of Figures

Figures in Application Track

Figure 1.	Politico.eu reports about China’s metaverse ambitions in August 2023 (Volpicelli, 2023).	14
Figure 2.	Magazine by Cointelegraph reports about South Korea's metaverse ambitions.....	16
Figure 3.	Global metaverse / virtual worlds strategies comparison	17
Figure 4.	Point cloud capturing, processing and visualization pipeline	24
Figure 5.	Tested application use cases.....	25
Figure 6.	VRTEACHER application main menu.....	28
Figure 7.	The three scenarios developed under the VRTEACHER project.	29
Figure 8.	Through the eyes of the teacher for all three scenarios and teacher’s inner voice.....	29
Figure 9.	Start button to begin the scenario.....	30
Figure 10.	VirWAIT use cases.....	32
Figure 11.	DPIAR-v1 system set-up.....	33
Figure 12.	The MR systems in use during procedure execution (LSS- Basement procedure) and its automatic reporting.....	33
Figure 13.	System architecture of the application.....	36
Figure 14.	Interaction design graph.....	37
Figure 15.	3D assembly instructions manipulation.....	38
Figure 16.	UR10e cobot cell for IKEA step stool assembly.....	39
Figure 17.	Part of the puzzle, where laser triggers the cannon via a mirror & Visualization of AI feedback.....	42
Figure 18.	Editing a card game in the game creation interface of the web application.....	46
Figure 19.	Playing a card game on a virtual table in VR & Playing a card game on a real table in MR.....	47
Figure 20.	Rendering of the VoirVR environment and the interaction concept.....	50
Figure 21.	Controls for manipulating facial palse	57
Figure 22.	User interface enabling users to control head and eye gaze of avatars	58
Figure 23.	User interface for the animation blending interface.....	59
Figure 24.	An overview of the pipeline that provides the user with pruning suggestions.....	61
Figure 25.	Comparison of grapevines.....	62
Figure 26.	Examples of keyframes.....	64
Figure 27.	AR interaction.....	67
Figure 28.	Computing the depth of the 3D character.....	68
Figure 29.	Collision and occlusion	68
Figure 30.	User study room.....	69
Figure 31.	System Architecture for AR Theatre.....	74
Figure 32.	Screenshot from AR customization UI. Right: screenshot from caption display.....	75
Figure 33.	AOI’s marked by colours for fixation classification.....	79
Figure 34.	Distribution of Technology order by cluster.....	81
Figure 35.	An overview of our setup.....	84
Figure 36.	Extended Reality-Based Behavioral Analysis Flowchart: Identifying Dietary Choices through Eye-Tracking, Heatmaps and Data Analysis.....	98
Figure 37.	Key Factors affecting food habits, and b) BMI of participants.....	99

Figure 38.	Mixed Reality Environment Setting with Data Storytelling Visuals.....	100
Figure 39.	Visual attention heatmaps generated within the MR environment for omnivorous participant	102
Figure 40.	A short conveyor was used for development.....	106
Figure 41.	An item of interest approaches the workstation of a worker wearing a HMD.....	107
Figure 42.	VESPA 2.0 Software on Virtual Room.....	112
Figure 43.	VESPA 2.0 Software on Android Tablet.....	112
Figure 44.	Structure of collaboration and artifacts in the transdisciplinary scenario-based design approach.	114
Figure 45.	Snow groomer exiting a garage during one of the phase 1 studies.....	114
Figure 46.	Business Meeting session with 2 VR participants and a Rainbow User.	118

Figures in Posters

Figure P. 1.	Overview of the XR4Human Project's outcomes and engagement activities.....	124
Figure P. 2.	Our device uses a two degree of freedom cylinder as the base.....	127
Figure P. 3.	Virtual scene in Unity.....	127
Figure P. 4.	A visualization of the six personhood types analysed against personhood qualifiers.	134
Figure P. 5.	Front-back confusion rate for yaw and pitch movements.	140
Figure P. 6.	Graphical representation of the cognitive affective model of immersive learning.....	145
Figure P. 7.	Aerodynamic condition simulation in VR.....	148
Figure P. 8.	View of our XR tool both in an indoor and outdoor setting.....	149
Figure P. 9.	DT framework for automated shotcrete planning and operation.	152
Figure P. 10.	Proposed System Architecture.	157
Figure P. 11.	Participants during the roller coaster simulation.....	161
Figure P. 12.	Landmark Identification as instantiated game objects	164
Figure P. 13.	Physical, Virtual & MediaPipe Metric Spaces.....	165
Figure P. 14.	Spherical Content Prototype Visualization.....	166
Figure P. 15.	Process diagram showing the digital twin (DT) creation.	167
Figure P. 16.	Incremental experiment design.....	172
Figure P. 17.	Illustration of the virtual environment and avatar representations.....	176
Figure P. 18.	Graphical representations of the impact of the experimental conditions SOLO.....	177
Figure P. 19.	Numbered components of a 3D devil's knot	180
Figure P. 20.	Specially designed cube to illustrate the 3D grid structure with guides for moving entire levels.	181
Figure P. 21.	System testing in natural environments.....	184
Figure P. 22.	Interact with virtual objects by gazing.....	185
Figure P. 23.	Gesture recognition evokes flat maps	185
Figure P. 24.	Immersive Cocooning Cube.....	188
Figure P. 25.	SWRP interface, with a QHSE rule.....	195
Figure P. 26.	Proposed architecture of an interoperable system between BIM - VR and QHSE rules.....	196
Figure P. 27.	3D visualisation of a room with switchgear compartments.....	200
Figure P. 28.	The relay cabinet with control bus circuit breakers and power bus circuit breakers. ..	200
Figure P. 29.	Stage 1 and Stage 2 of our Spatio-Temporal Diffusion Neural Architecture.....	202
Figure P. 30.	The Neural-Clipmap algorithm progressively converts a coarse octree	203
Figure P. 31.	Starting from a sequence of images	205
Figure P. 32.	Technology stack for WebAR applications.....	208

Figure P. 33.	Example application: Digital Construction Fence.....	209
Figure P. 34.	Examples of tracked off-highway vehicles studied within the THEIAXR project	212
Figure P. 35.	CAD design of the first prototype, showing the vibrotactile actuators.....	212
Figure P. 36.	Concepts for tactile feedback assistance for navigation.....	213
Figure P. 37.	Concept for torque feedback navigation assistance.....	214
Figure P. 38.	Frameworks for ethical impact assessment of AI in XR applications . Error! Bookmark not defined.	
Figure P. 39.	A room’s mesh is obtained by scanning using Microsoft HoloLens 2.	228
Figure P. 40.	Example of a task in experiment 1.....	229
Figure P. 41.	Mean of overall user ranking of the three boundary visualisation techniques	230
Figure P. 42.	Functional diagram.....	232
Figure P. 43.	Physical prototype on a human hand.....	233
Figure P. 44.	Smartphone MR interface	234
Figure P. 45.	PAX Methodology	238
Figure P. 46.	The “More Than One Milan” olfactory display prototype.	240
Figure P. 47.	User is testing the “More Than One Milan” olfactory display.	241
Figure P. 48.	More Than One Milan” app leveraging Augmented Reality.	241
Figure P. 49.	Main Menu including scene selection and tutorial area	251
Figure P. 50.	Table model with lighting adjustment	252
Figure P. 51.	GPS model spectating new Buildings on site.	252

Figures in Demos

Figure D. 1.	Basic structure of the environment with Creation Space and Launch Room... ..	258
Figure D. 2.	User interface of the Pneumatics-Case system showcasing the AR environment.	262
Figure D. 3.	Block diagram of system architecture for the Pneumatics-Case.....	264
Figure D. 4.	The simulator application running on VTT’s Powerwall.....	268
Figure D. 5.	Camera system’s preview images and their docking positions.	269
Figure D. 6.	AR guidance system elements	269
Figure D. 7.	When the user has to lock a container	270
Figure D. 8.	The user aligns the spreader’s twist locks to the container’s corner	270
Figure D. 9.	The user can “pan” through the 360° view.....	271
Figure D. 10.	Different views of the scene	274
Figure D. 11.	Example of the XR technician view.	275
Figure D. 12.	Overview of the XR4ED Intelligent XR Tutor.....	280
Figure D. 13.	Top-down view of the entire virtual environment and different task zones.....	282
Figure D. 14.	First-person perspective screenshots of each challenge.	283
Figure D. 15.	Target location mapping on motor array related to virtual position.	283
Figure D. 16.	Example haptic feedback intensities and ranges relative to user height.	283
Figure D. 17.	Mean completion times per challenge.	284
Figure D. 18.	Our augmented reality framework.....	288
Figure D. 19.	Architecture of the framework	289
Figure D. 20.	Images showing the Recyclage VR application.....	292
Figure D. 21.	Summary of proposed methodology.....	294
Figure D. 22.	Menu screen.....	298
Figure D. 23.	CD2 Task performance.....	299
Figure D. 24.	Snapshot of an IMD experience on UnityMol-Lite.	302
Figure D. 25.	Composite of the user interfaces (UI) for UnityMol-Lite.	303

Lists of Tables

Tables in Application Track

Table 1.	Descriptive Statistics.....	51
Table 2.	NASA-TLX/SUS Assessment.	70
Table 3.	Completion time in different positions.....	70
Table 4.	Centroids of the groups for each eye-tracking metric.....	80
Table 5.	Use Cases distribution.....	88
Table 6.	Concepts and terms of the draft XR taxonomy.....	92
Table 7.	Comparison of Dietary Groups Across Various Categories.....	98
Table 8.	Results of Number of Gazes.....	101
Table 9.	Number of User clicks.....	102
Table 10.	Summary of co-designed and operator-approved XR applications.....	115

Tables in Posters

Table P. 1.	Performance Comparison between Python-Only implementation (V1.0) and Unity+Python implementation (V2.0) (Latency calculated in seconds).....	158
Table P. 2.	Modules for parameter configuration and implementation of related real-time functionalities.....	165
Table P. 3.	Spearman's correlations between synchrony, body sway and cybersickness.....	177
Table P. 4.	Comparison of different WebAR frameworks.....	209

Tables in Demos

None

Title	EuroXR 2024 Proceedings of the 21st EuroXR International Conference
Author(s)	Kaj Helin, Despina Michael-Grigoriou, Tina Katika, Barbara Schiavi & Electra Tsaknaki (eds.)
Abstract	<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 2024, the 21th annual EuroXR conference, being hosted by ICCS (Institute of Communication and Computer System) from November 27th to November 29th 2024, in Athens, Greece. 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-association.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 978-951-38-8796-4 ISSN-L 2242-1211 ISSN 2242-122X (Online) DOI: 10.32040/2242-122X.2024.T432
Date	November 2024
Language	English
Pages	319 p.
Name of the project	
Commissioned by	
Keywords	
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 2024

Proceedings of the 21st EuroXR International Conference

The 21st EuroXR International Conference - EuroXR 2024 will be held on 27 to 29 November, 2024, organized by ICCS in Athens, Greece.

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), and Rotterdam (2023).

EuroXR 2024 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 978-951-38-8796-4
ISSN-L 2242-1211
ISSN 2242-122X (Online)
DOI: 10.32040/2242-122X.2024.T432



beyond the obvious