Jessie Y. C. Chen
Gino Fragomeni (Eds.)

LNCS 15790

Virtual, Augmented and Mixed Reality

17th International Conference, VAMR 2025
Held as Part of the 27th HCI International Conference, HCII 2025
Gothenburg, Sweden, June 22–27, 2025
Proceedings, Part III

3 Part III

HCI 2025
INTERNATIONAL

Springer

Lecture Notes in Computer Science **15790**

The series Lecture Notes in Computer Science (LNCS), including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI), has established itself as a medium for the publication of new developments in computer science and information technology research, teaching, and education.

LNCS enjoys close cooperation with the computer science R & D community, the series counts many renowned academics among its volume editors and paper authors, and collaborates with prestigious societies. Its mission is to serve this international community by providing an invaluable service, mainly focused on the publication of conference and workshop proceedings and postproceedings. LNCS commenced publication in 1973.

Jessie Y. C. Chen · Gino Fragomeni
Editors

Virtual, Augmented and Mixed Reality

17th International Conference, VAMR 2025
Held as Part of the 27th HCI International Conference, HCII 2025
Gothenburg, Sweden, June 22–27, 2025
Proceedings, Part III

Springer

Editors
Jessie Y. C. Chen
U.S. Army Research Laboratory
Adelphi, MD, USA

Gino Fragomeni
U.S. Army Combat Capabilities
Development Command Soldier Center
Orlando, FL, USA

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-031-93714-9 ISBN 978-3-031-93715-6 (eBook)
https://doi.org/10.1007/978-3-031-93715-6

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

The HCI International (HCII) conference was founded in 1984 by Gavriel Salvendy (Purdue University, USA, Tsinghua University, P.R. China, and University of Central Florida, USA) and the first event of the series, "1st USA-Japan Conference on Human-Computer Interaction", was held in Honolulu, Hawaii, USA, 18–20 August. Since then, HCI International is held jointly with several Thematic Areas and Affiliated Conferences, with each one under the auspices of a distinguished international Program Board and under one management and one registration. Twenty-seven HCI International Conferences have been organized so far (every two years until 2013, and annually thereafter).

Last year, we celebrated 40 years since the establishment of the HCII conference, which has been a hub for presenting groundbreaking research and novel ideas and collaboration for people from all over the world. Over the years, this conference has served as a platform for scholars, researchers, industry experts, and students to exchange ideas, connect, and address challenges in the ever-evolving HCI field. The conference has evolved itself, adapting to new technologies and emerging trends, while staying committed to its core mission of advancing knowledge and driving change.

The 27th International Conference on Human-Computer Interaction, HCI International 2025 (HCII 2025), was held as an 'on-site' conference at the Gothia Towers Hotel and Swedish Exhibition & Congress Centre, in Gothenburg, Sweden, on June 22–27, 2025, with the additional option for 'on-line' participation. It incorporated the 21 thematic areas and affiliated conferences listed below.

A total of 7972 individuals from academia, research institutes, industry, and government agencies from 92 countries submitted contributions. 1430 papers and 355 posters (as short research papers) are included in the volumes of the proceedings published just before the start of the conference, and which are listed below. The contributions thoroughly cover the entire field of human-computer interaction, highlight the evolving role of computers in diverse contexts, and demonstrate how HCI research is shaping and improving user experiences across a wide range of domains, influencing technological progress and its effective integration into various sectors.

The HCII conference also offers the option of presenting 'Late Breaking Work', both for papers and posters, with the corresponding proceedings volumes published after the conference. Full papers are included in the 'HCII 2025 - Late Breaking Papers' volumes of the proceedings published in the Springer LNCS series, while 'Poster Extended Abstracts' are included as short research papers in the 'HCII 2025 - Late Breaking Posters' volumes published in the Springer CCIS series.

I would like to thank the Program Board Chairs and the members of the Program Boards of all thematic areas and affiliated conferences for their contribution towards the high scientific quality and overall success of the HCI International 2025 conference. Their manifold support including paper reviews (via a single-blind review process, with a minimum of two reviews per submission), session organization, and their willingness to act as goodwill ambassadors for the conference is most highly appreciated.

This conference would not have been possible without the continuous and unwavering support and advice of Gavriel Salvendy, founder, General Chair Emeritus, and Scientific Advisor. For his outstanding efforts, I would like to express my sincere appreciation to Abbas Moallem, Communications Chair and Editor of HCI International News.

June 2025 Constantine Stephanidis

HCI International 2025 Thematic Areas and Affiliated Conferences

- HCI: Human-Computer Interaction Thematic Area
- HIMI: Human Interface and the Management of Information Thematic Area
- EPCE: 22nd International Conference on Engineering Psychology and Cognitive Ergonomics
- AC: 19th International Conference on Augmented Cognition
- UAHCI: 19th International Conference on Universal Access in Human-Computer Interaction
- CCD: 17th International Conference on Cross-Cultural Design
- SCSM: 17th International Conference on Social Computing and Social Media
- VAMR: 17th International Conference on Virtual, Augmented and Mixed Reality
- DHM: 16th International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management
- DUXU: 14th International Conference on Design, User Experience, and Usability
- C&C: 13th International Conference on Culture and Computing
- DAPI: 13th International Conference on Distributed, Ambient and Pervasive Interactions
- HCIBGO: 12th International Conference on HCI in Business, Government and Organizations
- LCT: 12th International Conference on Learning and Collaboration Technologies
- ITAP: 11th International Conference on Human Aspects of IT for the Aged Population
- AIS: 7th International Conference on Adaptive Instructional Systems
- HCI-CPT: 7th International Conference on HCI for Cybersecurity, Privacy and Trust
- HCI-Games: 7th International Conference on HCI in Games
- MobiTAS: 7th International Conference on HCI in Mobility, Transport and Automotive Systems
- AI-HCI: 6th International Conference on Artificial Intelligence in HCI
- MOBILE: 6th International Conference on Human-Centered Design, Operation and Evaluation of Mobile Communications

List of Conference Proceedings Volumes Appearing Before the Conference

1. LNCS 15766, Human-Computer Interaction - Part I, edited by Masaaki Kurosu and Ayako Hashizume
2. LNCS 15767, Human-Computer Interaction - Part II, edited by Masaaki Kurosu and Ayako Hashizume
3. LNCS 15768, Human-Computer Interaction - Part III, edited by Masaaki Kurosu and Ayako Hashizume
4. LNCS 15769, Human-Computer Interaction - Part IV, edited by Masaaki Kurosu and Ayako Hashizume
5. LNCS 15770, Human-Computer Interaction - Part V, edited by Masaaki Kurosu and Ayako Hashizume
6. LNCS 15771, Human-Computer Interaction - Part VI, edited by Masaaki Kurosu and Ayako Hashizume
7. LNCS 15772, Human-Computer Interaction - Part VII, edited by Masaaki Kurosu and Ayako Hashizume
8. LNCS 15773, Human Interface and the Management of Information: Part I, edited by Hirohiko Mori and Yumi Asahi
9. LNCS 15774, Human Interface and the Management of Information: Part II, edited by Hirohiko Mori and Yumi Asahi
10. LNCS 15775, Human Interface and the Management of Information: Part III, edited by Hirohiko Mori and Yumi Asahi
11. LNAI 15776, Engineering Psychology and Cognitive Ergonomics: Part I, edited by Don Harris and Wen-Chin Li
12. LNAI 15777, Engineering Psychology and Cognitive Ergonomics: Part II, edited by Don Harris and Wen-Chin Li
13. LNAI 15778, Augmented Cognition: Part I, edited by Dylan D. Schmorrow and Cali M. Fidopiastis
14. LNAI 15779, Augmented Cognition: Part II, edited by Dylan D. Schmorrow and Cali M. Fidopiastis
15. LNCS 15780, Universal Access in Human-Computer Interaction: Part I, edited by Margherita Antona and Constantine Stephanidis
16. LNCS 15781, Universal Access in Human-Computer Interaction: Part II, edited by Margherita Antona and Constantine Stephanidis
17. LNCS 15782, Cross-Cultural Design: Part I, edited by Pei-Luen Patrick Rau
18. LNCS 15783, Cross-Cultural Design: Part II, edited by Pei-Luen Patrick Rau
19. LNCS 15784, Cross-Cultural Design: Part III, edited by Pei-Luen Patrick Rau
20. LNCS 15785, Cross-Cultural Design: Part IV, edited by Pei-Luen Patrick Rau

21. LNCS 15786, Social Computing and Social Media: Part I, edited by Adela Coman and Simona Vasilache
22. LNCS 15787, Social Computing and Social Media: Part II, edited by Adela Coman and Simona Vasilache
23. LNCS 15788, Virtual, Augmented and Mixed Reality: Part I, edited by Jessie Y. C. Chen and Gino Fragomeni
24. LNCS 15789, Virtual, Augmented and Mixed Reality: Part II, edited by Jessie Y. C. Chen and Gino Fragomeni
25. LNCS 15790, Virtual, Augmented and Mixed Reality: Part III, edited by Jessie Y. C. Chen and Gino Fragomeni
26. LNCS 15791, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Part I, edited by Vincent G. Duffy
27. LNCS 15792, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Part II, edited by Vincent G. Duffy
28. LNCS 15793, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Part III, edited by Vincent G. Duffy
29. LNCS 15794, Design, User Experience, and Usability: Part I, edited by Martin Schrepp
30. LNCS 15795, Design, User Experience, and Usability: Part II, edited by Martin Schrepp
31. LNCS 15796, Design, User Experience, and Usability: Part III, edited by Martin Schrepp
32. LNCS 15797, Design, User Experience, and Usability: Part IV, edited by Martin Schrepp
33. LNCS 15798, Design, User Experience, and Usability: Part V, edited by Martin Schrepp
34. LNCS 15799, Design, User Experience, and Usability: Part VI, edited by Martin Schrepp
35. LNCS 15800, Culture and Computing: Part I, edited by Matthias Rauterberg
36. LNCS 15801, Culture and Computing: Part II, edited by Matthias Rauterberg
37. LNCS 15802, Distributed, Ambient and Pervasive Interactions: Part I, edited by Norbert A. Streitz and Shin'ichi Konomi
38. LNCS 15803, Distributed, Ambient and Pervasive Interactions: Part II, edited by Norbert A. Streitz and Shin'ichi Konomi
39. LNCS 15804, HCI in Business, Government and Organizations: Part I, edited by Fiona Fui-Hoon Nah and Keng Leng Siau
40. LNCS 15805, HCI in Business, Government and Organizations: Part II, edited by Fiona Fui-Hoon Nah and Keng Leng Siau
41. LNCS 15806, Learning and Collaboration Technologies: Part I, edited by Brian K. Smith and Marcela Borge
42. LNCS 15807, Learning and Collaboration Technologies: Part II, edited by Brian K. Smith and Marcela Borge
43. LNCS 15808, Learning and Collaboration Technologies: Part III, edited by Brian K. Smith and Marcela Borge

44. LNCS 15809, Human Aspects of IT for the Aged Population: Part I, edited by Qin Gao and Jia Zhou
45. LNCS 15810, Human Aspects of IT for the Aged Population: Part II, edited by Qin Gao and Jia Zhou
46. LNCS 15811, Human Aspects of IT for the Aged Population: Part III, edited by Qin Gao and Jia Zhou
47. LNCS 15812, Adaptive Instructional System: Part I, edited by Robert A. Sottilare and Jessica Schwarz
48. LNCS 15813, Adaptive Instructional System: Part II, edited by Robert A. Sottilare and Jessica Schwarz
49. LNCS 15814, HCI for Cybersecurity, Privacy and Trust: Part I, edited by Abbas Moallem
50. LNCS 15815, HCI for Cybersecurity, Privacy and Trust: Part II, edited by Abbas Moallem
51. LNCS 15816, HCI in Games, edited by Xiaowen Fang
52. LNCS 15817, HCI in Mobility, Transport and Automotive Systems: Part I, edited by Heidi Krömker
53. LNCS 15818, HCI in Mobility, Transport and Automotive Systems: Part II, edited by Heidi Krömker
54. LNAI 15819, Artificial Intelligence in HCI: Part I, edited by Helmut Degen and Stavroula Ntoa
55. LNAI 15820, Artificial Intelligence in HCI: Part II, edited by Helmut Degen and Stavroula Ntoa
56. LNAI 15821, Artificial Intelligence in HCI: Part III, edited by Helmut Degen and Stavroula Ntoa
57. LNAI 15822, Artificial Intelligence in HCI: Part IV, edited by Helmut Degen and Stavroula Ntoa
58. LNCS 15823, Human-Centered Design, Operation and Evaluation of Mobile Communications: Part I, edited by June Wei and George Margetis
59. LNCS 15824, Human-Centered Design, Operation and Evaluation of Mobile Communications: Part II, edited by June Wei and George Margetis
60. CCIS 2522, HCI International 2025 Posters - Part I, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy
61. CCIS 2523, HCI International 2025 Posters - Part II, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy
62. CCIS 2524, HCI International 2025 Posters - Part III, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy
63. CCIS 2525, HCI International 2025 Posters - Part IV, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy
64. CCIS 2526, HCI International 2025 Posters - Part V, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy
65. CCIS 2527, HCI International 2025 Posters - Part VI, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy

66. CCIS 2528, HCI International 2025 Posters - Part VII, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy
67. CCIS 2529, HCI International 2025 Posters - Part VIII, edited by Constantine Stephanidis, Margherita Antona, Stavroula Ntoa and Gavriel Salvendy

https://2025.hci.international/proceedings

Preface

With the recent emergence of a new generation of displays, smart devices, and wearables, the field of virtual, augmented, and mixed reality (VAMR) is rapidly expanding, transforming, and moving towards the mainstream market. At the same time, VAMR applications in a variety of domains are also reaching maturity and practical usage. From the point of view of user experience, VAMR promises possibilities to reduce interaction efforts and cognitive load, while also offering contextualized information, by combining different sources and reducing attention shifts, and opening the 3D space. Such scenarios offer exciting challenges associated with underlying and supporting technologies, interaction, and navigation in virtual and augmented environments, and design and development. VAMR themes encompass a wide range of areas such as education, aviation, social, emotional, psychological, and persuasive applications.

The 17th International Conference on Virtual, Augmented and Mixed Reality (VAMR 2025), an affiliated conference of the HCI International Conference, provided a forum for researchers and practitioners to disseminate and exchange scientific and technical information on VAMR-related topics in various applications. A considerable number of papers have explored user experience topics including avatar design, walking and moving in VR environments, immersive environments, multimodality and multisensory feedback. A key topic that emerged was interaction in immersive environments such as haptic interaction, tangible VR, and gestures. Furthermore, emphasis was given to the application domains of VAMR including collaboration, cultural heritage, education and learning, health and well-being, medicine, and games. We are thrilled to present this compilation of VAMR submissions encompassing a wide range of topics and exploring the current state of the art, while also highlighting future avenues in the design and development of immersive experiences.

Three volumes of the HCII 2025 proceedings are dedicated to this year's edition of the VAMR conference focusing on topics related to

- Virtual, Augmented and Mixed Reality - Part I: Designing and Developing Virtual Environments; and UX in Virtual Environments
- Virtual, Augmented and Mixed Reality - Part II: VR, Culture, Art and Entertainment; and Social Interaction and Wellbeing in Virtual Environments
- Virtual, Augmented and Mixed Reality - Part III: VR Games; Virtual Environments for Learning, Training and Professional Development; and Multimodal Interaction in Virtual Environments

The papers in these volumes were accepted for publication after a minimum of two single-blind reviews from the members of the VAMR Program Board or, in some cases, from members of the Program Boards of other affiliated conferences. We would like to thank all of them for their invaluable contribution, support, and efforts.

June 2025

Jessie Y. C. Chen
Gino Fragomeni

17th International Conference on Virtual, Augmented and Mixed Reality (VAMR 2025)

The full list with the Program Board Chairs and the members of the Program Boards of all thematic areas and affiliated conferences of HCII 2025 is available online at:

http://www.hci.international/board-members-2025.php

HCI International 2026 Conference

The 28th International Conference on Human-Computer Interaction, HCI International 2026, will be held jointly with the affiliated conferences at the Montréal Convention Centre (Palais des congrès de Montréal), in Montreal, Canada, 26–31 July 2026. It will cover a broad spectrum of themes related to Human-Computer Interaction, including theoretical issues, methods, tools, processes, and case studies in HCI design, as well as novel interaction techniques, interfaces, and applications. The proceedings will be published by Springer (part of Springer Nature) in a multi-volume set. More information will become available on the conference website: https://2026.hci.international/.

General Chair
Prof. Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
Email: general_chair@2026.hci.international

https://2026.hci.international/

Contents – Part III

Multimodal Interaction in Virtual Environments

VR Games

The Impact of Wind Experience on VR Game Immersion

Yung-Ting Chen(✉) and Meng-Shiuan Tsai

Department of Industrial Design, National Kaohsiung Normal University, Kaohsiung, Taiwan
{t4854,611272004}@mail.nknu.edu.tw

Abstract. Modern virtual reality (VR) devices primarily focus on visual sensory experiences while often underestimating the importance of haptic feedback in gaming. The lack of cross-sensory stimulation limits the overall VR experience. This study introduces wind-based haptic feedback in a VR environment by directing airflow to different parts of participants' bodies to examine how varying wind exposure affects their gaming experience. A total of 25 participants from National Kaohsiung Normal University took part in a VR motorcycle game under five different wind conditions. After each session, participants completed the Immersive Virtual Environment Experience Questionnaire, and their responses were analyzed using a one-way repeated measures ANOVA, along with their game performance data. The results revealed that presence, emotion, and game performance (number of mistakes) were significantly influenced by the wind-based haptic feedback. Specifically, wind stimulation notably enhanced the sense of presence, with the chest wind condition demonstrating the most substantial effect. Additionally, emotional scores significantly improved, indicating that wind stimulation enhances player enjoyment of the game environment. Regarding game performance, participants in the **head & neck wind condition** made significantly more mistakes than those in the **shins wind condition**, implying that wind directed at the head and neck may disrupt concentration and reduce control accuracy. The findings of this study have practical implications for controlled VR environments, such as arcade booths, exhibitions, and home settings by enhancing user satisfaction and immersion.

Keywords: Virtual Reality · Haptics · Game Immersion · Human-Computer Interaction

1 Introduction

Virtual Reality (VR) has garnered significant attention in both the entertainment and workplace sectors, with major technology companies competing to release VR-related products, such as Meta's Quest series, Apple's Vision Pro, and HTC's Vive series. According to a report by Global Information, Inc. (GII, 2024), the VR market is projected to grow from \$67.66 billion in 2024 to \$204.35 billion by 2029 at a compound annual growth rate (CAGR) of 24.74%, highlighting the increasing significance of VR technology in the future. Additionally, a report by Goldman Sachs (2016) estimated

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 3–17, 2025.
https://doi.org/10.1007/978-3-031-93715-6_1

that, by 2025, video games would account for 33% of the AR/VR application software market, making it the largest sector. This projection highlights the pivotal role of virtual gaming in shaping the industry's future development.

Currently, VR headsets primarily focus on enhancing visual experiences, while haptic feedback remains relatively underexplored. However, research indicates that users prefer gaming experiences that incorporate haptic feedback over those that completely lack tactile stimulation (Shen et al., 2022), underscoring the crucial role of haptic stimuli in enhancing immersion and realism. To address this, researchers have developed various haptic feedback devices to enhance VR immersion, including ultrasonic haptic feedback (Shen et al., 2022), liquid-based sensation (Peiris et al., 2018), suction-based haptics (Kameoka and Kajimoto, 2021), wind and temperature simulation (Ranasinghe et al., 2018), haptic gloves (Perret & Vander Poorten, 2018), and finger-string haptic devices (Fang et al., 2020). These devices can either be integrated with VR headsets or used as wearable accessories to provide targeted tactile stimulation to specific body areas. Although these technologies enhance the VR experience, they may also present challenges such as added weight, discomfort, and increased user burden, which could negatively impact gameplay experience and user adoption.

To address these challenges, this study develops an external wind-based haptic system that can be integrated with VR gaming to enhance immersion through wind stimuli. The device is positioned around the user and delivers targeted wind stimulation to different body parts to investigate its effects on game immersion and performance. The experimental setup dynamically adjusts the fan speed based on in-game parameters, controlling the intensity and distribution of wind stimuli to modulate wind haptic feedback. Designed as a low-cost and low-burden system, this approach effectively enhances haptic sensations in VR gaming while maintaining user comfort.

2 Review of Literature

2.1 Tactile Spatial Sensitivity

Tactile spatial sensitivity is commonly used in neurological research to assess the functionality of the dorsal column system. Mancini et al. (2014) conducted a Two-Point Threshold (TPT) test, in which two stimuli were applied either simultaneously or sequentially, and participants were asked to determine whether the stimuli originated from the same or different locations. This method was used to measure tactile spatial sensitivity across different regions of the human body.

Their study mapped tactile spatial acuity across multiple body parts, including the forehead, shoulders, forearms, back of the hand, palm, fingertips, back, mid-thigh, lower leg, dorsum of the foot, and sole (Fig. 1). The findings revealed that fingertips exhibited the highest tactile and pain spatial sensitivity, while the palm (hairless skin area) showed a consistent gradient in both tactile and pain sensitivity. However, in hairy skin areas of the upper limbs, the sensitivity gradients for touch and pain followed an inverse trend, which correlates with the distribution density of nerve endings in those regions. Additionally, tactile spatial sensitivity primarily relies on primary afferent tactile nerves, whereas pain spatial sensitivity does not solely depend on these nerves.

Given the significant variations in spatial sensitivity across different body regions, selecting appropriate test areas is crucial for this study. We first referred to the sensitivity mapping proposed by Mancini et al. (2014) and aligned our selection with the context of our VR motorcycle racing game. Since wind exposure areas during motorcycle riding are relatively distinct, we identified three primary evaluation regions: Hands (combined assessment of the palm and fingers), Head & Neck (including the forehead, classified as part of this region), Legs (specifically the front of the lower legs) (Fig. 2). Moreover, Mancini et al. (2014) did not include the chest area in their sensitivity study. However, since the chest is a primary wind-exposed region during motorcycle riding, we incorporated it into our research scope.

By analyzing the variations in wind haptic feedback across these body regions, this study aims to further evaluate VR game immersion and optimize VR game design to enhance realism and user immersion.

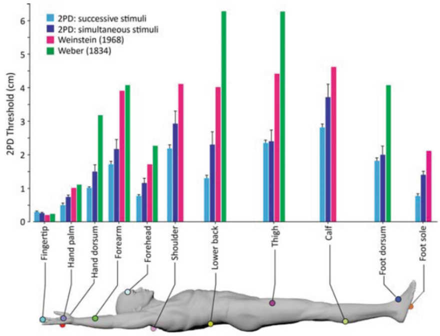


Fig. 1. Two-point discrimination thresholds for skin perception in various parts of the human body (Mancini et al., 2014).

2.2 Immersive Virtual Environment Experience Scale

In the field of virtual reality (VR) technology and immersive virtual environments, evaluating user experience has become crucial. Tcha-Tokey et al. (2016) proposed and validated an immersive virtual environment questionnaire (IVEQ) to assess user experience in immersive settings. This questionnaire encompasses ten dimensions, which are described as follows:

1. Presence: The extent to which users feel "physically present" in the virtual environment, as if they are truly inside it.
2. Engagement: The degree of user involvement, including their level of attention and sustained interest in the VR environment.
3. Immersion: The feeling of being fully surrounded by and absorbed in the virtual experience.
4. Flow: The perceived smoothness of interaction within the VR environment and the balance between challenge and skill.

5. Usability: The ease of use of the VR system, including the user-friendliness of the interface and the convenience of controls.
6. Emotion: The emotional responses elicited during the VR experience, such as joy, excitement, or frustration.
7. Skill: The user's ability and proficiency in completing tasks within the virtual environment.
8. Judgment: The ability to evaluate and interpret events and situations in the virtual setting.
9. Experience Consequence: The impact of the virtual experience on the user's real-world thoughts or reflections.
10. Technology Adoption: The user's willingness to accept and integrate VR technology into their daily activities.

Since this study primarily focuses on gaming experiences, the technology adoption dimension was excluded, while the remaining nine dimensions were retained to assess the immersive virtual environment experience. These dimensions not only provide a structured evaluation of user experience in immersive VR settings but also assist researchers and developers in refining virtual environment designs to enhance user engagement and satisfaction.

3 Research Method

3.1 Participants

This study investigates the impact of haptic stimulation on the VR gaming experience, specifically examining whether wind stimuli applied to different body regions influence participants' immersion and game performance. A total of 25 participants (13 males, 12 females) were recruited from National Kaohsiung Normal University, aged 19–31 years (M = 22.68, SD = 3.01). All participants had normal tactile perception abilities. Due to considerations of wind conditions and exposed body areas, participants were required to wear short-sleeved shirts and shorts (allowing their arms and lower legs to be exposed). They experienced a VR motorcycle game in a closed classroom environment and completed an experience evaluation after gameplay.

3.2 Experimental Design

The independent variable in this study was the wind haptic condition, which consisted of five variations: no-wind, head & neck wind, chest wind, hands wind, and shins wind (Fig. 2). The dependent variables included nine dimensions from the IVEQ, adapted from Tcha-Tokey et al. (2016), along with game performance metrics, such as game score and number of mistakes. The questionnaire was created using Google Forms and utilized a five-point Likert scale, except for the Judgment dimension, which was assessed using a bipolar adjective scale (Table 1).

To minimize order effects, participants were randomly divided into five groups of five and experienced the five wind conditions in different sequences, following a Latin Square Design (Table 2). After each VR session, participants completed the questionnaire to evaluate their experience.

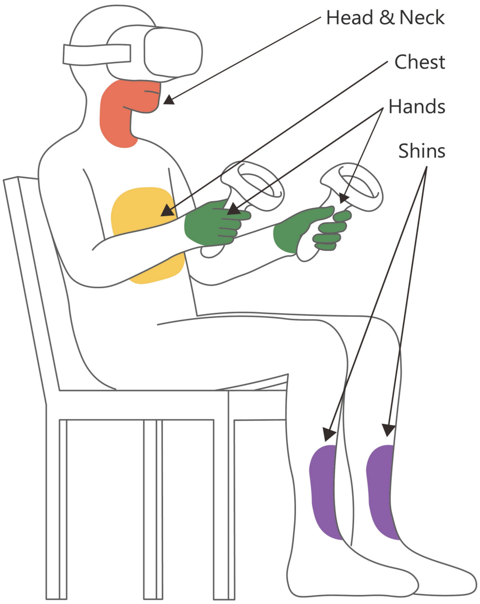


Fig. 2. Four types of wind experience areas except "no-wind" (Horizontal blowing on the front side of the body).

Table 1. Questionnaire Items.

Experience Dimension	Items	Rating Scale
Presence	My interaction with the game felt natural	1 (Low) – 5 (High)
Engagement	The movement within the game was engaging	
Immersion	I felt completely absorbed, as if inside the game rather than just controlling it	
Flow	I felt I could control my actions perfectly	
Usability	I found the VR headset and controller easy to use	
Emotion	I enjoyed being in this game	
Skill	I felt confident navigating the game with the controller	
Judgment	I found this game boring/exciting	
Experience Consequence	I felt fatigued after interacting with the game	

Table 2. Latin square design

Group	Order of Experiences (from left to right)				
1	No-wind	Head & Neck	Chest	Hands	Shins
2	Head & Neck	Chest	Hands	Shins	No-wind
3	Chest	Hands	Shins	No-wind	Head & Neck
4	Hands	Shins	No-wind	Head & Neck	Chest
5	Shins	No-wind	Head & Neck	Chest	Hands

3.3 Experimental Equipment

The experiment was conducted in a standardized classroom environment to ensure consistency in participant experiences. The setup included a VR headset, a gaming laptop, a custom-developed VR motorcycle game, and a custom-built fan system.

Participants wore the VR headset and controlled the motorcycle using a joystick on the right-hand controller, while holding the left-hand controller (non-functional) to simulate the experience of gripping handlebars during real motorcycle riding. In addition to viewing environmental changes through the VR display, participants experienced wind stimulation on different body regions.

The wind intensity dynamically adjusted based on the motorcycle's in-game speed, providing real-time wind haptic feedback. The specifications of the experimental equipment are detailed as follows.

1. Virtual Reality Device and Laptop

The VR system and laptop used in the experiment included the HTC Vive Focus 3 as the VR headset, connected to a computer via a Type-C interface to ensure stable gameplay performance and high-quality visuals. The laptop used was an MSI Katana 15 B13VGK, featuring a 13th Gen Intel® Core™ i9-13900H processor and 16 GB RAM, providing sufficient computing power to support VR rendering and experimental needs.

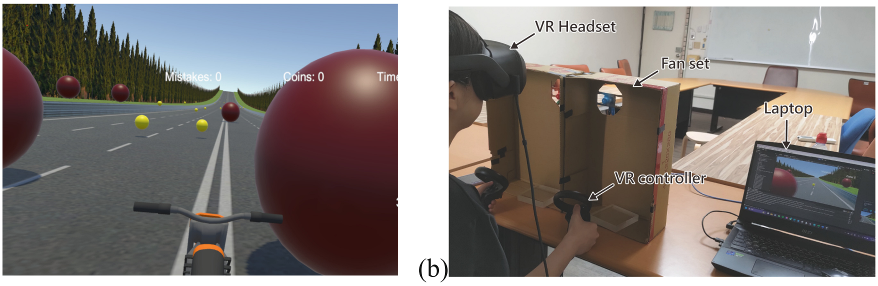
2. VR Motorcycle Game

The VR motorcycle game was developed using Unity 2022.3.22f1, incorporating the Motorbike Physics Tool (2021) from the Unity Asset Store to simulate realistic motorcycle physics and Kajaman's Roads – Free (2018) for track design. The game interface displayed remaining time, collected coins (score), and mistake counts, with randomly falling obstacles (red spheres) and collectible coins (yellow spheres) as interactive elements (Fig. 2a). The maximum motorcycle speed was set at 60 km/h, and players control acceleration, braking, and steering via the VR joystick. When colliding with obstacles, the interface displayed the mistake count, accompanied by an error sound. When collecting coins , the system displayed the updated coin count and triggered a positive sound effect. The game lasted 60 s per round, with the screen displaying the remaining time. After each round, the game automatically ended, and movements were paused.

3. Fan Equipment

The wind simulation system was developed using Arduino and L9110S fan modules, simulating real-time wind variations based on motorcycle speed. The system retrieved real-time speed data from Unity and transmitted it to Arduino via USB, where a PWM signal was converted into voltage control for fan speed. The fan modules were powered by a 9V DC source, enabling the motor to operate and adjust wind intensity according to the control signal. The speed mapping corresponded to the motorcycle's in-game speed, ranging from 0 to 60 km/h, which was converted into a PWM range of 90–255 in Arduino. If the calculated signal was below 90, the fan motor would not start; at 255, it reached maximum speed. This ensured a smooth and proportional control of wind speed from activation to full intensity.

The experimental setup is illustrated in Fig. 2b, where the fans were mounted on cardboard structures, allowing manual adjustment of placement and angle to direct airflow to specific body regions. A demonstration of the experimental setup and procedures can be viewed at: https://youtu.be/5_1N80mNB98.



(a) (b)

Fig. 3. Experimental environment and equipment (a); VR Game Screenshots (b). (Color figure online)

3.4 Experimental Procedure

1. The researcher first set up the experimental environment and equipment, ensuring that the room temperature remained stable at approximately 26 °C (78.8°F).
2. Upon arrival, participants were briefed on the experiment. The experiment proceeded only after obtaining informed consent from each participant.
3. The standardized experimental instructions provided to participants were as follows:

"*You will experience a VR motorcycle game with haptic wind feedback for five rounds, including no-wind, wind on the head & neck, chest, hands, and shins. After each round, you will complete a questionnaire. The entire session will take approximately 15 minutes, and you will receive a compensation of NT$300* (about

US$9). *Each round lasts for 1 minute, during which you will control the motorcycle using a VR joystick to avoid falling obstacles and collect coins. If you feel uncomfortable during the game, please close your eyes and immediately inform the staff to remove the VR headset, in which case you will receive NT$50 as compensation. If the screen flickers or malfunctions, close your eyes and notify the staff for adjustments.*"

4. After receiving the instructions, participants were assisted in adjusting the VR headset to ensure a comfortable fit. They then performed a practice session to familiarize themselves with the controls before proceeding to the main experiment.
5. After completing each round, participants filled out the Immersive Virtual Environment Experience Scale on a computer.
6. Steps 4 and 5 were repeated four more times until all five wind conditions were experienced.
7. Upon completion, the researcher thanked the participants, recorded their performance, and distributed the compensation, marking the end of the experiment.

3.5 Data Analysis

This study implemented five wind haptic conditions (independent variable) during the VR game sessions, collecting data on game performance (score and number of mistakes) and post-experience immersive virtual environment assessments (nine evaluation metrics). All collected data were aggregated and averaged, followed by a one-way repeated measures ANOVA to analyze differences across the five wind conditions. Additionally, the Sidak correction was applied for multiple comparisons to adjust the significance level and reduce the risk of Type I errors.

4 Results and Discussion

4.1 Results

This study examines the impact of wind haptic feedback applied to different body regions on the VR gaming experience. Table 3 summarizes the IVEQ scores and game performance metrics for 25 participants under five wind conditions during the VR motorcycle game.

A one-way repeated measures ANOVA was conducted to analyze the data in Table 3, with mean values calculated for each evaluation metric. The results are visualized in Fig. 4, which displays the scores for the nine evaluation metrics across the five wind conditions, with significant differences between groups indicated by connecting lines. The following section provides a detailed analysis of each metric.

1. Presence

The results of the within-subjects effect test indicate that wind haptic feedback had a significant effect on presence ($F(4, 96) = 2.902$, $p = .026$, Partial $\eta^2 = .108$). Pairwise comparisons showed that, compared to the no-wind condition (M = 3.68, SD = 1.069), all wind conditions significantly enhanced the sense of presence, with the following effect sizes:

- Chest wind (M = 4.20, SD = 0.816): Mean Difference (MD) = −0.520 (p = .009)
- Head & neck wind (M = 4.16, SD = 0.943): MD = −0.480 (p = .043)
- Hands wind (M = 4.08, SD = 1.038): MD = −0.400 (p = .038)
- Shins wind (M = 4.04, SD = 0.841): MD = −0.360 (p = .047)

These findings indicate that, regardless of the wind stimulation location, all wind conditions significantly enhanced the sense of presence compared to the no-wind condition (chest, head and neck, hands, shins > no-wind condition).

2. Engagement

The within-subjects effect test indicated that wind haptic feedback did not have a statistically significant effect on engagement (F(4, 96) = 2.134, p = .082, Partial η^2 = .082). However, pairwise comparisons showed that the shins wind condition (M = 4.28, SD = 0.678) resulted in significantly higher engagement compared to the no-wind condition (M = 3.88, SD = 0.781), with MD = 0.400 (p = .005). Additionally, the shins wind condition also had a significantly higher engagement rating than the chest wind condition (M = 4.00, SD = 0.764), with MD = 0.280 (p = .050). These findings suggest that, compared to the no-wind and chest wind conditions, the shins wind condition significantly enhanced engagement (shins > chest, no-wind condition).

3. Immersion

The within-subjects effect test indicated that wind haptic feedback did not have a significant effect on immersion (F(4, 96) = 0.595, p = .667, Partial η^2 = .024). Pairwise comparisons also revealed no significant differences in mean scores among the wind conditions, suggesting that wind stimulation had a limited impact on enhancing immersion.

4. Flow

The within-subjects effect test showed that wind haptic feedback did not significantly influence flow (F(4, 96) = 0.771, p = 547, Partial η^2 = 031). Pairwise comparisons further indicated no significant differences across the wind conditions, implying that wind stimulation had minimal impact on the perceived smoothness of gameplay.

5. Usability

Since Mauchly's test of sphericity was violated (p = 02), the Greenhouse-Geisser correction was applied to adjust the degrees of freedom. The results showed that wind haptic feedback had no significant effect on usability (F(2.759, 66.226) = 0.790, p = 495, Partial η^2 = 032), and pairwise comparisons revealed no significant differences among conditions. These findings suggest that wind stimulation did not significantly influence participants' usability ratings of the VR system.

6. Emotion

Since Mauchly's test of sphericity was violated (p = 004), the Greenhouse-Geisser correction was applied to adjust the degrees of freedom. The results showed that wind haptic feedback had a significant effect on emotion (F(2.523, 60.545) = 5.326, p = 004, Partial η^2 = 182).

Pairwise comparisons revealed that, compared to the no-wind condition (M = 3.92, SD = 0.909), the following three wind conditions significantly enhanced participants' emotional ratings, ranked in the following order:

- Chest wind (M = 4.44, SD = 0.583): MD = −0.520 (p = .001)
- Shins wind (M = 4.28, SD = 0.678): MD = −0.360 (p = .017)
- Head & neck wind (M = 4.28, SD = 0.792): MD = −0.360 (p = .026)

Additionally, the emotional rating for the chest wind condition was significantly higher than the other two conditions, with both comparisons reaching a significant level (vs. Head & neck wind, MD = 0.160, p = .043; vs. Shins wind, MD = 0.160, p = .043). (Chest wind > Head & neck wind, Shins wind > No-wind condition).

7. Skill

The within-subjects effect test indicated that wind haptic feedback did not significantly affect skill performance ($F(4, 96) = 0.690$, $p = .601$, Partial $\eta^2 = .028$). Pairwise comparisons also showed no significant differences between conditions, suggesting that wind stimulation applied to different body regions did not influence participants' skill execution in the VR game.

8. Judgment

Since Mauchly's test of sphericity was violated (p = .008), the Greenhouse-Geisser correction was applied. The results indicated that wind haptic feedback had no significant effect on judgment ($F(2.662, 63.899) = 1.101$, $p = .351$, Partial $\eta^2 = .044$). Pairwise comparisons also showed no significant differences among the wind conditions, suggesting that wind stimulation did not influence participants' subjective evaluation of game excitement or engagement.

9. Experience Consequence

Since Mauchly's test of sphericity was violated (p = .037), the Greenhouse-Geisser correction was applied. The results indicated that wind haptic feedback had no significant effect on experience consequence ($F(2.904, 69.685) = 1.093$, $p = .357$, Partial $\eta^2 = .044$).

Pairwise comparisons also revealed no significant differences among the wind conditions, suggesting that wind stimulation had a limited impact on participants' post-experience consequences, with no notable distinctions across conditions.

The study then conducted a one-way repeated measures ANOVA to analyze game performance, with the results described as follows:

1. Score

The within-subjects effect test indicated that wind haptic feedback had no significant effect on game score ($F(4, 96) = 0.412$, $p = .800$, Partial $\eta^2 = .017$). Pairwise comparisons also showed no significant differences between conditions, suggesting that wind stimulation did not impact participants' game scores.

2. Mistakes

Table 3. Descriptive Statistics

Item		No-wind	Head & Neck	Chest	Hands	Shins
Experience Dimension	**Presence (p = .004)**	M = 3.68 (SD = 1.069)	4.16 (0.943)	4.20 (0.816)	4.08 (1.038)	4.04 (0.841)
	Engagement	3.88 (0.781)	4.12 (0.833)	4.00 (0.764)	4.16 (0.800)	4.28 (0.678)
	Immersion	3.68 (0.988)	3.92 (0.862)	3.88 (0.881)	3.88 (1.054)	3.84 (1.028)
	Flow	3.48 (1.046)	3.28 (1.173)	3.40 (1.000)	3.52 (1.122)	3.56 (1.044)
	Usability	3.80 (1.041)	3.88 (1.130)	3.84 (1.068)	3.80 (1.190)	4.00 (1.041)
	Emotion (p = .026)	3.92 (0.909)	4.28 (0.792)	4.44 (0.583)	4.24 (0.723)	4.28 (0.678)
	Skill	3.68 (0.945)	3.60 (1.080)	3.52 (1.159)	3.76 (1.234)	3.72 (1.100)
	Judgment	3.72 (0.891)	3.88 (0.971)	3.84 (1.028)	4.12 (0.927)	4.00 (0.816)
	Experience Consequence	2.56 (1.261)	2.20 (1.291)	2.32 (1.180)	2.32 (1.249)	2.40 (1.155)
Game Performance	Score	10.44 (3.083)	10.72 (2.654)	10.40 (4.062)	9.80 (3.841)	10.16 (2.779)
	Mistakes	1.96 (2.835)	3.40 (3.069)	2.12 (2.242)	2.92 (2.914)	1.68 (2.249)

Since Mauchly's test of sphericity was violated ($p = .019$), the Greenhouse-Geisser correction was applied. The results showed that wind haptic feedback had a marginally significant effect on the number of mistakes ($F(2.669, 64.048) = 2.644, p = .063$, Partial $\eta^2 = .099$). Pairwise comparisons revealed that the head & neck wind condition (M = 3.40, SD = 3.069) resulted in significantly more mistakes compared to other conditions, ranked as follows:

- Chest wind (M = 2.12, SD = 2.242): MD = 1.280 ($p = .044$)
- No-wind condition (M = 1.96, SD = 2.835): MD = 1.440 ($p = .012$)
- Shins wind (M = 1.68, SD = 2.249): MD = 1.720 ($p = .002$)

These results suggest that wind stimulation directed at the head & neck significantly increased the number of mistakes compared to the no-wind, chest, and shins conditions, indicating that wind exposure in this area may disrupt focus and reduce control accuracy (head & neck > chest, no-wind, shins).

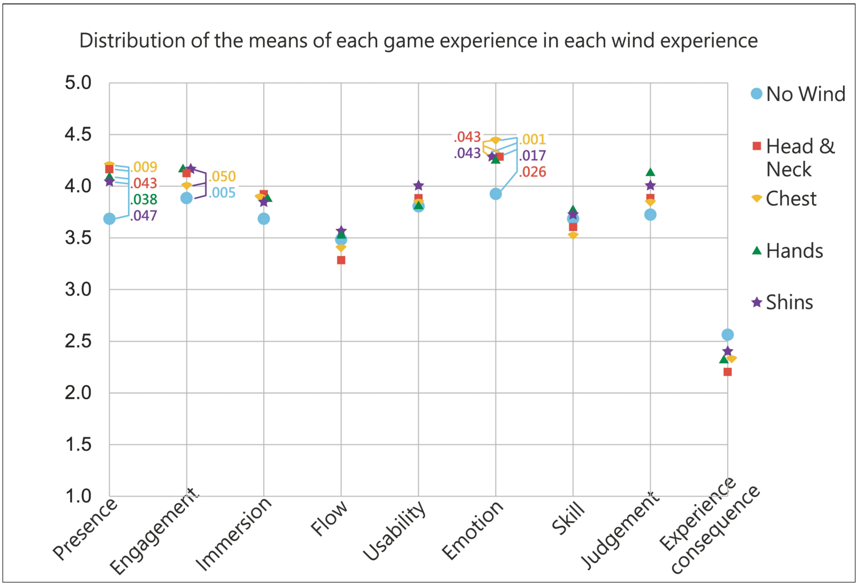


Fig. 4. Scatter plot of mean scores for the nine evaluation metrics across different wind conditions, with significant pairwise comparisons indicated by colored lines. (Color figure online)

4.2 Discussion

This study investigated the impact of wind haptic feedback applied to different body regions on the VR gaming experience. The results indicated that, among the nine immersive experience metrics, only presence and emotion were significantly affected by wind stimulation, while the other seven metrics showed no significant changes overall. The possible reasons for these findings are discussed below.

- **Presence**, defined as “My interaction with the game felt natural,” was significantly enhanced under head & neck, chest, hands, and shins wind conditions compared to the no-wind condition, with chest wind producing the highest scores. This can be attributed to the motorcycle racing context used in the experiment, where wind exposure is a natural and expected phenomenon during riding. Simulating wind through haptic feedback likely reinforced the sense of real-world presence for participants. Additionally, as discussed earlier, the chest was selected because it is a primary wind-exposed area during motorcycle riding, making it a key contributor to enhancing environmental consistency and presence perception.
- **Emotion**, defined as “I enjoyed being in this game,” was also significantly increased under head & neck, chest, and shins wind conditions compared to the no-wind condition, with chest wind again showing the strongest effect. This suggests that wind haptic feedback made the game experience more engaging and enjoyable by enhancing realism and natural sensory integration. Additionally, since the wind intensity in the experiment dynamically adjusted based on the motorcycle’s speed, it may

have further contributed to the game's excitement and emotional engagement. However, this study did not explicitly investigate the effect of varying wind intensity on emotional arousal and enjoyment, which could be explored in future research.

- Although wind haptic feedback did not significantly affect engagement at the statistical level, pairwise comparisons showed that the shins wind condition significantly enhanced engagement compared to the no-wind and chest wind conditions. This finding suggests that wind applied to the shins may enhance the perception of movement, allowing players to better sense their interaction with the virtual environment, thereby increasing their overall game involvement.

For the six metrics—immersion, flow, usability, skill, judgment, and experience consequence—wind haptic feedback did not produce significant effects in overall analysis or pairwise comparisons. This suggests that in motorcycle racing games, these aspects may be more influenced by other factors rather than variations in wind exposure across different body regions.

For example, flow, defined as "I feel I can perfectly control my actions," likely depends more on the game's responsiveness and operational efficiency rather than wind stimulation. Reducing latency and improving response speed could be more crucial in enhancing flow perception than adding haptic wind feedback. Similarly, usability (defined as "I find the VR headset and joystick easy to use") and skill (defined as "I feel confident using the joystick to navigate the game") are strongly tied to hardware design. Well-optimized ergonomic design can enhance wearability, comfort, and control ease, which directly improves the user experience.

Regarding game performance, the analysis showed that wind conditions did not significantly affect game scores but did have a notable impact on the number of mistakes. Specifically, the head & neck wind condition led to significantly more errors compared to the shins wind condition, suggesting that wind exposure in this area may cause greater distraction, thereby affecting gameplay accuracy. This effect is likely due to the higher sensory sensitivity of the head and neck region—when exposed to wind, it may disrupt the player's focus and stability, leading to an increased number of errors. Additionally, Fig. 4 illustrates that the head & neck wind condition resulted in the lowest flow scores, further indicating that this condition may negatively affect operational stability during gameplay.

5 Conclusion

This study explored the impact of wind haptic feedback applied to different body regions on the VR gaming experience by analyzing 25 participants' immersive virtual environment experience ratings across five wind conditions. The findings revealed significant differences in the effects of wind feedback on various gaming experience metrics. The main conclusions are as follows:

1. Wind haptic feedback significantly influenced presence and emotion.

Presence: Wind stimulation applied to the head & neck, chest, hands, and shins significantly enhanced presence compared to the no-wind condition, with chest wind having the strongest effect.

Emotion: Wind stimulation at the head & neck, chest, and shins significantly increased emotional ratings, with chest wind again demonstrating the most substantial effect.

2. Limited impact on game operation and skill execution.

Wind feedback did not show significant effects on engagement, immersion, flow, usability, skill, judgment, or experience consequence, suggesting that while wind feedback serves as an environmental reinforcement mechanism, its impact is primarily on player interaction and emotional engagement rather than operational fluency or skill performance.

3. Minimal impact on game performance, but head & neck wind may affect control accuracy.

Game scores were not significantly different across wind conditions. However, mistake counts were significantly higher in the head & neck wind condition compared to the shins wind condition. This may be due to the greater sensory sensitivity of the head & neck region, where wind exposure disrupts player focus and increases operational errors. These findings suggest that wind feedback could be used as an external condition to increase game difficulty, potentially serving as a game design element in future applications.

Based on the findings of this study, the following recommendations and considerations are proposed for future research:

1. Equipment and Wind Direction: The small motorized fans used in this study had limited wind speed, which was particularly noticeable for the chest wind condition, as clothing could obstruct airflow and reduce its effectiveness. Additionally, only two fans were used, lacking multi-directional wind control, and requiring manual adjustments, which affected the convenience of the setup.
2. Environmental Factors: The noise generated by the fan motors may have influenced experience ratings, particularly in quiet environments. Additionally, heat buildup inside the VR headset could have affected participant comfort, potentially impacting their overall experience.
3. Accuracy of Wind Exposure Areas: The head & neck wind condition may have unintentionally affected the cheeks, making it difficult to precisely isolate its impact. Similarly, chest wind exposure may have also influenced the neck area, which could affect comparisons between conditions. Since this study relied on participant-perceived wind exposure, future research should consider more precise wind positioning and defined exposure areas.
4. Experimental Design: Although participants were given breaks between trials, they still completed five consecutive game sessions, which may have led to fatigue accumulation. Future studies should consider longer rest intervals to further refine the experimental design.
5. Future Research Directions: Enhancing fan wind speed, introducing multi-directional wind control, and accounting for environmental and individual physiological differences could improve data accuracy and reliability. Expanding the sample size

would increase the stability and generalizability of the findings. Additionally, integrating multiple sensory modalities, such as auditory stimulation, could provide deeper insights into how wind feedback affects VR immersion. Incorporating physiological indicators (e.g., heart rate and galvanic skin response) could also contribute to a more comprehensive evaluation of user experiences.

Ultimately, this study provides empirical evidence for the design of wind haptic feedback in VR gaming, particularly in enhancing presence and emotional engagement. The findings offer practical insights into the application of wind feedback to different body regions, paving the way for future advancements in VR haptic interaction design.

References

Fang, C., Zhang, Y., Dworman, M., Harrison, C.: Wireality: Enabling complex tangible geometries in virtual reality with worn multi-string haptics. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–10. Association for Computing Machinery (2020). https://doi.org/10.1145/3313831.3376470

GII. Virtual Reality (VR) – Market Share Analysis, Industry Trends & Statistics, Growth Forecasts (2024–2029) (2024). https://www.giiresearch.com/report/moi1441594-virtual-reality-vr-market-share-analysis-industry.html

Goldman Sachs. Virtual & Augmented Reality: The Next Big Computing Platform. Goldman Sachs Global Investment Research (2016). https://www.gspublishing.com/content/research/en/reports/2016/01/13/eb9acad9-3db9-485c-864d-321372a23726.pdf

Kameoka, T., Kajimoto, H.: Tactile transfer of finger information through suction tactile sensation in HMDs. In: 2021 IEEE World Haptics Conference (WHC), pp. 949–954 (2021)

Mancini, F., et al.: Whole-body mapping of spatial acuity for pain and touch. Ann. Neurol. **75**(6), 917–924 (2014). https://doi.org/10.1002/ana.24179

Peiris, R.L., Chan, L., Minamizawa, K.: LiquidReality: wetness sensations on the face for virtual reality. In: Prattichizzo, D., Shinoda, H., Tan, H., Ruffaldi, E., Frisoli, A. (eds.) Haptics: Science, Technology, and Applications. EuroHaptics 2018. Lecture Notes in Computer Science, vol. 10894. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93399-3_32

Perret, J., Vander Poorten, E.: Touching virtual reality: a review of haptic gloves. In: ACTUATOR 2018; 16th International Conference on New Actuators, Bremen, Germany, pp. 1–5 (2018)

Ranasinghe, N., et al.: Season traveller: multisensory narration for enhancing the virtual reality experience. In: 2018 CHI Conference, pp. 1–13 (2018). https://doi.org/10.1145/3173574.3174151

Shen, V., Shultz, C., Harrison, C.: Mouth haptics in VR using a headset ultrasound phased array. In: CHI Conference on Human Factors in Computing Systems (2022). https://doi.org/10.1145/3491102.3501960

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Richir, S.: Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. Int. J. Virtual Reality **16**(1), 33–48 (2016). https://doi.org/10.20870/ijvr.2016.16.1.2880

Mathematical Models with War Games: Symbolism and Numerology in Game

Yijun Chen[1], Shaofeng Duan[2], and Donglin Wang[3](✉)

[1] School of Design, Central Academy of Fine Arts, Beijing 100020, China
[2] School of Arts and Humanities, Tianjin Academy of Fine Arts, Tianjin 300141, China
[3] School of Architecture and Design, Beijing Jiaotong University, Beijing 100044, China
315539439@qq.com

Abstract. This text explores the mathematical models and war games, focusing on the "representation" and "numbers" in the game. It discusses how ancient war games often contained the ultimate imagination of civilization's cosmic order. The text analyzes two ancient Chinese games, Liubo and Go, which represent different cognitive paradigms and philosophical concepts. Liubo is seen as a game that simulates nature and reflects the concept of "heavenly destiny," while Go is regarded as an abstract mathematical model that emphasizes human rationality and strategy. The text also examines the cultural and philosophical implications of these games, as well as their impact on modern game design. It proposes a modern war game theory that combines the "heavenly destiny randomness" of Liubo and the "strategic equilibrium" of Go, aiming to provide new ideas for game design in the digital age.

Keywords: Mathematical models · War game theory · Game Cultural design

1 Introduction

The rules of ancient games have been shown to reflect civilizations' ultimate imaginations of the order of the universe [1]. This is evident in the oracle bone inscriptions of the Shang Dynasty, which contain references to games, as well as in the star charts of six games on the lacquered chests of Zeng Houyi's tomb. These games have been interpreted as a means of rationalizing heaven and earth and regulating yin and yang. This is due to the existence of another strand in Chinese culture: primitive thinking, Yin-Yang and Taoism, and Taoist culture, represented by the art of counting and squaring [2]. The isomorphism between the game of Liubo and the style board (an ancient divination tool) demonstrates that the gaming model is, in fact, a figurative expression of numerical thinking.

Many modern war games lose their heavy philosophical framework and focus on sensory experience. On the one hand, to lower the threshold, military strategies are over-simplified, resulting in a disconnect from the logic of real war; on the other extreme, high-fidelity games struggle to attract mass players due to their complex controls and steep learning curve; and entertaining games are often criticised for their lack of tactical depth.

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 18–29, 2025.
https://doi.org/10.1007/978-3-031-93715-6_2

In the contemporary context, where digital technology is subverting the cognitive paradigm and reshaping human interaction, the "heaven-earth-human" ternary structure of traditional games offers a distinctive theoretical framework for modern game de-sign. The reinterpretation of these models carries dual significance:

At the technical philosophy level, the term offers cultural and genetic inspiration for game algorithms in the age of artificial intelligence, and it forms an isomorphism between the Chinese mathematical universe model and the contemporary data universe model. At the level of experience design, the term, with the help of extended reality technology, upgrades the two-dimensional chessboard of "Heaven and Earth" into an interactive universe narrative space, which provides a new way of thinking about con-temporary spatial computing and immersive experience.

2 Liubo: The Mapping of Probability Space and the Concept of Fate

'Game' in ancient China refers to two types of games, namely 'Liubo' and 'Go', as the early strategic war games in China, and later the two types of games are combined to refer to the game of chess. The two games were later combined to refer to chess games.

2.1 Liubo Game Board

Liubo is a two- or four-player chess game, popular in China from the Warring States period to the Jin dynasty, in which the winner is the one with more chips, and the game simulates the behaviour of birds such as owls hunting fish in a pond. In 1973, a complete set of six-player chess in a lacquer box was unearthed in the Mawangdui Tomb No. 3 of the Western Han Dynasty in Changsha, including a chess board, a chess set and a piece, a dice set and other artefacts.

The game design of Liubo is a vivid imitation of the natural environment. The Liubo board is square in shape, simulating the earth, with the centre and surrounding boxes marked with inlays. The central area of the board is a box called the 'pool', which usually contains fish-like chips, symbolising the limited natural resources to be fought over and the core area they represent. The perimeter of the centre is made up of 'L' or 'T' shaped chess paths, called curved paths, with a total of twelve paths and four nodes with bird patterns at the corners. These path nodes mimic the different directions of 'up, down, left, right and centre', with shortcuts going in different directions, as well as 'dangerous paths' that are not conducive to the movement of the game, and even paths that mimic the nature of nature, such as 'water'. There are shortcuts going in different directions, as well as 'dangerous paths' that are not conducive to playing chess, and there are also obstacles that mimic natural mountains and rivers.

In addition to the board as a simulation of nature, the pattern itself is a geometric expression of the cosmology of yin and yang and the five elements [3], and the form of its construction is also modelled on divination instruments [4], which are a model of the universe that can be used for divination [5] (Fig. 1).

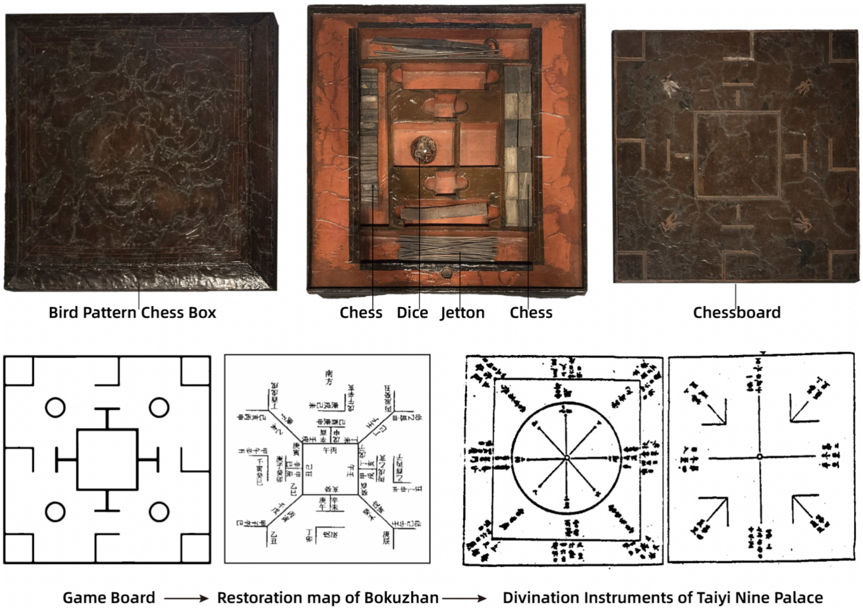


Fig. 1. The similarity of Liubo chess sets and divination patterns unearthed from Mawangdui No. 3 Western Han Tomb in Changsha.

2.2 Liubo Game Dice

Another important instrument in the Bo set is a spherical octahedral die. Each of the sixteen sides is engraved with a number from one to sixteen, while the two opposite sides are engraved with the character 'Jiao' in seal script on one side and the character 'Qiwei' on the other. During the game, the player determines the movement of the pieces by the number of dice points, similar to the mechanism of Monopoly.

If the board simulates geography, the pieces simulate creatures and the dice simulate the sky. The action of rolling the dice is symbolic of asking the will of Heaven, and the result of the dice is the judgement of Heaven. The mechanism of the dice game allows the variable factor of 'heaven' to be projected into the objective 'geography' and 'biology', while the randomness of its variables reflects the uncontrollability and uncertainty of the outcome caused by nature (Fig. 2).

Ancient China has a deep understanding of the relationship between 'heaven' and 'War' – 'The great events of the state are in ritual and military affairs [6]. The most important thing we do before a war is the rituals and the military. One of the most important things we do before a war is to examine the role of the sky before and during the war. Through divination and rituals, we ask Heaven whether it is right to start a war, whether it is the right time to start a war, and whether the war process will go smoothly and the result will be successful.

The Liubo game, as a reenactment of ancient wars, can re-conceptualise the role of dice by focusing on the interplay between 'sacrifice' and its underlying representation

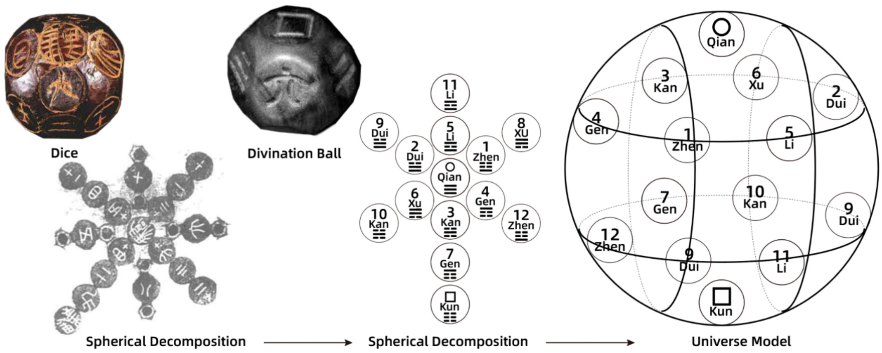


Fig. 2. Analysis of the Liubo dice in relation to the 'Meibuwan' and the model of the universe.

of 'heaven's will' and 'war'. The role of dice can be re-conceptualised by emphasising 'ritual' and the interplay between 'destiny' and 'war'.

The octahedral design of the Liubo dice (engraved with the numbers 1–16 and divination inscriptions) is not a simple probability tool for games, but a miniature model of the universe. The prototype of the dice is similar to the 14-sided 'meibuwan' [7] used for divination, and people regard the dice as a three-dimensional spherical heaven and earth, and stand inside the dice to observe heaven and earth, with the upper and lower hemispheres forming a three-dimensional model of the positive and negative eight trigrams of heaven. In addition, the number on the surface of the dice corresponds to the eight trigrams. For example, 1 is Zhen Gua, 2 is Dui Gua, 3 is Kan Gua, 4 is Burgundy Gua, 5 is Li Gua, 6 is Xun Gua, the circle is Qian Gua and the square is Kun Gua. You can start a hexagram by touching the numbers on the pills. The first touch is for the lower hexagram, the second touch is for the upper hexagram and the third touch is for the moving lines. If you touch the square, the trigrams are static; if you touch the circle, the six lines are all moving.

3 Go: The Clash of Digital Models and Royal Philosophy

3.1 The 'Humane' Turn of the Go Model

Although both Go and Liubo belong to the same ancient Chinese game system, they are rooted in very different cognitive paradigms and zeitgeists, and both show profound philosophical differences, from rule design to cultural metaphors. Liubo was born at a time when the practice of the square arts was widespread and the concept of heavenly destiny dominated social thought; its game mechanism projected the uncontrollability of 'heaven's will' through dice, and the layout of the board mimicked mountains and rivers and astrological phenomena, which was essentially a figurative interpretation of the cosmological concept of 'the interaction between heaven and man'. -The movement of the chess pieces is limited by the number of dice points, just like the limited mobility of man before the laws of nature.

On the contrary, Go matured during the Spring and Autumn Period and the Warring States Period, when the rites and music of the game were in ruins, coinciding with the rise of rationalism and the deconstruction of the aristocracy, and its extremely simplified apparatus (the grid board and homogeneous pieces) stripped away natural mimicry and random variables, constructing instead a series of purely abstract playing spaces. In this space, victory or defeat no longer depends on the intervention of 'heavenly fate', but on the player's overall control of the topological network, and every move is a mathematical deconstruction of 'connectivity' and 'boundaries'. Each move is a mathematical deconstruction of 'connectivity' and 'boundaries', alluding to the rational spirit of 'Zhou Yi': 'Change and simplicity lead to the truth of the world'.

The egalitarian qualities of Go are particularly worthy of study: all the grid spaces are of the same shape, with no distinction between centre and boundary and path of travel; all the pieces are of equal value, with no hierarchical differentiation between 'lords' and 'dispersed', and no pivot of survival or death of the 'king'; victory or defeat depends only on the topological effectiveness of the radiation of power. There is no hierarchical differentiation between 'lords' and 'dispersed', and there is no pivot of the 'king'; victory or defeat depends only on the topological effectiveness of the radiation of power. This design is not only a tacit subversion of feudal hierarchy, but also reflects the fusion of Taoism's idea of 'qi matter' and the military school's theory of 'potential' – the underlying logic of winning in Go through the dominance of space rather than the consumption of sub-power. The underlying logic of winning in Go is through the dominance of space rather than the consumption of pieces [8]. The assertion of Yinwenzi "that 'those who seek with intellect, the metaphor is like a game, the decision-making in Go is all down to me' further sublimates Go as a metaphor of individual rationality against chaos: players must break through the non-linear 'hand muscles' (rather than linear path planning) in the discrete topology of 361 intersections to This process is highly abstract, but implies strict graph-theoretic laws (e. e.g. the essence of 'qi' is the computation of degrees of freedom of connected branches).

The difference between the two also reflects the dual nature of ancient Chinese thought: Liubo inherited the primitive mathematical tradition and regarded the game as a ritual for exploring heavenly opportunities; while Go represents the humanistic awakening of the Axial Age, transforming the game into a purely mental exercise. This shift from 'heavenly destiny' to 'humanism' is, in fact, a spiritual slice of the evolution of Chinese civilisation from magician culture to ethical rationality.

3.2 Game Models Mirroring the Social Turn

The situation of Go in the Spring and Autumn and Warring States Periods was actually a civilisational conflict in miniature – its flat game structure completely dismantled the hierarchical order of feudal rites and became the spiritual mirror image of the era of vassal usurpation and sons competing for supremacy. When aristocratic politics collapsed in the chaos of 'rites and music and vassal conquests', Go eliminated the 'lord' and 'dispersed' hierarchical order with the rule design of 'no nobility or inferiority of equal sons'. The rule design of 'equal pieces without nobility or inferiority' eliminates the power differential between 'lords' and 'scatterers', and shifts the criterion of victory and defeat from the protection of 'symbols of the king's power' (e.g. the 'lords' of Liubo) to the quantitative competition for power in abstract space. This decentralised game paradigm resonates with the utilitarianism of Legalism, which is based on the principle of ‘following the name and responsibility’ and ‘not distinguishing affinity’, but touches the ethical nerve of Confucianism, which is based on the principle of ‘proximity and respect’. The record of Han Fei Zi that ‘Liubo’s noble lord, the Confucian, is harmful to justice’ is a vivid footnote to the clash of the two philosophies of governance: Liubo determines victory or defeat by killing the ‘lord’, alluding to the legalist idea of ‘catching the thief and capturing the king’; the power and tactics of Liubo are the same as those of legalism, ‘catching the thief and capturing the king’. Confucianism rejected it as ‘harmful to righteousness’, but was deeply disturbed by the disintegration of the patriarchal order of the Zhou dynasty.

The stigmatisation of Go as a ‘tricky way’ reveals a deeper rift in the Axial Age’s concept of heavenly ways. While the traditional philosophy of ‘the way of kings’ attributed the legitimacy of war to the mandate of ‘heaven’, Go stripped away the sacred rituals of sacrifice and divination, and attributed victory and defeat purely to calculation and strategy, which was tantamount to proclaiming the separation of ‘heaven is too far away and humanism is very close’ [9]. It is tantamount to proclaiming that ‘the way of heaven is too far, and the way of man is very near’ [9], the separation of heaven and man. Pi Rixiu’s ‘The Original Game’ asserts that the emergence of Go ‘must have originated in the Warring States’ [10], precisely because the core of its rules is highly compatible with the pragmatism of Zonghongjia’s ‘power for power’ – the game board The use of cheating, the estimation of the power of change, it is Su Qin and Zhang Yi, as zongheng hengjiazi tactical projection. This attempt to abstract war into a ‘calculable game’ not only dissolved the sanctity of ‘honouring the heavens and respecting the ancestors’, but also foreshadowed the prototype of the ‘Confucianism outside legalism inside’ ruling technique in the imperial era.

Interestingly, the development of the rules of Go and Liubo during the Warring States period illustrates the dual paths of power reconstruction: Go builds a rational authority based on spatial control by eliminating hierarchy; Liubo symbolises violent games by strengthening the mechanism of ‘killing the lord’. The two seem to be opposites, but in fact they share the spirit of the times of ‘demoralisation’ – when Confucius lamented ‘propriety disintegration’, the game had quietly turned into a metaphorical theatre of power competition.

But it is this transformation of game thinking that has led to an update of the perception of the cosmic model. The grid-based system and two-dimensional mathematical perception of Go have led to a number of digital models. For example, Chinese cities have already formed a grid-based layout; the Song Dynasty’s ‘Yu Signs Map’ has already formed a grid-based geographical perception and mapping; and the layout of military formations and in military education has also been presented as a way of labelling chess boards and chess pieces. In modern times, the progress of the revolution has been analysed from the dual perspective of space and network of relations, and the layout of Go has been used to illustrate the ‘encirclement of the city by the countryside’ and the ‘theory of protracted war’.

4 Computational Philosophy of Traditional Game Models: Archetypal Systems in the Numerology

Liubo and Go represent two extreme but complementary paradigms of ancient game models:

Liubo is a 'natural variable' model dominated by heavenly fate: In terms of the core mechanism, the dice mechanism simulates natural chaos and reflects 'the unpredictability of heaven' through randomness; in terms of the cultural logic, the outcome of the dice is given the meaning of 'heavenly destiny', reflecting the cosmological view of 'heaven and man'; in terms of the rule characteristics, the choice of paths is subject to dice selection, reflecting the cosmological view of 'heaven and man'; in terms of the rule characteristics, the choice of paths is subject to dice selection. In terms of cultural logic, the outcome of the dice is given the meaning of 'heaven's destiny', reflecting the cosmology of 'heaven and man'; in terms of rules and characteristics, the choice of paths is limited by the number of points on the dice, and players must adjust their strategies under limited control.

The intellect-led 'strategic equilibrium' model of Go: in terms of core mechanism, global equilibrium removes the random factor, reflecting 'what people can do'; in terms of cultural logic, it reflects the Confucian idea of 'harmony', emphasising rational planning and moral constraints. In terms of cultural logic, it reflects the Confucian idea of 'neutralisation', emphasising rational planning and moral constraints; in terms of rule characteristics, it builds a network of connections by dropping stones, and victory or defeat depends on the balanced allocation of resources over the long term.

Based on the game models of Liubo and Go, we can propose a modern war game theory that integrates 'randomness of fate' and 'strategic equilibrium'. The essence of the fusion model is to reconfigure human perception of nature and war through the interaction of dynamic environmental variables and the strategy network. By deconstructing the dice mechanism of Liubo and the topological network of Go, a 'culture-technology' dual-driven game framework is constructed. The model not only breaks the traditional data-driven one-dimensional confrontation, but also integrates natural variables, ethical constraints and surreal narratives into the system design, creating an immersive war experience that combines unpredictability and strategic depth (Fig. 3).

There are two key combinations, on this basis, the entire path from cultural logic to technical implementation is realised:

Cultural-technological closure: the algorithmic closure of traditional cultural logic is achieved through mechanisms such as divination event chains and moral scoring.

Dynamic equilibrium: the interaction between natural variables (randomness) and strategic equilibrium (controllability) is the core tension of the model.

– Theoretical layer:

Natural Variables: Inheriting the dice logic of Liubo, the 'unpredictable fate' is transformed into controllable algorithmic parameters such as resource fluctuations and climate events.

Strategic equilibrium: Based on the topological network of Go, emphasising the balance between global connectivity control and dynamic blocking.

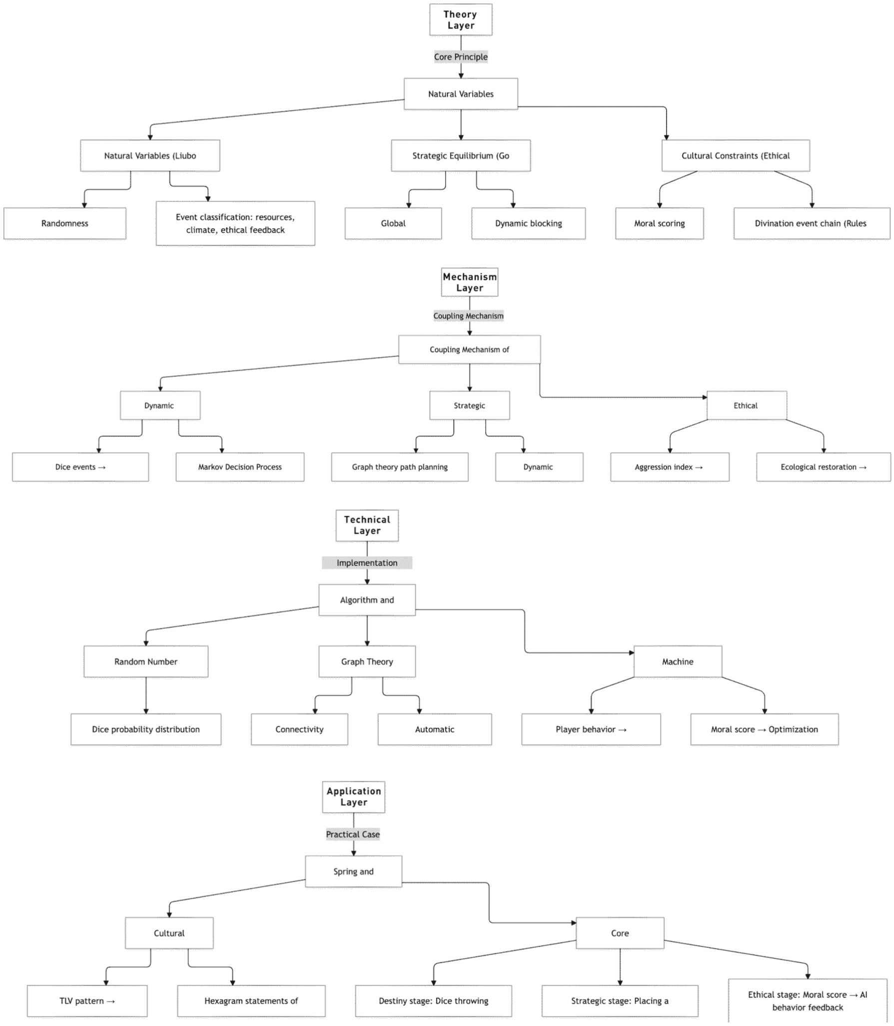


Fig. 3. Logical thinking of prototype system.

Cultural Constraints: The interaction between 'Sacrifice and Rong' is regulated by a chain of moral scoring and divination events.

– Mechanism layer:

Dynamic environment engine: state transfer through dice events and MDP model to ensure coupling of randomness and strategy.

Strategy network: graph theory algorithm to support path planning, heat map to visualise resource distribution and ecological pressure.

Ethical feedback: quantifies player behaviour (e.g. aggression index) and dynamically adjusts game difficulty and event triggering probability.

– Technology layer:

RNG algorithm: accurately controls the probability distribution of dice results (e.g. ‘Qian Gua’ and ‘Kun Gua’ event weights).

Graph theory tools: ensure real-time and robust path planning (e.g. network reconfiguration after dynamic blocking).

Machine learning: train AI opponents using player behaviour data to achieve adaptive strategy feedback.

– Application layer:

Cultural symbol translation: Translate abstract symbols such as TLV patterns and trigrams into interactive interface elements and narrative text.

Game Process: Realise the closed-loop interaction of ‘destiny-strategy-ethics’ in stages to verify the feasibility of the theory (Table 1).

Table 1. Comparison with conventional models.

Dimension	Traditional War Game Model	New Model
Randomness source	Fixed probability tables or pseudo-randomised algorithms	·Dice event chain (cultural logic driven)
Strategy Depth	Path planning or resource management uni-dimensional optimisation	Global connectivity network + dynamic blocking + ethical feedback
Cultural expression	Surface symbols (e.g., historical skin)	Regularised cultural logics (e.g. chains of divinatory events)

5 Go Experiment Based on Earth Model

Combined with the above framework, a board game based on the earth model can be formed, with the earth’s latitude and longitude network to construct spatial coordinates, and with curved surfaces to break the boundary limitations of the two-dimensional plane, forming an infinite space, and with the land and the sea to construct the boundaries for the movement of the figures ‘sea, land and air’, forming different areas with different rules of fall, forming a three-dimensional borderless space, and thinking a comprehensive war strategy under different natural conditions. In the border space, under different natural conditions, we can think of comprehensive war strategies (Fig. 4).

The following are the innovations of Earth Go:

– Rule Innovation

Borderless board: Unlike traditional Go, which is a rectangular board with fixed boundaries, Earth Go does away with the traditional method of using the four outermost edges of a Go board as boundaries, and plays squarely on the board, with the top and bottom two edges considered as one edge, and the left and right two edges, which also

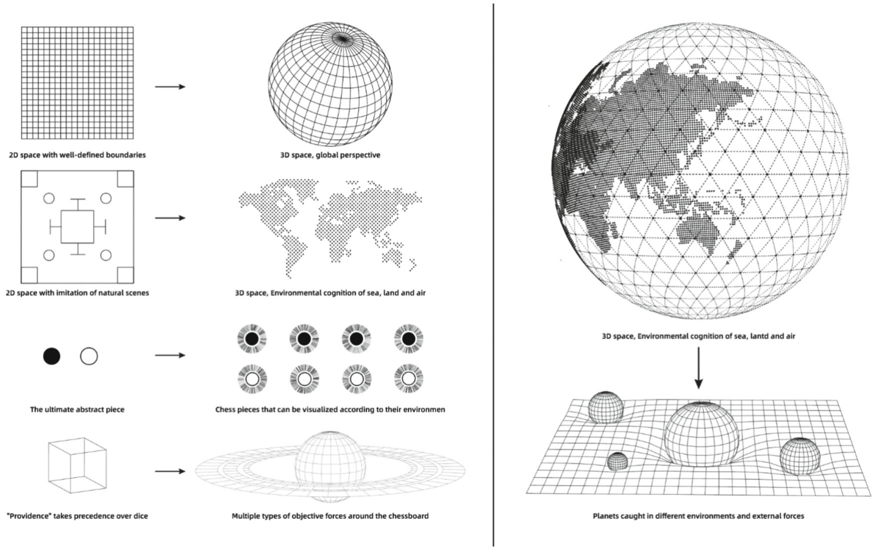


Fig. 4. The basic logic of the construction of the game model and the integration of various elements.

perform the same action when a move is made. This means that the outer edges of the board are connected, symbolising the circle of the earth and providing a more continuous, seamless playing area without the constraints of traditional boundaries.

Addition of gas points: EarthGo has added gas points to make the game more flexible and interesting. This change increases the complexity and strategic possibilities of the game, and players must take these additional chi points into account when making moves and calculating the life and death of their pieces.

– Game innovations

3D Thinking: The design of the EarthGo board is based on Einstein's Theory of Relativity and has a 3D mindset. Players must not only think on a traditional 2D plane, but also consider the 3D spatial relationships between pieces, requiring a more complete and in-depth understanding of the structure and strategy of the game.

New strategies: As the board and rules have changed, new strategies have emerged. For example, players can take advantage of the lack of borders to attack from unexpected directions, or use the addition of air points to create more complex life and death positions for pieces, making the game more unpredictable and challenging.

– Cultural and conceptual innovations

Global Perspective: The design of the EarthGo board, with its Earth-like lines of latitude and longitude, and the concept of land and sea borders, allows players to think globally. This is not just a local battle on a small board, but a global strategic game, a new concept in the field of Go.

– Technological innovation

Integration with digital technologies: Earth Go is integrated with digital technologies such as Virtual Reality (VR) and Augmented Reality (AR) to provide a more immersive gaming experience. Players can see the different natural states of the 3D Earth board in the virtual environment, as well as the actual shape of the pieces in this state, and make moves through motion-controlled devices, which is a new attempt to combine traditional board games with modern technology.

AI application: AI technology can be applied to Earth Go to support the abstract pieces' concrete changes in response to the environment; AI can analyse the game in real time and make moves according to the natural rules of the Earth's environment and human strategies, providing players with a more challenging and interesting gaming experience.

6 Conclusion

By analysing the philosophical core and cultural logic of the ancient Chinese game models Liubo and Go, this paper proposes a theoretical framework for modern war games that integrates the 'randomness of fate' and the 'balance of strategy'. It is found that Liubo and Go represent the two poles of the ancient cognitive paradigm: the former simulates natural chaos and unpredictable fate through a dice mechanism, which is in line with the cosmology of 'heavenly and human feeling'; the latter constructs rational authority through a topological network, which embodies the practical wisdom of 'humanism can be done'. The latter uses topological networks to build rational authority, reflecting the practical wisdom of 'humanity can be done'. The integration of the two is not only an innovation in game mechanics, but also a deep exploration of the dialectical relationship between controllability and uncontrollability in the nature of war.

6.1 Key Contributions

– Technical translation of the cultural logic:

Liubo's dice mechanism is reconstructed as a dynamic environment engine that simulates natural chaos through random events and at the same time combined with blockchain technology to generate an irreversible "chain of heavenly fate and causality" that algorithms the logic of interaction between "ritual and military".

The topological network of Go is upgraded to a three-dimensional strategic space, and through graph theory algorithms and quantum computation, the non-linear pattern evolution is realised, and its 'no all-same' rule is extended into an ethical constraint system to quantify the impact of players' moral choices on the outcome of war.

– Construction of a dynamic cognitive interface:

The fusion model transforms traditional symbols into interactive narrative interfaces through meta-universe technology. The player's decision not only affects the virtual battlefield, but also generates a dynamic moral topology through neural interface technology, realising the multi-dimensional feedback of 'strategy-ethics-history'.

– Innovation of the wargame paradigm:

The proposed case of EarthGo simulates the globalised features of modern war with a borderless board and 3D space strategy, combining AI and VR technologies to break through the planar limitations of traditional games and provide a new path for immersive war experience.

6.2 Theoretical Significance and Practical Value

– Interdisciplinary revelation

The model combines military simulation, artificial intelligence and cultural studies to provide methodological references for the field of digital humanities. Liubo's divination event chain can be applied to modelling uncertainty in complex systems, and Go's equilibrium strategy can optimise resource allocation algorithms.

– Social Criticism Function

Through the moral scoring and ethical feedback mechanism, the game becomes a testing ground for reflecting on war violence. Players' choices directly map the power, resources and human dilemmas in reality, responding to the social controversy over the glorification of game violence.

6.3 Summary

The fusion model of Liubo and Go is not only a digital activation of the ancient cognitive paradigm, but also an innovation of the modern war game design paradigm. It upgrades war from a 'data-driven confrontation' to an 'experiment in civilisation evolution', and reconstructs human perceptions of violence, rationality and nature in the interweaving of technology and culture. Future research can further explore the integration of cross-civilisation gaming genes, promote the paradigm shift of games from entertainment tools to 'cognitive prisms', and provide a new vision for globalised digital humanities research.

References

1. Ge, Z.: History of Chinese Thought, 3 vols., 2nd ed. Fudan University Press, Shanghai (2013)
2. Li, L.: A Study of Chinese Magic and Divination. Zhonghua Book Company, Beijing (2000)
3. Tseng, L.L.-Y.: Representation and appropriation, rethinking the TLV mirror in Han China. Early China **29**, 163–215 (2004)
4. Chuang, H.Z.: Gambling games in the Pre-Qin and Han dynasties: Bo and Sei. J. Chin. Soc. Hist. Sci. Technol. **27**, 93–104 (2022)
5. Li, Z.: Liubo chess games unearthed in the tomb of King Zhongshan of Trek – a comparison of designs with Yinwan's 'Bo Guan Zhan'. Chin. Hist. Antiquit. **1**, 8–15.5 (2002)
6. Wang Xuejun, F., He Weili, S.: The original context and change of meaning of 'the greatest events of the state are sacrifice and Rong.' Ancient Civil. **2**, 92–98 (2012)
7. BILIBILI. https://www.bilibili.com/video/BV1X44y1Y7Y9/. Accessed 25 November 2021
8. Sun Wu, F., Lionel Giles, T.: The Art of War. Luzac & Co., London (1910). (Original work written around 500 BCE)
9. Zuo, Z. Zhonghua Shubu, Beijing (1981). (Original work written around 500 BCE)

Play and Learning Across Realities: Design Strategies for a Permeable Magic Circle

Erik D. van der Spek(✉)

Department of Industrial Design, Eindhoven University of Technology, Groene Loper 3, 5612 AE Eindhoven, The Netherlands
e.d.vanderspek@tue.nl

Abstract. The rise of Mixed and Virtual Reality technologies can stimulate more intense game experiences, offer new opportunities for learning and have made new interaction paradigms possible. However, it is important to understand how players engage with mixed and virtual realities, and how to design ways to subtly pull people in and out of the game world as the context requires. This paper proposes a framework based on concepts of the magic circle and diegesis, to provide four Mixed Reality interaction design strategies, and subsequently gives a game design exemplar for each of these categories to highlight how they can be operationalized.

Keywords: Mixed Reality · Interaction Design · Magic Circle · Presence · Play

1 Introduction

The idea of Virtual Reality may have started out as a science fiction fantasy in the early twentieth century by noted authors like Stanley Weinbaum or Philip K. Dick, it became a reality in the mid-1960s with the development of the first actual Head Mounted Display (HMD) [1]. Virtual Reality (VR) has taken flight in the past decade with affordable Virtual Reality HMDs for home settings such as the Oculus Rift and more recently as time of writing, Meta Quest 3 [2]. Interaction technologies have also developed over the years, from only allowing static viewing in the beginning [1], to controllers and precise hand and finger tracking with modern commercial HMDs [2, 3]. Other interaction technologies developed to improve immersion within Mixed (MR) and Virtual Reality settings, range from tactile suits to eye trackers and treadmills, among others [4].

Immersion is one of the key concepts when it comes to VR HMDs, providing stronger feelings of being present in a mediated fantasy world when putting on the headset as opposed to 2D screens [5]. Virtual reality offers many of the same potential benefits as conventional entertainment and serious games when it comes to for instance wellbeing [6] and learning [7], while also providing interesting opportunities to train affective skills and empathy [8]. A number of these benefits, such as student engagement and learning [9], or motivation to exercise [10] appear amplified when the VR game is more immersive than in low immersive settings (although the degree of immersiveness is inconsistently qualified in the literature as e.g. resolution, field of view or amount of senses stimulated, among others [cf. 11]).

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 30–40, 2025.
https://doi.org/10.1007/978-3-031-93715-6_3

1.1 Transitioning to and from the Virtual World

VR gaming is generally more immersive than conventional gaming because a larger part of the visual field of view is covered by the screen, but also because it is easier to imagine the self in the virtual world than when it is mediated through a screen further away [11]. One would imagine that donning a HMD immediately transports the player to the virtual world. Interestingly, the process doesn't seem to be quite that immediate. Researchers found that creating a transitional state between the real world and the virtual world, where for instance the real world is slowly faded out [12], or participants walk through a virtual portal [13] led to increased feelings of presence and virtual body ownership [12, 13].

Conversely, especially for VR games for serious applications, the immersiveness of VR can lead to problems with learning. Players who are too immersed can lose the capacity for reflection and critical thinking. In games this is related to the Flow paradox, where players become too engaged "in the moment" to reflect on their actions [14]. In theater and cinema, Brechtian estrangement (Verfremdung) techniques have been developed to temporarily break immersion, so viewers regain the capacity for critical thinking, provoking them to reflect on the narrative [15, 16]. Although less scrutinized than transitions *into* immersive Virtual Reality, recently some research has been done on techniques to satisfactorily transition out of VR, and transfer learning along with it [8, 17].

1.2 Mixed Reality and the Magic Circle

In parallel, new technologies have been developed to believably bring the virtual world to the real world. In the context of gamification this is done so that the positive engaging qualities of games can be integrated into real world situations, and improve, among others, motivation for learning and adopting positive behaviors. These technologies, which are called Augmented Reality (AR, when information is overlayed on the real world) or Mixed Reality (when AR and VR are combined [18]), have a history just as old as Virtual Reality, but have recently become more popular with the advent of smartphones, as well as with HMDs such as Microsoft HoloLens, Apple Vision Pro and Meta Quest 3. These technologies provide the ability to fade the amount of immersion that the user experiences and can subsequently bring the user into the virtual world and back into the real world to various degrees.

At this point it's important to note that physical immersion is correlated with, but not the same as the feeling of presence, or the feeling of being there in the virtual world. The first describes the physical capacity of the system to envelop the user's senses, the second is a psychological state as the result of a mental activity, where the virtual world is considered plausible and the self a part of it [11]. A separate but related notion in the study of games revolves around the so-called magic circle, a metaphor for an alternate bubble of reality that we adopt when we become playful, and where we let the rules of the game world take precedence over the rules of the real world [19]. Both of these mental activities in turn are related to the concept of aesthetic illusion, the suspension of disbelief and transportation into the game world [20]. In addition, the aesthetic distance denotes the perceived distance we experience between the real world and the game world inside the magic circle. It is hypothesized that a higher feeling of presence leads to a lower aesthetic distance, as a result of which a game is experienced more intensely [20].

As MR and VR technologies are becoming more commonplace, and lowering the aforementioned aesthetic distance may be good for the game experience but at times bad for learning, designers of educational MR experiences will look for strategies to control presence and the aesthetic distance, to push and pull (or transgress) players in and out of the magic circle, similar to the Brechtian Verfremdung techniques in theater and cinema. For the past eight years, students that I supervise have been prototyping Mixed and Virtual Reality entertainment and serious games, sometimes with the express purpose to trace the edges of the magic circle. However, over time it became clear that we lack a vocabulary to categorize different MR design strategies. Here, I will propose a framework based on designerly intent; how the design is trying to stimulate immersion into which area of the game world, and subsequently show design exemplars for each of the four resulting categories.

2 A Framework for the Design of Interfaces into the Game World

2.1 Related Work

Recently, a number of frameworks and categorizations for MR and VR interactions have already been developed. Next to tracking technology and learning pedagogy frameworks that are not relevant for the purpose of this paper, Papadopoulos and colleagues describe an extensive framework based on the interaction modality that can be used, such as visual (eye-tracking, surface detection, etc.), audio (speech recognition, music feedback, etc.), haptic (multi-touch, force feedback, etc.) and sensor-based interactions (pressure, data monitoring, etc.) [21]. Malinverni and colleagues make a distinction between providing a window on the world versus the world providing a support for the interaction, and contrasts this with whether the tracking uses markers or not [22]. Meanwhile, Steffen et al. take a more meta perspective on the types of Extended Reality technologies, their affordances and their subsequent ideal application domain [23]. An older framework by Rogers et al. describes the different interactions that MR affords in terms of physical effects having a digital counterpart and vice versa [24]. To the best of my knowledge however, no framework has tried to describe directly how the design of an interaction supports the experience of transgressing into the game world.

A type of game that has historically grappled with the boundaries of the magic circle, is that of pervasive games. In her research on tracing the magic circle, Nieuwdorp describes two metaphorical membranes that a person needs to transgress in order to enter the magic circle: a paratelic interface and a paraludic interface [25]. The paratelic interface denotes that the person first needs to become playful, and the paraludic interface posits that the person should then accept the semiotic domain of the game world as meaningful [25]. While this model implies a certain directionality, as far as I am aware, there is little empirical evidence that these interfaces need to be transgressed in this order. Perhaps one can also engage with the game world first as a motivator to become playful, or the two state transgressions operate in a positive feedback loop, strengthening each other.

In research on the design of persuasive games, Kors and colleagues define strategies to stimulate different kinds of empathy, based on different perspectives into the game world and its characters: a third-person (observer) perspective, a second-person (partaker) perspective, and a first-person (victim) perspective [8, 26]. Through this, they identify three conceptual universes that the player can mentally inhabit: the extrafictional universe (i.e. real world) that the player is in while playing; as well as inside the game world, the extradiegetic universe and the intradiegetic universe [26]. In the extradiegetic universe (relevant for the third-person observer perspective), the player is inside the game world, but spectating the events happening. In the intradiegetic universe (relevant for the second-person partaker and first-person victim perspective), the player is inside the game world and interacts in the main story line, either directly as the protagonist, or as another character that can influence the proceedings of the narrative. Similar to the concept of aesthetic distance, the experienced distance between the outside player and the main protagonist can lead to differential learning gains.

2.2 Paratelic and Paraludic Interfaces into Extradiegetic and Intradiegetic Universes

From these two concepts, viz. the paratelic and paraludic membranes that one needs to transgress to enter the magic circle, and the extradiegetic and intradiegetic universes that one can inhabit once mentally inside the game world, I propose a framework to describe MR and VR interaction design strategies. It should be noted that the above concepts need to be reinterpreted slightly.

In the case of transgressing into the magic circle, this is because, firstly, these concepts describe experiential outcomes of a mental activity, and designers can only create the mechanics in the hope of engendering certain aesthetics in the user [27], therefore they should be seen more as designerly intent for a certain experience to happen. Secondly, and related to this, bringing someone from a telic (serious) to a paratelic (playful) state, or having them be immersed in the game world's semiotics, is among others dependent on the fantasy proneness of the user [28] and likely not a matter of a single interaction, but due to a complex interplay of multiple interactions and aesthetical experiences.

In the case of extradiegetic and intradiegetic universes, there is not one clear definition of what constitutes diegesis in games [29]. In the strictest sense, diegesis relates to the story told within the game world [30], but this would exclude games without a clear focus on storytelling. In the broadest sense, diegesis relates to the understanding of the game text, which makes it very similar to the paraludic interface described above. For the purpose of this framework we interpret the terms extradiegetic and intradiegetic thusly: Extradiegetic interactions affect the believability and richness (or 'aesthetic illusion' [20]) of the game world; Intradiegetic interactions affect the main goals and mechanics of the game, which could be the story, the procedural rhetoric or the learning content.

The framework then describes the following four categories of mixed and virtual reality interactions:

1. Paratelic Extradiegetic interactions, or, interactions that (are intended to) playfully strengthen the aesthetic illusion of the game world.
2. Paraludic Extradiegetic interactions, or, interactions that strengthen the meaningfulness of the game world.

3. Paratelic Intradiegetic interactions, or, interactions that add playfulness to the main goals and mechanics of the game.
4. Paraludic Intradiegetic interactions, or, interactions that make the main goals and mechanics of the game more meaningful.

Next, I will briefly show design exemplars for each of these categories. Note that the framework is a result of these experimental prototypes and not a precursor to them; they were first made for a variety of purposes over multiple years and then analyzed and grouped together. As such their development did not follow a uniform systematic process, nor have they been evaluated uniformly. Where results are mentioned, these should be interpreted as no more than interesting hypotheses for follow-up research within the context of this framework.

3 Paratelic Extradiegetic: Vivezza

In the Paratelic Extradiegetic category, playful interactions are created that help the player transgress into a paratelic state, to more readily accept the virtual world. Either as a liminal interface into virtual reality, or to help strengthen the believability of a pervasive mixed reality. An example of such a playful interaction can be seen in the VR game "Vivezza" (Figs. 1, 2, and 3).

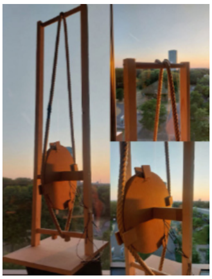


Fig. 1. The rope that participants could see and pull on in the real world.

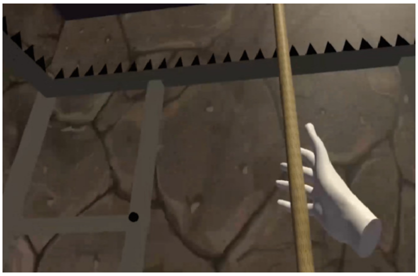


Fig. 2. Pulling on the rope in the virtual world hoists the player character into the arena.

Fig. 3. A screenshot of fighting enemies in the arena.

Vivezza is a VR arena fighting game where the player uses crossbows and swords to fight fantastical monsters. Inspired by the elevators that in ancient times would hoist wild animals into Rome's colosseum, the player has to step into a virtual elevator and pull themselves up with a rope into the arena. This rope pull system was created in real life, with a rotary encoder that sensed and transmitted the speed with which the player pulled into the game engine, and acted as a transdiegetic object [8] into the virtual world. This is an example of a Paratelic Extradiegetic interaction, because it is a playful activity that adds to the aesthetic illusion of the game world, without being part of the core activity in the game, i.e. fighting in an arena.

Participants of a small experiment would see the rope pull system when they entered the room, put on the headset and then find the rope in the same place in the virtual world. After hoisting themselves up, they could turn around and fight monsters. This was compared with a control condition that had no physical rope pull system, but still had to pull themselves up with the virtual rope. There was no significant effect of adding a physical rope on the feeling of presence $t(16) = .589$, $p = .564$, although there were too few participants (9 per group) to conclude anything conclusively.

4 Paraludic Extradiegetic: MathBuilder

Paraludic Extradiegetic interactions probably describe the most common forms of Augmented Reality outside of QR codes. Here, one example is a visual prop related to the game world that acts as an image target, but after scanning the targets, the game itself plays on the device with the camera. MathBuilder [31] is an exemplar of such a type of interaction. In this game, elementary school children learn about mathematics by playing as construction workers. They have a map in front of them with different construction sites and roads connecting them together (Fig. 4). The children play different roles (e.g. a supervisor, a carpenter, a mason) and scanning the construction sites gives them asymmetric information, which they have to combine in a collaborative fashion to solve mathematics puzzles (Fig. 5).

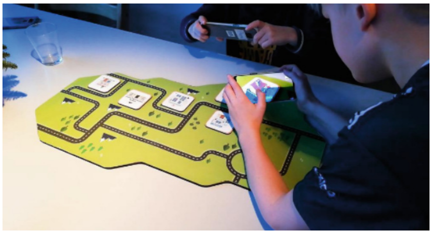


Fig. 4. Props of the game world in the real world that can be scanned as image targets.

Fig. 5. Scanning the image target starts an asymmetric educational game.

The map in this case again acts as a transdiegetic item, providing a meaningful context in the real world that the players take with them into the digital game. It grounds and supports the transition into the game world, even though the game mechanics and the goals that need to be completed themselves are all virtual. A qualitative study with eight 5th graders showed that students enjoyed learning mathematics through this type of collaboration and the process of scanning items and seeing the buildings grow in the virtual world [31].

5 Paratelic Intradiegetic: Heist Extravaganza

Paratelic Intradiegetic interactions are generally tangible playful mechanisms that are used to control the main action in the game. Theoretically any kind of game controller could fit here, but a better exemplar would be an interaction that has playful qualities by itself. Heist Extravaganza is an asymmetric multiplayer mixed reality game, where a group of friends set out to rob a museum after closing time. The game was designed to create a strong role differentiation between player types and VR and MR experiences. The VR player sneaks around the museum in room-scale, and has to stay out of sight of patrolling guards, dodge lasers, crack locks and ultimately open the safe (Fig. 6). They are being helped by three other players. One plays an engineer with a ‘super gadget’, a device that can be used to crack locks and open the safe (Fig. 7), another person has a switchboard to control the cameras (Fig. 8) and the last person is the mastermind who has a playbook detailing among others the guards on duty (Fig. 9).

Fig. 6. The view of the VR player as they skulk through the museum.

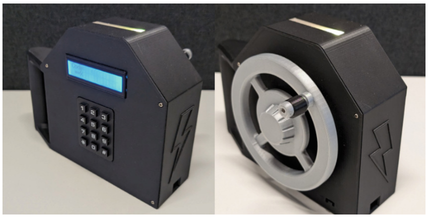


Fig. 7. The super gadget that the engineer uses to crack locks.

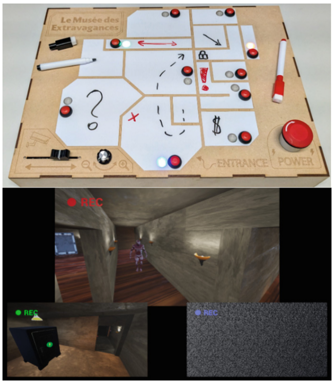


Fig. 8. The switchboard that can be used to control the camera feeds.

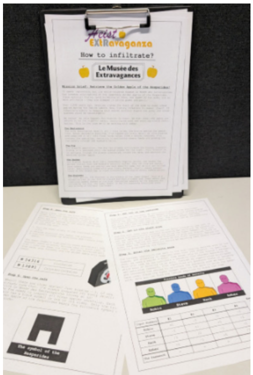


Fig. 9. The flipboard with instructions for the Mastermind.

All interactions offer a Paratelic Intradiegetic experience, where the separate interactions themselves are playful but also meaningfully contribute to the game story and the main goal of robbing the museum. The best exemplar is the super gadget however, since it offers minigames on the device itself (tapping the right pattern on the keypad, carefully moving the safe dial to feel the bolt unlock) to reach the main goals in the VR game. Two user tests with four players each showed high scores for social presence (Empathy M = 3.125, SD = .461, Negative feelings M = .271, SD = .377, and Behavioral engagement M = 3.109, SD = .430, on a range of 0–4) on the Social Presence in Gaming Questionnaire [32].

6 Paraludic Intradiegetic: ChemiKami AR

Paraludic Intradiegetic interactions are similar to Paraludic Extradiegetic interactions, but this time the props do play a meaningful part in the story itself or solving the main challenges in the game. An example of this is the educational game ChemiKami AR, where players mix and match image targets of chemical elements and related real world contexts. For instance, the element Carbon needs to be matched with a picture of a pencil that is out of graphite. If the correct pairs are scanned, an anthropomorphic Carbon element fills up the pencil with flashy special effects (Fig. 10). ChemiKami AR was used for an experiment to see if fantasy and anthropomorphization (Fig. 11) improves learning, compared with an AR game that showed more realistic elements and applications. This was found to be the case: F(1,96) = 9.08, $p = 0.003$, $\eta p^2 = 0.08$ [33].

Fig. 10. Matching the right chemical element with its real-world context leads to flashy special effects.

Fig. 11. Anthropomorphized chemical elements improve recall of knowledge.

Unlike Paratelic interactions, the Paraludic Extradiegetic and Intradiegetic interactions in these cases don't offer much interesting interaction potential outside of the virtual or augmented reality. However, that doesn't mean that they don't play a role here. In this case, the cards used to make the digital game possible, offer nice transdiegetic mementos that can be collected and traded outside of the game. In this way, the magic circle can be made permeable and extended into the real world, or in other words the aesthetic illusion can linger outside of the game world, and subsequently possibly improve the transfer of learning to the real world.

7 Discussion and Conclusion

The rise of Mixed and Virtual Reality technologies have made new interaction paradigms possible that can stimulate play and learning. However, current frameworks on the way we can interact with these new technologies seem to focus primarily on modalities and technological affordances, and not on design strategies how to facilitate drawing people in and out of the virtual game world. Here I tried to create a framework based on work on pervasive games and diegesis in persuasive virtual reality, leading to a categorization into four interaction types: Paratelic Extradiegetic, Paraludic Extradiegetic, Paratelic

Intradiegetic and Paraludic Intradiegetic interactions. I've showed design exemplars for each of these categories that I hope can serve as an inspiration for Mixed and Virtual Reality researchers and designers.

This framework is very much a first attempt in charting the design space of Mixed Reality interactions according to these dimensions. There are still a number of limitations. The framework is based on prototypes developed with Research through Design processes for different purposes, and retrospective reflection on the designs generated. These designed prototypes did not follow a systematically similar approach, and it is therefore impossible to make generalized, let alone quantified, comparisons between the efficacy of the different design strategies to bring people into the magic circle and craft a convincing aesthetic illusion of the game world. So far the design exemplars also primarily show how to improve the aesthetic illusion within the interaction categories. It would be interesting to see if the same categories can be used to temporarily disengage players from the magic circle as a means to improve reflection, or whether additional categories need to be added, for mechanisms such as Brechtian Verfremdung, or cognitive and ludonarrative dissonance [34]. Above all I hope that other researchers take this framework as a basis and add new interaction strategies, to improve the design and our understanding of how players interact in a Mixed Reality future.

Acknowledgments. This paper builds on the work done by Bachelor, Master and PhD students in the TU/e Industrial Design Game/Play Lab, and could not have been written without their excellent work. Vivezza was made by then Bachelor students Sören de Kok en Remy Schunck in 2022. MathBuilder was made by then Master students Almar van der Stappen, Yunjie Liu, Jiangxue Xu and Xiaoyu Yu in 2019. Heist Extravaganza was made by then Master student Jules van Gurp in 2025. ChemiKami AR was made by then PhD candidate (and currently professor at HKUST) Tengjia Zuo in 2021.

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

References

1. Boas, Y.A.G.V.: Overview of virtual reality technologies. In: Proceedings of the Interactive Multimedia Conference, vol. 2013 (2013)
2. Aros, M., Tyger, C.L., Chaparro, B.S.: Unraveling the meta quest 3: an out-of-box experience of the future of mixed reality headsets. In: Stephanidis, C., Antona, M., Ntoa, S., Salvendy, G. (eds.) HCI International 2024 Posters. HCII 2024. Communications in Computer and Information Science, vol. 2116. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-61950-2_1
3. Juan, M.C., Elexpuru, J., Dias, P., Santos, B.S., Amorim, P.: Immersive virtual reality for upper limb rehabilitation: comparing hand and controller interaction. Virt. Real. **27**(2), 1157–1171 (2023)
4. Novacek, T., Jirina, M.: Overview of controllers of user interface for virtual reality. Presence Virt. Augment. Real. **29**, 37–90 (2020)

5. Buttussi, F., Chittaro, L.: Effects of different types of virtual reality display on presence and learning in a safety training scenario. IEEE Trans. Vis. Comput. Graph. **24**(2), 1063–1076 (2017)
6. Halbrook, Y.J., O'Donnell, A.T., Msetfi, R.M.: When and how video games can be good: a review of the positive effects of video games on well-being. Perspect. Psychol. Sci. **14**(6), 1096–1104 (2019)
7. Wouters, P., Van Nimwegen, C., Van Oostendorp, H., Van Der Spek, E.D.: A meta-analysis of the cognitive and motivational effects of serious games. J. Educ. Psychol. **105**(2), 249 (2013)
8. Kors, M.J., et al.: The curious case of the transdiegetic cow, or a mission to foster other-oriented empathy through virtual reality. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2020)
9. Chen, J., Fu, Z., Liu, H., Wang, J.: Effectiveness of virtual reality on learning engagement: a meta-analysis. Int. J. Web Based Learn. Teach. Technol. **19**(1), 1–14 (2024)
10. Mouatt, B., Smith, A.E., Mellow, M.L., Parfitt, G., Smith, R.T., Stanton, T.R.: The use of virtual reality to influence motivation, affect, enjoyment, and engagement during exercise: a scoping review. Fron. Virt. Real. **1**(564664), 1–23 (2020)
11. Cummings, J.J., Bailenson, J.N.: How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol. **19**(2), 272–309 (2016)
12. Jung, S., Wisniewski, P.J., Hughes, C.E.: In limbo: the effect of gradual visual transition between real and virtual on virtual body ownership illusion and presence. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 267–272. IEEE (2018)
13. Steinicke, F., Bruder, G., Hinrichs, K., Steed, A., Gerlach, A.L.: Does a gradual transition to the virtual world increase presence? In: 2009 IEEE Virtual Reality Conference, pp. 203–210. IEEE (2009)
14. Squire, K.D.: Educating the fighter: buttonmashing, seeing, being. On the Horizon **13**(2), 75–88 (2005)
15. Bahng, S., Kelly, R.M., McCormack, J.: Reflexive VR storytelling design beyond immersion: facilitating self-reflection on death and loneliness. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2020)
16. Jiang, J., Ahmadpour, N.: Beyond immersion: designing for reflection in virtual reality. In: Proceedings of the 33rd Australian Conference on Human-Computer Interaction, pp. 208–220 (2021)
17. Knibbe, J., Schjerlund, J., Petraeus, M., Hornbæk, K.: The dream is collapsing: the experience of exiting VR. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2018)
18. Kaplan, A.D., Cruit, J., Endsley, M., Beers, S.M., Sawyer, B.D., Hancock, P.A.: The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. Hum. Fact. **63**(4), 706–726 (2021)
19. Zuo, T., Hu, J., van der Spek, E., Birk, M.: Understanding the effect of fantasy in augmented reality game-based learning from a player journey perspective. In: Proceedings of the Eleventh International Symposium of Chinese CHI, pp. 55–60 (2023)
20. de Wit, J.: A unified model of game design, through the lens of user experience. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–6 (2021)
21. Papadopoulos, T., Evangelidis, K., Kaskalis, T.H., Evangelidis, G., Sylaiou, S.: Interactions in augmented and mixed reality: an overview. Appl. Sci. **11**(18), 8752 (2021)
22. Malinverni, L., Valero, C., Schaper, M.M., Pares, N.: A conceptual framework to compare two paradigms of augmented and mixed reality experiences. In: Proceedings of the 17th ACM Conference on Interaction Design and Children, pp. 7–18 (2018)
23. Steffen, J.H., Gaskin, J.E., Meservy, T.O., Jenkins, J.L., Wolman, I.: Framework of affordances for virtual reality and augmented reality. J. Manag. Inf. Syst. **36**(3), 683–729 (2019)

24. Rogers, Y., Scaife, M., Gabrielli, S., Smith, H., Harris, E.: A conceptual framework for mixed reality environments: designing novel learning activities for young children. Presence **11**(6), 677–686 (2002)
25. Nieuwdorp, E.: The pervasive interface: tracing the magic circle. In: Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play (2005)
26. Kors, M.J., Van der Spek, E.D., Ferri, G., Schouten, B.A.: You; the observer, partaker or victim. delineating three perspectives to empathic engagement in persuasive games using immersive technologies. In: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, pp. 493–501 (2018)
27. Hunicke, R., LeBlanc, M., Zubek, R.: MDA: a formal approach to game design and game research. In: Proceedings of the AAAI Workshop on Challenges in Game AI, vol. 4, no. 1, pp. 1722–1727 (2004)
28. Zuo, T., Feijs, L., Van der Spek, E.D., Hu, J.: A classification of fantasy in serious games. In: Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, pp. 821–828 (2019)
29. Kleinman, E., Carstensdottir, E., Seif El-Nasr, M.: A model for analyzing diegesis in digital narrative games. In: Interactive Storytelling: 12th International Conference on Interactive Digital Storytelling, ICIDS 2019 Proceedings 12, pp. 8–21. Springer International Publishing. (2019)
30. Prestopnik, N.R., Tang, J.: Points, stories, worlds, and diegesis: comparing player experiences in two citizen science games. Comput. Hum. Behav. **52**, 492–506 (2015)
31. van der Stappen, A., Liu, Y., Xu, J., Yu, X., Li, J., Van Der Spek, E.D.: MathBuilder: a collaborative AR math game for elementary school students. In: Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, pp. 731–738 (2019)
32. De Kort, Y.A., IJsselsteijn, W.A., Poels, K.: Digital games as social presence technology: development of the Social Presence in Gaming Questionnaire (SPGQ). In: Proceedings of PRESENCE, 195203, pp. 1–9 (2007)
33. Zuo, T., Birk, M.V., Van der Spek, E.D., Hu, J.: The effect of fantasy on learning and recall of declarative knowledge in AR game-based learning. Entertain. Comput. **46**, 100563 (2023)

The Influence of User Experience Satisfaction in VR Serious Games: Flow Experience and Self-efficacy as Mediating Effects

Qi Xiao(✉)

Royal College of Art, School of Communication, London, UK
xiao_cirilla@163.com

Abstract. VR serious games are interactive entertainment experiences with educational purposes, and the application of training memory and cognition has expanded in a range of health and academic settings in recent years. However, there is still little literature on the interaction between user experience characteristics and user satisfaction in serious VR memory training games. This study discussed in detail the relationship between user experience characteristics and user satisfaction with VR serious games, focusing on the influence of the game's overall experience on the learning effect and satisfaction from the user's perspective. The study developed a theoretical model through a literature review that identified seven key underlying factors: play experience, learning experience, adaptability, fidelity, flow experience, self-efficacy, and satisfaction. Based on the stimulus-organic-response (SOR) paradigm, this study conducted online and offline experiments, inviting 320 respondents to use VR serious games as stimuli and, finally, 315 valid data. The results show that game experience, learning experience, and fidelity directly and significantly affect user satisfaction. However, adaptability has no substantial effect on satisfaction. Flow experience and self-efficacy are important mediating effects of VR profound game experience and profoundly impact user satisfaction. The results of this study further emphasize the user experience characteristics of serious VR games, especially the importance of the gaming experience, learning experience, and fidelity to improve user satisfaction. It also emphasizes the importance of serious VR games in satisfying flow experiences and self-efficacy. In general, this study lays a foundation for qualitative analysis and research on the types of severe game applications in VR. It also provides new insights into the future design and development of serious games in VR immersive environments, emphasizing the importance of optimizing user experience and flow experience to improve user satisfaction.

Keywords: VR serious game · memory perception · satisfaction · user experience · SOR model · PLS-SEM

1 Introduction

The term serious games was coined as early as 1970 (Abt 1987). Although serious games are entertaining and playful, they are primarily educational with a "serious" purpose centered on training, skill development, education, or attitudinal behavior change (Ge

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 41–62, 2025.
https://doi.org/10.1007/978-3-031-93715-6_4

and Ifenthaler 2018). In recent years, serious games have been widely used in education (Agbo et al. 2023), health care (Wang et al. 2022), clinical psychology (Dewhirst, Laugharne, and Shankar 2022), and many other fields. Research has shown that serious games potentially benefit psychological and behavioral change or symptom relief, especially in mental cognition. Serious games make therapy more engaging, potentially increasing the impact of digital and in-person interventions in health through edutainment and gamification by providing learning opportunities, safe skill drills, and alternative therapeutic experiences during play (Fleming et al. 2016). Journey to Wild Divine (Wikipedia 2023), for example, utilizes biofeedback and video game systems to promote stress management and overall health through breathing, meditation, and relaxation exercises. By engaging in visual processing and memory processes, visual-spatial games such as Tetris (Kessler et al. 2020) may help reduce intrusive memory' effects. Studies have proved that serious games can effectively interfere with cognitive and memory functions through mental training (Al-Thaqib et al. 2018).

With the development of technology, the future of serious games combined with virtual reality (VR) environments is very bright (Checa and Bustillo 2020). Research shows that an immersive virtual reality environment can provide users with more immersive, private, flexible, safe and other psychological treatment environments or characteristics (Li 2022). VR can provide better depth perception than 2D or even 3D games, and VR's "immersive" nature can improve the efficiency of learning, training, psychotherapy, and other aspects (Grendelgames 2024). For example, the Ayahuasca VR game's (Astrea 2020) psychedelic experience in virtual reality invites users to experience and meditate on a simulated spiritual journey. Fujii (Funktronic Labs 2019) gives users a healing experience from the game through visual and musical wonderland. At the same time, using the body to participate in serious VR games can enhance players' positive emotions and reduce negative emotions and anxiety (Pallavicini and Pepe 2020). Research shows that VR serious game training models to enhance cognitive abilities are still evolving (Liang and Dong 2022). In short, the application and research of serious VR games in support of health and learning continues to expand.

From a user usage perspective, user experience goes beyond application usability and is at the heart of technology usability (Mccarthy and Wright 2004). At the same time, user experience is a key factor in the development and use of computer games, and a high-quality and practical user experience determines the acceptability of digital games (Jakubowski 2015). Research has shown that user experience satisfaction and usage intent in serious games affects application effectiveness (Espinosa-Curiel et al. 2020, Cohard 2019), and the balanced design of serious games is a key factor in learning effectiveness (Victoria and Marian 2012). However, most research has focused on topics and characteristics related to serious games themselves, and only a few studies have discussed the psychological factors and their impact on users during the game experience (Heiden et al. 2019). To fill a gap in this area, this study explores how the user experience of serious games in VR affects satisfaction and usage effectiveness. The results provide insights for game designers and user experience designers.

The research framework of this paper is as follows: First, literature review, introducing the theoretical background and research progress involved in this study; Second, research methods, put forward the theoretical framework and hypothesis development;

Third, data collection and analysis; Fourth, research results display and discussion; Finally, the significance of the research results to the field of VR serious games are summarized and discussed.

2 Literature Review and Hypothesis

2.1 Stimulus-Organ-Response (SOR) Theory

The stimulus-organic-response (SOR) model, first proposed by American psychologists Mehrabian and Russell (Mehrabian and Russell 1974), consists of three parts: S (Stimulus) refers to the properties of the system environment; O (Organism) refers to an individual's emotional and cognitive state; R (Response) refers to an individual's behavioral response (Qiu et al. 2024). This model was initially used in the cognitive framework of environmental psychology and is now widely used in studying user behavior between the environment and people in mobile applications and information systems (Hlee et al. 2022, Lee and Chen 2022). Numerous studies have shown that the unique nature of product interaction in VR games can affect players' psychological responses, affecting their continued intent and satisfaction (Fromm et al. 2021, Jin et al. 2021, Checa and Bustillo 2020). In the VR environment, Nguyen et al. 's research shows (Nguyen, Le, and Chau 2023) that the stimulation of VR (vividness and interactivity) can trigger the flow response of users, which ultimately affects the satisfaction with VR (Nguyen, Le, and Chau 2023). In the study of Han et al. (Han, Kim, and An 2023), vividness, interactivity, and control constitute the stimulus environment; Presence and playfulness are seen as organisms; User intent is considered a reaction.

Based on the literature, this study developed a research model based on the extended SOR framework (see Fig. 1). In this study, interactive attributes (S) of serious VR games include game experience, learning experience, adaptability, and fidelity, which will affect users' emotional and cognitive psychological activities, such as flow experience and self-efficacy (O), and ultimately affect users' satisfaction (R).

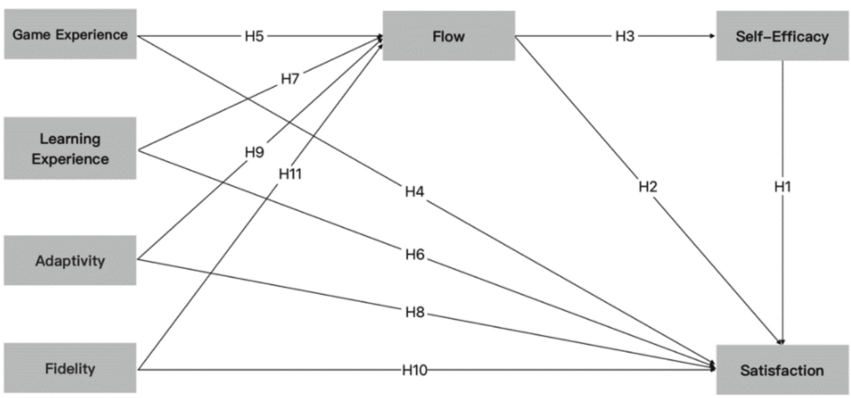


Fig. 1. Research model and hypotheses.

2.2 Satisfaction

The satisfaction of VR serious games is the focus of current research. Satisfaction refers to the user's feelings after using serious VR games and is related to the user's tendency to participate in serious VR games again, which is different from the initial willingness to accept (Qiu et al. 2024). Research has shown that satisfaction with serious VR games is directly related to user happiness and performance (Checa, Miguel-Alonso, and Bustillo 2023). Guldager et al. (Guldager et al. 2023) used quantitative research methods to demonstrate that user experience and satisfaction in VR games are closely related to the effectiveness of serious games. In recent years, many experts and scholars have studied various factors that affect user satisfaction after using VR serious games, including demographic characteristics (such as age and gender), design characteristics of VR serious games (such as narrative, creativity, sound aesthetics, visual aesthetics, etc.) and interactive characteristics (such as game experience, learning experience, adaptability, fidelity, etc.) (Shelstad, Smith, and Chaparro 2017, Moizer et al. 2019). Dumblekar et al.'s (Dumblekar, Antony, and Dhar 2024) research shows that in simulation games, player satisfaction consists of excitement, challenge, learning experience, team victory, and self-discovery.

2.3 Self-efficacy

Self-efficacy is an individual's belief in his or her ability to complete a task or achieve a goal and links an individual's engagement (MSEd 2024), learning, and well-being to the type of experience to which he or she is exposed (Kuznetcova et al. 2023). Experiencing self-efficacy and success in the game can satisfy the user's ability needs (Reer et al. 2022). There are many reasons to combine self-efficacy theory with studying serious games in VR. First, VR serious games provide unique ways of interacting to change and adjust the perception of task difficulty, motivation, and point of control, all closely related to self-efficacy. Second, numerous studies have confirmed that serious VR games can improve users' self-efficacy (Kuznetcova et al. 2023, Perlwitz and Stemmann 2022, Power, Lynch, and McGarr 2020). For example, Ali Mousavi et al. (Mousavi et al. 2023) examined how self-efficacy and user simulation performance can predict learning gains in serious VR games. Several studies consistently highlight that serious games can be further improved to improve users' self-efficacy and associated positive outcomes. This paper proposes the following hypotheses:

H1: Self-efficacy positively affects user satisfaction in serious VR games.

2.4 Flow Experience

Psychologist Csikszentmihalyi, who first proposed the concept of "flow experience" in the 1970s, is a mental state that combines cognitive, physical, and emotional aspects (Biasutti 2017). Flow describes an optimal state of mind in which the user fully engages in an activity and experiences high focus, control, and enjoyment (Bodzin et al. 2020). This is directly related to beneficial job-related outcomes, such as improved performance, energy levels, or creativity (Demerouti et al. 2012, Engeser and Rheinberg 2008). Research has shown that flow experiences are enhanced when users perform gaming tasks in a virtual environment (Schaik and Vallance 2012). When players experience pleasure and satisfaction in serious VR games, it stimulates their creativity and imagination and gives them intense satisfaction with the game simulation (Alrehaili and Osman 2019). At the same time, the flow effect can increase learning motivation, which in turn helps promote immersive learning (Paras and Bizzocchi 2005). Fang's (Fang 2024) research found that the virtual reality interaction mode performed well regarding flow experience and acceptance, triggering strong positive emotions. Emotions can influence a range of human attributes, such as confidence, self-efficacy, and attitude, and can be a learning outcome in serious games (Garris, Ahlers, and Driskell 2002). Therefore, this paper proposes the following hypothesis:

H2: Flow experience positively affects user satisfaction in serious VR games.
H3: Flow experience positively affects user self-efficacy in serious VR games.

2.5 Game Experience

Game experience refers to the one-to-one relationship between the player and the game, including the feelings and engagement generated during play (Caroux 2023). Fluency, immersion, emotion, challenge, and skill development are at the heart of the game experience, and the flow experience generated during the process is primarily related to learning through the game (Hamari et al. 2016). Virtual environments can provide a multi-dimensional gaming experience, especially in terms of enhanced immersion and presence (Yao and Kim 2019). Immersion in virtual reality games has a positive impact on learning (Thompson et al. 2021). Martinez et al.'s (Martinez, Menéndez, and Bustillo 2020) research shows that game experience is directly related to user satisfaction, and the higher the user satisfaction in VR serious games, the higher the game experience is promoted. Therefore, this paper proposes the following hypothesis:

H4: Gaming experience positively affects user satisfaction in serious VR games.
H5: Gaming experience positively affects the flow experience of serious VR game users.

2.6 Learning Experience

Serious games are games with educational purposes, in which the learning experience is closely related to the effectiveness of serious games (Cook et al. 2012). Feedback in the game experience enables participants to learn and reflect on the experience, thereby creating integrated knowledge in the game that can then be applied in the real world (Dzeng, Lin, and Wang 2014). Dumblekar et al. (Dumblekar, Antony, and Dhar 2024) studied the learning experience as an important component of player satisfaction in simulation games. Therefore, this paper proposes the following hypothesis:

H6: Learning experience positively affects user satisfaction in serious VR games.
H7: The learning experience positively affects the flow experience of serious VR game users.

2.7 Adaptability

In the context of a game, adaptation describes the automatic adaptation of game elements, such as content, user interface, game mechanics, game difficulty, etc., to customize or personalize interactive experiences (Streicher and Smeddinck 2016). Harefa et al. (Nifataro et al. 2024) have shown that adaptive adjustment of each player's performance and mood changes the game's difficulty in real-time, creating a more personalized and engaging gaming experience, thereby increasing player satisfaction. Ceja, Fullagar, et al. (Ceja and Navarro 2011, Fullagar and Kelloway 2009) believe that individuals differ in their tendency to experience flow, and flow states also depend on tasks, situations, and time. Hence, the adaptability of serious games is an important factor in enhancing users' flow experience (Karen et al. 2024). Therefore, this paper proposes the following hypothesis:

H8: Adaptability positively affects user satisfaction in serious VR games.
H9: Adaptability positively affects the flow experience of serious VR game users.

2.8 Fidelity

Fidelity in serious games is related to the level of realism in the virtual environment provided by the user, and high fidelity is considered an important factor when transferring knowledge learned in games to the real world (Moizer et al. 2019). Game design and development with assurance in mind can improve learning in serious games and thus enhance user satisfaction (Ye et al. 2019, Rooney 2012). Ye et al. 's (Ye et al. 2019) research shows that higher fidelity leads to better user satisfaction and thus performance in augmented reality environments. For example, Peschel et al.' s (Anne et al. 2024) research proves that higher visual fidelity can improve explicit memory performance. High-fidelity gaming environments provide individuals with a more immersive experience that enhances their sense of flow, while at the same time positively impacting performance, learning and engagement (Perttula et al. 2024). Therefore, this paper proposes the following hypothesis:

H10: Fidelity positively affects user satisfaction in serious VR games.
H11: Fidelity positively affects the flow experience of VR serious game users.

3 Methodology

3.1 Data Collection and Sample Characteristics

To scientifically verify the theoretical framework proposed in this paper, 320 respondents were invited to conduct a questionnaire survey after experiencing serious VR games online and offline. Data collection is divided into the following: first, players play serious games online (if they have a device) or offline in Shanghai; second, respondents fill out questionnaires to record their experience. The questionnaire was recruited through the online platform, and 320 responses were collected. Among them, attention test questions were set in the questionnaire to ensure the quality of data collection. Finally, 315 valid samples were collected for the final evaluation and analysis. In this study, the SOR model is adopted, and serious games as external stimuli affect the emotions, perceptions, and attitudes of respondents during the experience, thus triggering specific behavioral responses, namely, filling out questionnaires.

3.2 Measurement Development

This study selects ADAPTS measures from the classical literature in the past to suit the background of this paper. The study involved seven measurement items: play experience (GE), Learning Experience (LE), adaptability (ADA), fidelity (FID), flow experience (FE), self-efficacy (SE), and satisfaction (SAT). The scale of game experience is mainly based on the scale of Moizer J. et al. and Qiu et al. (Moizer et al. 2019), Qiu et al. 2024). The learning experience is mainly based on the scale of Lopez et al., Chambilla et al., and Rosenthal et al. (López et al. 2021, Chambilla et al. 2020). Adaptability is mainly based on the scale of Streicher et al. and Lopes et al. (Streicher and Smeddinck 2016, Lopes and Bidarra 2011). Fidelity is based on the scale of Ye et al. and Chambilla et al. (Ye et al. 2019, Chambilla et al. 2020). The flow experience is based on the assessment scale of Tenenbaum et al. (Tenenbaum, Fogarty, and Jackson 1999). The Linares et al.

Table 1. List of constructs and their items.

Constructs	Items	Source
Game Experience	GM1: (Immersive) VR serious games give me an immersive experience that I'm always focused on	Modified from (Moizer et al. 2019, Qiu et al. 2024)
	GM2: The overall experience of the (emotional) game was positive, and the time I spent in the VR serious game was enjoyable	
	GM3: I think VR serious games in general look very appealing	
	GM4: When playing serious games in VR, the interaction is intuitive and clear, and the overall experience is positive	
Learning Experience	LM1: VR serious games have clear goals, and from start to finish, I understand the tasks to be accomplished in the simulation	Modified from (López et al. 2021, Chambilla et al. 2020)
	LM2: Playing serious games in VR helps me pick up skills faster	
	LM3: VR serious games can help me improve my learning efficiency in reality	
	LM4: The VR serious game experience is rich in content, and I can learn and gain new things from it	

(*continued*)

Table 1. (*continued*)

Constructs	Items	Source
Adaptivity	ADA1: The VR serious game system is not fixed, and can adapt to the needs of users and changes in the environment	Modified from (Streicher and Smeddinck 2016, Lopes and Bidarra 2011)
	ADA2: VR serious games can integrate player data and provide personalized customization based on my background, needs and goals	
	ADA3: In general, serious VR games can trigger real-time adaptive interventions based on my actions, such as repeating tasks, increasing or decreasing difficulty levels, and adjusting the speed of the game	
Fidelity	FID1: The interface of VR serious games is very realistic, and the game system allows me to experience the sensory experience in the real world	Modified from (Ye et al. 2019, Chambilla et al. 2020)
	FID2: VR serious games have a good resemblance to reality	
	FID3: VR serious games encourage me to use cognitive processes that I would use in the real world	
Flow	FE1: I enjoy the VR memory training serious game experience, the whole process is very rewarding	Modified from (Tenenbaum, Fogarty, and Jackson 1999)
	FE2: During the game, I have a clear goal and I know what I want to achieve	
	FE3: During the game, I can feel that the ability matches the challenge and I have the ability to meet the requirements	

(*continued*)

Table 1. (*continued*)

Constructs	Items	Source
	FE4: In a VR memory training serious game, I have complete control over my own behavior, completely focus on what is happening, and don't care about others. When playing serious games in VR, my confidence is very high	
	FE5: In the game, my behavior and consciousness merge, and I know how well I'm obviously doing by my performance	
	FE6: During the game, I was completely immersed in it and couldn't feel the passing of time	
Self-Efficacy	SE1: Using VR serious games has improved my sense of ability to do things	Modified from (Linares et al. 2021)
	SE2: Using VR serious games has improved my sense of ability to achieve complex goals	
	SE3: Using VR serious games allows me to do activities as I wish	
	SE4: My confidence level is very high when playing serious games in VR	
Satisfaction	SAT1: This game simulation is enjoyable	Modified from (Chambilla et al. 2020)
	SAT2: I like to use this game for training and learning	
	SAT3: Overall, I'm happy with the simulation	

scale was considered for self-efficacy (Linares et al. 2021). Satisfaction is mainly based on the scale of Chambilla et al. (Chambilla et al. 2020). A 5-point Likert scale was used to measure items, ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Details of the project content and references can be found in the Table 1.

4 Analysis and Results

4.1 Measurement Model

This paper uses structural equation modeling technology and Smart PLS 4.0 software package to analyze the sample data by partial least squares (PLS). PLS analysis was chosen because it combines comprehensive development functions such as exploratory research and theoretical construction (Ramli, Latan, and Nartea 2024). According to the research suggestions of Hair et al. and Kurtaliqi et al. (Kurtaliqi et al. 2024), this paper adopts the following methods for PLS analysis: analyzing the measurement model and analyzing the structural model. Cronbach's alpha (α), composite reliability (rho_a and rho_c), and mean-variance extraction (AVE) were used to evaluate the reliability and validity of the construction, and the reliability, convergence validity, and discriminant validity of the model variables were evaluated.

In the analysis of the measurement model stage, in the reliability test process, according to the previous scientific research recommendations, the α value should be >0.70 (Hair, Anderson, and Tatham 1986); Internal consistency Cronbach's Alpha > 0.70 (Fornell and Larcker 1981); Composite Reliability CR > 0.70 (Hair, Anderson, and Tatham 1986); factor loadings >0.70 and cross-loadings (Nunnally 1994); average variance extracted (AVE) > 0.50 (Henseler, Ringle, and Sinkovics 2009). As shown in the Table 2, Cronbach's alpha value and CR value of all factors in the conceptual model proposed in this study were between (0.80–1.00) and (0.80–1.00), which obviously exceeded the critical value of 0.70, so the data met the test standard. In the process of discriminant validity, according to the research suggestion of Fornell et al. (Fornell and Larcker 1981), the square root of AVE should exceed its highest correlation coefficient with items in different structures (see Table 3). Cross-loading results of items in this study also show that the model has good discriminant validity. In summary, all the results prove that the structure of this study is reliable and meets the acceptable criteria of internal consistency reliability, item reliability, convergence validity, and discriminant validity, indicating that the structure proposed in this paper is appropriate and the measurement model is effective.

4.2 Structural Modeling and Hypothesis Testing

In evaluating the construction model, bootstrapping technology was adopted in SmartPLS4.0 to generate 5000 random samples to test the structural model. As suggested by Hair et al. (Hair Jr. et al. 2021), the evaluation takes into account the path coefficient (β), the significance level of the structural model effect (P-value), the variance interpretation ratio (R^2), and Stone-Geisser's $\mathbf{Q^2}$ values. At the same time, Hair et al. (Hair Jr. et al. 2021) suggest that r-squared values should exceed the minimum acceptable level of 0.10 with $R^2 > 0.20$ for good explanatory power. As shown in the Fig. 2, R^2 calculated by the PLS algorithm above is satisfaction (0.543), self-efficacy (0.294), and flow experience (0.470) among the three variables, which all exceed the threshold value of 0.20 and have adequate explanatory power (Ajamieh et al. 2016).

In terms of direct impact on satisfaction, flow experience ($\beta = 0.197$, $p < 0.001$) had the largest impact, followed by self-efficacy ($\beta = 0.184$, $p < 0.001$), gaming experience

Table 2. Scales for reliability and validity of measurement model.

Construct	Item	M	SD	FL	α	CR	AVE
Game Experience	GE1	3.771	1.074	0.820	0.877	0.897	0.731
	GE2	3.387	1.043	0.843			
	GE3	3.419	1.000	0.862			
	GE4	3.438	1.020	0.894			
Learning Experience	LE1	3.721	0.861	0.845	0.864	0.873	0.709
	LE2	3.746	0.861	0.852			
	LE3	3.657	0.958	0.863			
	LE4	3.803	0.901	0.807			
Adaptivity	ADA1	3.559	0.966	0.871	0.850	0.855	0.770
	ADA2	3.502	1.052	0.907			
	ADA3	3.686	0.862	0.853			
Fidelity	FID1	3.644	0.851	0.862	0.858	0.865	0.778
	FID2	3.667	0.943	0.885			
	FID3	3.692	0.974	0.898			
Flow	FE1	3.686	0.862	0.809	0.904	0.905	0.677
	FE2	3.689	0.928	0.822			
	FE3	3.648	0.932	0.827			
	FE4	3.794	0.962	0.812			
	FE5	3.727	0.892	0.871			
	FE6	3.776	1.046	0.794			
Self-Efficacy	SE1	3.749	0.938	0.860	0.885	0.888	0.744
	SE2	3.727	0.957	0.839			
	SE3	3.781	0.982	0.883			
	SE4	3.648	0.901	0.867			
Satisfaction	SA1	3.765	1.085	0.857	0.867	0.867	0.790
	SA2	3.717	1.163	0.898			
	SA3	3.714	1.155	0.911			

Note: M = mean; SD = standard deviation; FL = factor loading; α = Cronbach's alpha; CR = composite reliability; AVE = average variance extracted.

($\beta = 0.152$, $p < 0.05$), and learning experience ($\beta = 0.150$, $P < 0.001$). $p < 0.05$) comparable. However, adaptability ($\beta = 0.103$, $p > 0.05$) showed that this variable had no direct significant effect on satisfaction. For the direct effect of flow experience, gaming experience ($\beta = 0.247$, $p < 0.001$) and learning experience ($\beta = 0.230$, $p < 0.01$) had more significant positive effects, followed by fidelity ($\beta = 0.194$, $p < 0.05$). Adaptability ($\beta = 0.168$, $p < 0.05$) had the weakest effect in this respect. Meanwhile, flow experience

Table 3. Discriminant validity and Correlation Matrix

Construct	FID	LE	FE	GE	SA	SE	ADA
FID	0.882						
LE	0.570	0.842					
FE	0.555	0.573	0.823				
GE	0.571	0.577	0.579	0.855			
SA	0.576	0.592	0.603	0.591	0.889		
SE	0.536	0.573	0.542	0.559	0.592	0.862	
ADA	0.537	0.542	0.527	0.533	0.544	0.535	0.877

Note: Bold-faced diagonal elements are the square roots of AVEs. The off-diagonal elements are the correlations between constructs. FID = Fidelity, LE = Learning Experience, FE = Flow, GM = Game Experience, SA = Satisfaction, SE = Self-Efficacy, ADA = Adaptivity.

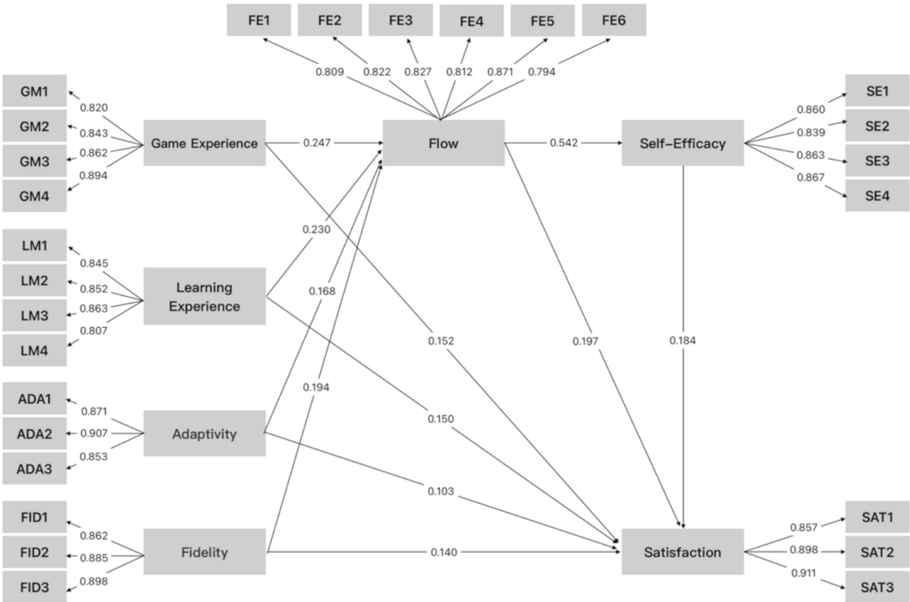


Fig. 2. Structural Model Results.

($\beta = 0.542$, $p < 0.001$) directly and significantly affected self-efficacy. In summary, most of the hypotheses (H1, H2, H3, H4, H5, H6, H7, H9, H10) are valid, only hypothesis H8 is not supported (Table 4).

4.3 Mediation Effect Test

The paper went on to examine the mediating variables to validate the structural relationships of the model and clarify the direct and indirect relationships between the variables.

Table 4. The results of hypothesis testing.

Hypothesis	Path	Dependent Variable	R^2	ß	P Value	Hypothesis Supported
H11	FID → FE	FE	0.470	0.194	0.002**	Yes
H10	FID → SA	SA	0.543	0.140	0.022**	Yes
H7	LE → FE	FE	0.470	0.230	0.000***	Yes
H6	LE → SA	SA	0.543	0.150	0.012*	Yes
H2	FE → SA	SA	0.543	0.197	0.001**	Yes
H3	FE → SE	SE	0.294	0.542	0.000***	Yes
H5	GE → FE	FE	0.470	0.247	0.000***	Yes
H4	GE → SA	SA	0.543	0.152	0.022*	Yes
H1	SE → SA	SA	0.543	0.184	0.001**	Yes
H9	ADA → FE	FE	0.470	0.168	0.005**	Yes
H8	ADA → SA	SA	0.543	0.103	0.072	No

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

It involved a total of 5000 guide samples to estimate the significance of the mediating effects more accurately. This study examines the mediating effects of flow experience and self-efficacy on the relationship between the design elements of VR serious game interactive experience (game experience, learning experience, adaptability, fidelity) and user satisfaction. The Bootstrap method proposed by Shrout et al. (Shrout and Bolger 2002) was wholly removed for evaluation. The Table 5 detailed the results of the mediating effect between variables. The research results showed that flow experience was an important mediating factor connecting the interactive experience of serious VR games with user satisfaction. In particular, gaming experience and learning experience have significant effects on satisfaction through flow experience and self-efficacy, respectively (learning experience → flow experience → self-efficacy → satisfaction; $\beta = 0.023$, $p < 0.05$) (game experience → flow experience → Self-efficacy → satisfaction; $\beta = 0.025$, $p < 0.05$). It is worth mentioning that adaptability has significant indirect effects on both self-efficacy and satisfaction through flow experience (adaptability → flow experience → self-efficacy; $\beta = 0.091$, $p < 0.05$) (adaptability → Flow experience → satisfaction; $\beta = 0.033$, $p < 0.05$); However, adaptability had no significant indirect effect on satisfaction through the mediation effects of flow experience and self-efficacy (adaptability → flow experience → self-efficacy → satisfaction; $\beta = 0.0017$, $p > 0.05$). In general, flow experience and self-efficacy play an important mediating role in the design attributes of VR serious game interactive experience on user experience satisfaction and are important factors.

Table 5. Mediating effect test.

Indirect effects	Path	Dependent Variable	ß	P Value
Partial mediation effect	FE → SE → SA	SA	0.099	0.002
Partial mediation effect	FID → FE → SA	SA	0.038	0.038
Partial mediation effect	FID → FE → SE	SA	0.105	0.003
Partial mediation effect	LE → FE → SA	SA	0.045	0.020
Partial mediation effect	LE → FE → SE	SA	0.124	0.000
Partial mediation effect	GE → FE → SA	SA	0.049	0.010
Partial mediation effect	GE → FE → SE	SA	0.134	0.000
Partial mediation effect	ADA → FE → SA	SA	0.033	0.0035
Partial mediation effect	ADA → FE → SE	SA	0.091	0.006
Partial mediation effect	FID → FE → SE → SA	SA	0.019	0.027
No mediation effect	ADA → FE → SE → SA	SA	0.017	0.058
Partial mediation effect	LE → FE → SE → SA	SA	0.023	0.023
Partial mediation effect	GE → FE → SE → SA	SA	0.025	0.012

5 Discussion

The intervention of VR serious games in the field of digital health continues to expand, and the application of VR serious games is developing rapidly. However, most of the current research focuses on the theme and design features related to serious games themselves, and there is still little literature on the interaction between user experience characteristics and user satisfaction with serious VR games. To fill the gap in this area, this paper focuses on the impact of the overall experience of a game on learning (self-efficacy) and satisfaction from the user perspective because, from the user use perspective, user experience goes beyond the usability of serious games, and user experience satisfaction and usage intention affect the effectiveness of an application. This study not only verifies the direct impact of VR serious game user experience elements (game experience, learning experience, fidelity, adaptability) on user satisfaction but also discusses the important impact of flow experience and self-efficacy on user satisfaction in VR serious games and how user experience elements indirectly affect user satisfaction through flow experience and self-efficacy. The detailed discussion results are as follows.

First, this study confirms the significant impact of flow experience and self-efficacy on user satisfaction in serious VR games (H1, H2). This result is consistent with research on forgetting (Lemmens and Münchhausen 2023, Kim and Ko 2019, Reer et al. 2022), and our results also show that flow experiences and self-efficacy increase user satisfaction. Specifically, when users enjoy a more substantial flow experience during the experience, it will bring a higher sense of pleasure and trigger a stronger emotional response, thus enhancing the satisfaction of the user experience (Fang 2024, Kim and Ko 2019). This result is because VR devices' sense of presence and telepresence support the flow experience (Coelho et al. 2006, Shelstad, Smith, and Chaparro 2017). Higher flow experience is closely related to users' self-efficacy and satisfaction in virtual environments.

Secondly, the results show that when VR serious games have a better game experience, users' flow experience and satisfaction will be higher, and the game experience of VR serious games is positively correlated with flow experience and satisfaction (H4, H5). This result is consistent with the findings of Hamari et al. (Hamari et al. 2016), who suggest that when the game experience of a serious game in VR is more challenging, the user can continue to grow in ability and learning. This way of influencing learning through increased game engagement leads to a better flow experience and satisfaction. Previous research has established that games provide a strong sense of presence in the gaming experience, which promotes a player's flow experience. A more substantial flow experience is associated with better performance, higher physiological arousal, and more fun and user satisfaction (Lemmens and Münchhausen 2023). Overall, gaming experiences in serious VR games are positively correlated with flow experiences and satisfaction.

Third, the learning experience in serious VR games is also an important factor affecting user satisfaction, and our hypotheses H6 and H7 have been confirmed. Checa and Bustillo's study proposed that trainees feel they can control the interactive learning process in serious games, promoting active and critical learning. VR serious games improve user experience, enhancing knowledge acquisition ability (Checa and Bustillo 2020). Previous research has also demonstrated that VR technology can effectively enhance user satisfaction by increasing gaming enjoyment through natural mapping and resulting satisfaction with the need for ability and learning experience (Reer et al. 2022).

Fourth, the fidelity of serious VR games is positively correlated with flow experience and satisfaction (H10, H11). Lowell et al.'s (Lowell and Tagare 2023) research also proves that high-fidelity experiences enable users to self-regulate, increase self-efficacy, and enhance satisfaction. In addition, high-fidelity VR serious games represent a strong fidelity of interaction and scenes, and the fidelity of virtual scenes significantly enhances the sense of presence and usability, thereby enhancing users' flow experience (Luo et al. 2023). In short, fidelity plays an important role in the user experience of serious games in VR.

Finally, this study found that adaptability positively correlates with flow experience in serious VR games, supporting H9 but not directly related to satisfaction, contrary to our hypothesis H8. However, previous research has suggested that greater adaptability and fluency evoke better user performance and enjoyment user performance and enjoyment (Lemmens and Münchhausen 2023). However, with the satisfaction results, it may be because when users use VR, adaptability may be the latter consideration in interactive experience. However, in any case, the impact of adaptability on the flow experience is well established.

6 Theoretical and Practical Significance

First, this paper provides a theoretical contribution to the impact of serious VR games on user satisfaction in interactive experiences. In previous studies, this topic mainly focused on the review topics and features related to serious games themselves, and there is still little literature on the interaction between user experience characteristics and user satisfaction in serious games of VR memory training. Especially from the designers' perspective, it is imperative to establish a model of VR serious game interactive experience design elements to evaluate the overall user experience. Therefore, starting from the interactive experience design elements of VR serious games, this paper further analyzes and studies the user satisfaction model of this topic through the SOR theoretical model. The model in this paper includes the features of interaction design elements, such as game experience, learning experience, fidelity, and adaptability, and further analyzes the flow experience and self-efficacy characteristics unique to VR devices to analyze better and explain the experience process and feelings of users when using VR serious games. In addition, this article provides practical guidance on interaction design for designers and developers of serious VR games. For interaction designers, game experience, learning experience, and fidelity are important factors that enhance user satisfaction. Designers can consider these factors in interaction design, process design, and other related aspects.

7 Limitations

This article still has some limitations. This paper only covers the user group in China, which may have some potential problems. In future research, appropriate consideration should be given to expanding the research scope and comprehensively considering the differences in the influence of different regions, nationalities, and other cultural differences on the satisfaction of VR serious game interactive experience. Secondly, the data in this study are all cross-sectional, and longitudinal data can enrich the research on the effectiveness of VR's serious game interactive experience design on user satisfaction.

8 Conclusion

The results of this study further emphasize the user experience characteristics of serious VR games, especially the importance of the gaming experience, learning experience, and fidelity to improve user satisfaction. It also emphasizes the importance of serious VR games in satisfying flow experiences and self-efficacy. In general, this study lays a foundation for qualitative analysis and research on the types of severe game applications in VR. It also provides new insights into the future design and development of serious games in VR immersive environments, emphasizing the importance of optimizing user experience and flow experience to improve user satisfaction.

Ethical Review. This article has obtained the informed consent of participants and ensured data privacy without any ethical concerns.

References

1. Abt, C.C.: Serious Games. University Press of America (1987)
2. Ge, X., Ifenthaler, D.: Designing engaging educational games and assessing engagement in game-based learning. In: Gamification in Education: Breakthroughs in Research and Practice, pp. 1–19. IGI Global (2018). https://doi.org/10.4018/978-1-5225-5198-0.ch001
3. Agbo, F.J., Olaleye, S.A., Bower, M., Oyelere, S.S.: Examining the relationships between students' perceptions of technology, pedagogy, and cognition: the case of immersive virtual reality mini games to foster computational thinking in higher education. Smart Learn. Environ. **10**(1), 16 (2023). https://doi.org/10.1186/s40561-023-00233-1
4. Wang, Y., Wang, Z., Liu, G., et al.: Application of serious games in health care: scoping review and bibliometric analysis. Front. Public Health **10**, 896974 (2022). https://doi.org/10.3389/fpubh.2022.896974
5. Dewhirst, A., Laugharne, R., Shankar, R.: Therapeutic use of serious games in mental health: scoping review. BJPsych Open **8**(2), e37 (2022). https://doi.org/10.1192/bjo.2022.4
6. Fleming, T.M., Bavin, L., Stasiak, K., et al.: Serious games and gamification for mental health: current status and promising directions. Front. Psychiatry **7** (2016) https://doi.org/10.3389/fpsyt.2016.00215
7. Wikipedia. Journey to wild divine. In: Wikipedia (2023). https://en.wikipedia.org/w/index.php?title=Journey_to_Wild_Divine&oldid=1150265384. Accessed 27 September 2024
8. Kessler, H., Schmidt, A.-C., James, E.L., et al.: Visuospatial computer game play after memory reminder delivered three days after a traumatic film reduces the number of intrusive memories of the experimental trauma. J. Behav. Ther. Exp. Psychiatry **67**, 101454 (2020). https://doi.org/10.1016/j.jbtep.2019.01.006
9. Al-Thaqib, A., Al-Sultan, F., Al-Zahrani, A., et al.: Brain training games enhance cognitive function in healthy subjects. Med. Sci. Monit. Basic Res. **24**, 63–69 (2018). https://doi.org/10.12659/MSMBR.909022
10. Checa, D., Bustillo, A.: A review of immersive virtual reality serious games to enhance learning and training. Multimed. Tools Appl. **79**(9), 5501–5527 (2020). https://doi.org/10.1007/s11042-019-08348-9
11. Li, S.: Healing the mind and body in the 'metaverse': how virtual reality is transforming the mental health circuit (2022). https://m.thepaper.cn/newsDetail_forward_16833051. Accessed 27 September 2024
12. Grendelgames. Virtual Reality (VR) for Serious Games – Grendel Games (2024). https://grendelgames.com/vr/. Accessed 27 September 2024
13. Astrea. Ayahuasca (2020). https://store.steampowered.com/app/1212940/Ayahuasca/. Accessed 27 September 2024
14. Funktronic Labs. Fujii (2019). https://store.steampowered.com/app/589040/Fujii/. Accessed 27 September 2024
15. Pallavicini, F., Pepe, A.: Virtual reality games and the role of body involvement in enhancing positive emotions and decreasing anxiety: within-subjects pilot study. JMIR Serious Games **8**(2), e15635 (2020). https://doi.org/10.2196/15635

16. Liang, H., Dong, X.: Enhancing cognitive ability through a VR serious game training model mixing Piaget's epistemological methodology and Lumosity concept. Vis. Comput. **38**(9), 3487–3498 (2022). https://doi.org/10.1007/s00371-022-02552-9
17. Mccarthy, J., Wright, P.: Technology as experience. Interactions **11** (2004). https://doi.org/10.1145/1015530.1015549
18. Jakubowski, M.: User experience as a crucial element of future simulation and gaming design. In: Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL Conference, p. 42 (2015). https://absel-ojs-ttu.tdl.org/absel/article/view/2981. Accessed 27 September 2024
19. Espinosa-Curiel, I.E., Pozas-Bogarin, E.E., Martínez-Miranda, J., Pérez-Espinosa, H.: Relationship between children's enjoyment, user experience satisfaction, and learning in a serious video game for nutrition education: empirical pilot study. JMIR Serious Games **8**(3), e21813 (2020). https://doi.org/10.2196/21813
20. Cohard, P.: Evaluation of serious game user experience: the role of emotions. Electron. J. Inf. Syst. Eval. **22**(2), 128–141 (2019). https://doi.org/10.34190/EJISE.19.22.2.005
21. Victoria, G.-N., Marian, A.-C.: Serious games and learning effectiveness: the case of It's a Deal! Comput. Educ. **58**(1), 435–448 (2012). https://doi.org/10.1016/j.compedu.2011.07.015
22. von der Heiden, J.M., Braun, B., Müller, K.W., Egloff, B.: The association between video gaming and psychological functioning. Front. Psychol. **10**, 1731 (2019). https://doi.org/10.3389/fpsyg.2019.01731
23. Mehrabian, A., Russell, J.A.: An approach to environmental psychology (1974). https://psycnet.apa.org/record/1974-22049-000. Accessed 22 August 2024
24. Qiu, T., Li, H., Chen, Y., Zeng, H., Qian, S.: Continuance intention toward VR games of intangible cultural heritage: a stimulus-organism-response perspective. Virt. Real. **28**(3), 149 (2024). https://doi.org/10.1007/s10055-024-01043-7
25. Hlee, S., Park, J., Park, H., Koo, C., Chang, Y.: Understanding customer's meaningful engagement with AI-powered service robots. Inf. Technol. People **36**(3), 1020–1047 (2022). https://doi.org/10.1108/ITP-10-2020-0740
26. Lee, J.C., Chen, X.: Exploring users' adoption intentions in the evolution of artificial intelligence mobile banking applications: the intelligent and anthropomorphic perspectives. Int. J. Bank Market. **40**(4), 631–658 (2022). https://doi.org/10.1108/IJBM-08-2021-0394
27. Fromm, J., Radianti, J., Wehking, C., Stieglitz, S., Majchrzak, T.A., vom Brocke, J.: More than experience? On the unique opportunities of virtual reality to afford a holistic experiential learning cycle. Internet High. Educ. **50**, 100804 (2021). https://doi.org/10.1016/j.iheduc.2021.100804
28. Jin, B., Kim, G., Moore, M., Rothenberg, L.: Consumer store experience through virtual reality: its effect on emotional states and perceived store attractiveness. Fashion Text. **8**(1), 19 (2021). https://doi.org/10.1186/s40691-021-00256-7
29. Nguyen, T.B.T., Le, T.B.N., Chau, N.T.: How VR technological features prompt tourists' visiting intention: an integrated approach. Sustainability **15**(6), 4765 (2023). https://doi.org/10.3390/su15064765
30. Han, S.L., Kim, J., An, M.: The role of VR shopping in digitalization of SCM for sustainable management: application of SOR model and experience economy. Sustainability **15**(2), 1277 (2023). https://doi.org/10.3390/su15021277
31. Checa, D., Miguel-Alonso, I., Bustillo, A.: Immersive virtual-reality computer-assembly serious game to enhance autonomous learning. Virt. Real. **27**(4), 3301–3318 (2023). https://doi.org/10.1007/s10055-021-00607-1
32. Guldager, J.D., Hrynyschyn, R., Kjær, S.L., Dietrich, T., Majgaard, G., Stock, C.: User experience, game satisfaction and engagement with the virtual simulation VR FestLab for alcohol prevention: a quantitative analysis among Danish adolescents. PLoS ONE **18**(5), e0286522 (2023). https://doi.org/10.1371/journal.pone.0286522

33. Shelstad, W., Smith, D., Chaparro, B.: Gaming on the rift: how virtual reality affects game user satisfaction. Proc. Hum. Fact. Ergonom. Soc. Annu. Meet. **61**, 2072–2076 (2017). https://doi.org/10.1177/1541931213602001
34. Moizer, J., Lean, J., Dell'Aquila, E., et al.: An approach to evaluating the user experience of serious games. Comput. Educ. **136**, 141–151 (2019). https://doi.org/10.1016/j.compedu.2019.04.006
35. Dumblekar, V., Antony, S.P., Dhar, U.: Openness to experience and player satisfaction in a simulation game. Simul. Gaming **55**(3), 479–501 (2024). https://doi.org/10.1177/10468781241234131
36. MSEd, K.C.: How Self Efficacy Helps You Achieve Your Goals. Verywell Mind (2024). https://www.verywellmind.com/what-is-self-efficacy-2795954. Accessed 23 October 2024
37. Kuznetcova, I., Glassman, M., Tilak, S., et al.: Using a mobile Virtual Reality and computer game to improve visuospatial self-efficacy in middle school students. Comput. Educ. **192**, 104660 (2023). https://doi.org/10.1016/j.compedu.2022.104660
38. Reer, F., Wehden, L.-O., Janzik, R., Tang, W.Y., Quandt, T.: Virtual reality technology and game enjoyment: the contributions of natural mapping and need satisfaction. Comput. Hum. Behav. **132**, 107242 (2022). https://doi.org/10.1016/j.chb.2022.107242
39. Perlwitz, P., Stemmann, J.: Flow and self-efficacy in a serious game for STEM education. In: Söbke, H., Spangenberger, P., Müller, P., Göbel, S. (eds.) Serious Games, pp. 3–16. Springer International Publishing (2022). https://doi.org/10.1007/978-3-031-15325-9_1
40. Power, J., Lynch, R., McGarr, O.: Difficulty and self-efficacy: an exploratory study. Br. J. Educ. Technol. **51**(1), 281–296 (2020). https://doi.org/10.1111/bjet.12755
41. Mousavi, S.M.A., Powell, W., Louwerse, M.M., Hendrickson, A.T.: Behavior and self-efficacy modulate learning in virtual reality simulations for training: a structural equation modeling approach. Front Virtual Real. **4**, 1250823 (2023). https://doi.org/10.3389/frvir.2023.1250823
42. Biasutti, M.: Flow and Optimal Experience (2017). https://doi.org/10.1016/B978-0-12-809324-5.06191-5
43. Bodzin, A., Junior, R.A., Hammond, T., Anastasio, D.: Investigating engagement and flow with a placed-based immersive virtual reality game. J. Sci. Educ. Technol. (2020). https://doi.org/10.1007/s10956-020-09870-4
44. Demerouti, E., Bakker, A.B., Sonnentag, S., Fullagar, C.J.: Work-related flow and energy at work and at home: a study on the role of daily recovery. J. Organ. Behav. **33**(2), 276–295 (2012). https://doi.org/10.1002/job.760
45. Engeser, S., Rheinberg, F.: Flow, performance and moderators of challenge-skill balance. Motiv. Emot. **32**(3), 158–172 (2008). https://doi.org/10.1007/s11031-008-9102-4
46. Schaik, P., Vallance, M.: Measuring flow experience in an immersive virtual environment for collaborative learning. J. Comput. Assist. Learn. 28 (2012). https://doi.org/10.1111/j.1365-2729.2011.00455.x
47. Alrehaili, E.A., Osman, H.A.: A virtual reality role-playing serious game for experiential learning. Interact. Learn. Environ. **30**(5), 922 (2019). https://doi.org/10.1080/10494820.2019.1703008
48. Paras, B., Bizzocchi, J.: Game, Motivation, and Effective Learning: An Integrated Model for Educational Game Design (2005)
49. Fang, Y.M.: Exploring usability, emotional responses, flow experience, and technology acceptance in VR: a comparative analysis of freeform creativity and goal-directed training. Appl. Sci. **14**(15), 6737 (2024). https://doi.org/10.3390/app14156737
50. Garris, R., Ahlers, R., Driskell, J.E.: Games, Motivation, and Learning: A Research and Practice Model (2002). https://doi.org/10.1177/1046878102238607
51. Caroux, L.: Presence in video games: a systematic review and meta-analysis of the effects of game design choices. Appl. Ergon. **107**, 103936 (2023). https://doi.org/10.1016/j.apergo.2022.103936

52. Hamari, J., Shernoff, D.J., Rowe, E., Coller, B., Asbell-Clarke, J., Edwards, T.: Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. Comput. Hum. Behav. **54**, 170–179 (2016). https://doi.org/10.1016/j.chb.2015.07.045
53. Yao, S., Kim, G.: The effects of immersion in a virtual reality game: presence and physical activity. In: Fang, X. (ed.) HCI in Games, pp. 234–242. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-22602-2_18
54. Thompson, M., Uz-Bilgin, C., Tutwiler, M.S., et al.: Immersion positively affects learning in virtual reality games compared to equally interactive 2d games. Inf. Learn. Sci. **122**(7/8), 442–463 (2021). https://doi.org/10.1108/ILS-12-2020-0252
55. Martinez, K., Menéndez-Menéndez, M.I., Bustillo, A.: Considering user experience parameters in the evaluation of VR serious games. In: De Paolis, L.T., Bourdot, P. (eds.) Augmented Reality, Virtual Reality, and Computer Graphics, pp. 186–193. Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-58465-8_14
56. Cook, N.F., McAloon, T., O'Neill, P., Beggs, R.: Impact of a web based interactive simulation game (PULSE) on nursing students' experience and performance in life support training – a pilot study. Nurse Educ. Today **32**(6), 714–720 (2012). https://doi.org/10.1016/j.nedt.2011.09.013
57. Dzeng, R.J., Lin, K.Y., Wang, P.R.: Building a construction procurement negotiation training game model: learning experiences and outcomes. Br. J. Educ. Technol. **45**(6), 1115–1135 (2014). https://doi.org/10.1111/bjet.12189
58. Streicher, A., Smeddinck, J.D.: Personalized and adaptive serious games. In: Dörner, R., Göbel, S., Kickmeier-Rust, M., Masuch, M., Zweig, K. (eds.) Entertainment Computing and Serious Games: International GI-Dagstuhl Seminar 15283, Dagstuhl Castle, Germany, July 5–10, 2015, Revised Selected Papers, pp. 332–377. Springer International Publishing (2016). https://doi.org/10.1007/978-3-319-46152-6_14
59. Seandillon Nifataro, H., Joseph Bryan, S., Gunawan, A.A.S., Syahputra, M.E.: Analysis on Adaptive Difficulty Adjustment for Player Satisfaction in Game: A Systematic Literature Review (2024). https://ieeexplore.ieee.org/abstract/document/10762577. Accessed 31 December 2024
60. Ceja, L., Navarro, J.: Dynamic patterns of flow in the workplace: characterizing within-individual variability using a complexity science approach. J. Organ. Behav. **32**(4), 627–651 (2011). https://doi.org/10.1002/job.747
61. Fullagar, C.J., Kelloway, E.K.: Flow at work: an experience sampling approach. J. Occup. Organ. Psychol. **82**(3), 595–615 (2009). https://doi.org/10.1348/096317908X357903
62. Karen, B., Michael, T.K., Christof, W., Gabriele, O.: Using mental contrasting to promote flow experiences at work: a just-in-time adaptive intervention. Comput. Hum. Behav. Rep. **16**, 100488 (2024). https://doi.org/10.1016/j.chbr.2024.100488
63. Ye, X., Backlund, P., Ding, J., Ning, H.: Fidelity in Simulation-Based Serious Games (2019). https://doi.org/10.1109/TLT.2019.2913408. Accessed 23 September 2024
64. Rooney, P.: A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process. IJGBL **2**(4), 41–60 (2012). https://doi.org/10.4018/ijgbl.2012100103
65. Anne, O.P., Darius-Aurel, F., Daniel, D.B., Sascha, S.: Visual fidelity in the metaverse matters for memory performance. Technol. Forecast. Soc. Change **205**, 123511 (2024). https://doi.org/10.1016/j.techfore.2024.123511
66. Perttula, A., Kiili, K., Lindstedt, A., Tuomi, P.: Flow experience in game based learning – a systematic literature review. ResearchGate (2024). https://doi.org/10.17083/ijsg.v4i1.151
67. López, F.R., Arias-Oliva, M., Pelegrín-Borondo, J., Marín-Vinuesa, L.M.: Serious games in management education: an acceptance analysis. Int. J. Manage. Educ. **19**(3), 100517 (2021). https://doi.org/10.1016/j.ijme.2021.100517

68. Chambilla, R., Tomiuk, D., Marcotte, S., Plaisent, M., Bernard, P.: Factors Affecting Satisfaction with Serious Games – Direct, Mediated and Higher-Order Constructs (2020). https://doi.org/10.1109/IEMCON51383.2020.9284862
69. Lopes, R., Bidarra, R.: Adaptivity Challenges in Games and Simulations: A Survey (2011). https://ieeexplore.ieee.pubapi.xyz/document/5765665. Accessed 23 September 2024
70. Tenenbaum, G., Fogarty, G.J., Jackson, S.A.: The flow experience: a Rasch analysis of Jackson's Flow State Scale. J. Outcome Meas. (1999). https://www.semanticscholar.org/paper/The-flow-experience%3A-a-Rasch-analysis-of-Jackson's-Tenenbaum-Fogarty/10253a6fb5b70d34ba504170ea96bdbe88a9cb17. Accessed 23 September 2024
71. Linares, M., Gallego, M.D., Bueno, S.: Proposing a TAM-SDT-based model to examine the user acceptance of massively multiplayer online games. Int. J. Environ. Res. Public Health **18**(7), 3687 (2021). https://doi.org/10.3390/ijerph18073687
72. Ramli, N.A., Latan, H., Nartea, G.: Why should PLS-SEM be used rather than regression? Evidence from the capital structure perspective. In: ResearchGate (2024). https://doi.org/10.1007/978-3-319-71691-6_6
73. Kurtaliqi, F., Miltgen, C.L., Viglia, G., Pantin-Sohier, G.: Using advanced mixed methods approaches: combining PLS-SEM and qualitative studies. J. Bus. Res. **172**, 114464 (2024). https://doi.org/10.1016/j.jbusres.2023.114464
74. Hair, J.F., Anderson, R.E., Tatham, R.L.: Multivariate Data Analysis with Readings, 2nd ed. Macmillan Publishing Co. Inc. (1986)
75. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. J. Mark. Res. **18**(1), 39–50 (1981). https://doi.org/10.1177/002224378101800104
76. Nunnally, J.: Psychometric Theory, 3rd ed. (1994). https://cir.nii.ac.jp/crid/1370002219408110722. Accessed 05 September 2024
77. Henseler, J., Ringle, C.M., Sinkovics, R.R.: The use of partial least squares path modeling in international marketing. In: New Challenges to International Marketing, vol. 20, pp. 277–319. Emerald Group Publishing Limited (2009). https://doi.org/10.1108/S1474-7979(2009)0000020014
78. Hair Jr., J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P., Ray, S.: Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook. Springer Nature (2021). https://doi.org/10.1007/978-3-030-80519-7
79. Ajamieh, A., Benitez, J., Braojos, J., Gelhard, C.: IT infrastructure and competitive aggressiveness in explaining and predicting performance. J. Bus. Res. **69**(10), 4667–4674 (2016). https://doi.org/10.1016/j.jbusres.2016.03.056
80. Shrout, P., Bolger, N.: Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol. Methods **7**, 422–445 (2002). https://doi.org/10.1037/1082-989X.7.4.422
81. Lemmens, J.S., von Münchhausen, C.F.: Let the beat flow: how game difficulty in virtual reality affects flow. Acta Physiol (Oxf.) **232**, 103812 (2023). https://doi.org/10.1016/j.actpsy.2022.103812
82. Kim, D., Ko, Y.J.: The impact of virtual reality (VR) technology on sport spectators' flow experience and satisfaction. Comput. Hum. Behav. **93**, 346–356 (2019). https://doi.org/10.1016/j.chb.2018.12.040
83. Coelho, C., Tichon, J., Hine, T.J., Wallis, G.: Media presence and inner presence: the sense of presence in virtual reality technologies. ResearchGate (2006). https://www.researchgate.net/publication/312980588_Media_presence_and_inner_presence_The_sense_of_presence_in_virtual_reality_technologies. Accessed 13 January 2025

84. Lowell, V.L., Tagare, D.: Authentic learning and fidelity in virtual reality learning experiences for self-efficacy and transfer. Comput. Educ. **2**, 100017 (2023). https://doi.org/10.1016/j.cexr.2023.100017
85. Luo, Y., Ahn, S., Abbas, A., Seo, J., Cha, S.H., Kim, J.I.: Investigating the impact of scenario and interaction fidelity on training experience when designing immersive virtual reality-based construction safety training. Dev. Built Environ. **16**, 100223 (2023). https://doi.org/10.1016/j.dibe.2023.100223

Enhancing VR Immersion Through Avatar Scaling and Sensor Fusion with Mediapipe

Shidao Zhao(✉) and Tomochika Ozaki

Graduate School of Integrated Science and Technology, Nagasaki University, Nagasaki City 852-8521, Japan
bb54424210@ms.nagasaki-u.ac.jp

Abstract. Virtual Reality (VR) has gained increasing attention in recent years due to its strong immersion and realistic simulation. VR devices often include a Head Mounted Display (HMD) and VR controllers, which enhance VR immersion and interaction. However, the combination of HMD and controllers can only track the position of the user's head and hands. For the rest of the body, there is no widely used method, as full-body motion capture devices are cumbersome and expensive. In this paper, we propose combining Mediapipe with VR devices to enhance VR immersion. We use VR devices to adjust the avatar scale and Mediapipe's output. Considering that 2D camera full-body motion capture results are jittery, we apply filtering to reduce the jitter. Also, Mediapipe's outputs are pixel positions, so it is not suitable for controlling the avatar directly. To solve this problem, we calculate the direction vectors of the body parts, which are calculated from Mediapipe's output, to control the avatar. We evaluated our method through both visual and quantitative comparisons with other methods. It shows that our method performs better than the pure Mediapipe method and the Kinect method. As a result, we believe that our method can be an effective solution for full-body motion capture for general users and enhances VR immersion.

Keywords: virtual reality · motion capture · sensor fusion

1 Introduction

Since the release of Vision Pro and Meta Quest3, VR devices have increasingly gained attention. Compared to traditional displays, VR devices provide a superior sense of immersion and engagement, enabling more realistic and interactive experiences. However, as VR devices primarily consist of headsets and joysticks, they can only capture the user's head and hand movements. Therefore, the most widely used method today is IK-based control. However, because they are not able to obtain the position information of the knees, legs and elbows, it is necessary to rely on other methods to obtain, such as the methods based on the motion capture suit [1, 2] and the methods based on the trackers [3, 4].

However, the additional dedicated motion capture suit and trackers will make the process of using VR every time cumbersome and expensive, thus reducing the user's

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 63–75, 2025.
https://doi.org/10.1007/978-3-031-93715-6_5

interest in using VR equipment. Employing depth cameras like Azure Kinect to perform 3D human body recognition and integrate it with VR devices offers full-body motion capture without complicated preparation work. However, professional depth cameras like Azure Kinect are expensive and not widely adopted, making this method impractical for broad usage among VR users.

The combination of 3D human body recognition with VR using a monocular RGB camera is the most widely accepted by users. Therefore, we propose a method that combines a 2D camera-based 3D motion capture neural network with VR devices for motion capture. Considering the real-time requirements of VR for accurate gesture tracking, we chose Mediapipe as our 2D camera-based motion capture framework. However, when using image recognition-based motion capture, noticeable jitter in the character's movements may be experienced from the first-person perspective of the avatar. Therefore, we also adopt the moving least squares (MLS) method to eliminate jitters. The size discrepancy between the avatar's and user's sizes can misalign the avatar's joints, even when the gestures are identical, reducing immersion in the VR game. Therefore, we propose dynamically adjusting the avatar's scale to match the user's body proportions and leveraging sensor fusion to refine Mediapipe's results, significantly enhancing immersion and improving the VR experience. Through our method, the user's immersion can be greatly improved when playing VR games.

2 Related Works

2.1 Trackers

Trackers refer to devices attached to the human body to track parameters such as position, acceleration, velocity, and direction [3]. These devices often include Inertial Measurement Units (IMUs) to measure body direction and angular velocity, enabling the calculation of body gestures. They are often combined with computer vision-based motion capture to calibrate the IMUs' position, improving motion capture accuracy [5].

However, wearable devices require users to wear and set them up before use, increasing the time cost required before operation. Moreover, IMUs are not particularly useful for common users due to their lack of additional functionalities. This results in low user intent to purchase IMU devices, leading to limited market penetration. More importantly, IMUs have limited availability and high costs, which may discourage common users from acquiring them.

2.2 Computer Vision

Computer vision is the most widely used method for pose estimation due to its ease of setup and wide availability.

Motion capture methods based on depth cameras like Kinect [6] are effective as they capture users' depth information and build 3D models to improve accuracy [7]. However, depth cameras are not widely purchased by common users, discouraging their use for motion capture.

Among consumer motion capture methods, 2D camera-based approaches are the most widely used [8, 9]. However, these methods often require high computational resources [10], or alternatively, they achieve lower computational demands at the cost of reduced accuracy [11].

2.3 Inverse Kinematics

Inverse kinematics (IK) is a mathematical process used to calculate the joint parameters required to position the end of a kinematic chain [12]. It is widely used in VR games, such as Assassin's Creed Nexus VR and VRChat.

The precise tracking of VR headsets and controllers ensures the accuracy of IK end effectors' position coordinates. Consequently, in VR, methods combining controllers and HMDs with IK achieve effective hand control.

However, as IK control relies on end effectors, it cannot determine the real positions of intermediate points, such as elbows and legs, leading to unnatural movement and an inability to accurately represent arm rotations.

3 Enhancing Immersion

In our method, we combine VR devices with Mediapipe. The VR devices collect the user's head and hands positions and orientation data. This information will be used for Scale adjustment and Mediapipe adjustment process to calculate scale factors and correction coefficients, respectively. The adjustments are required to be made at the start of the VR software.

The 2D camera will also take the user's full-body image for Mediapipe to recognize the user body's key joints in 3D positions. The 3D positions are then adjusted using scale factors and correction coefficients. Afterward, these adjusted 3D positions are used to calculate the direction vectors of the avatar's body parts. Finally, the avatar's body parts are controlled using these direction vectors and positions and orientation data collected by VR devices.

3.1 Scale Adjustment Process

In our method, the avatar's size needs to be scaled before the VR software starts. The scale adjustment process is illustrated in Fig. 1.

Fig. 1. Scale adjustment process

To adjust the avatar's size, users should perform the following steps each time the VR software starts:

1. Stand in a T-pose with VR devices equipped.

2. Press the measure button to determine the height H_p, which represents the distance from the HMD to the ground.

However, the HMD's height corresponds to the user's eye level, whereas the avatar's height H_c is defined as the distance from the top of the head to the feet. Thus, H_p should be increased by the adjustment distance De, a constant representing the distance from the user's eyes to the top of their head. The overall body scale S_b can then be calculated as follows:

$$S_b = \frac{H_p + D_e}{H_c} \tag{1}$$

The arm scale will be adjusted to S_a after the height adjustment. The algorithm for adjusting the arm scale is as follows:

$$S_a = \frac{|C_L - C_R|}{|W_L - W_R|} \tag{2}$$

Here, C_L and C_R represent the positions of the left and right controllers, while W_L and W_R denote the positions of the avatar's left and right wrists. Unity then updates the avatar's arm scale to S_a.

3.2 Mediapipe Adjustment Process

Although high-speed 3D motion capture methods, such as Mediapipe, can be used for full-body motion capture, they fail to align the avatar with the real body from a first-person perspective. However, by combining the precise tracking of VR controllers with the flexibility of Mediapipe, we can achieve more accurate body alignment and smoother motion capture.

In our method, correction coefficients are calculated to adjust the Mediapipe output. The calculation process is illustrated in Fig. 2.

First, we instruct the user to perform the gesture shown in Fig. 3, ensuring the wrists are in different spatial positions with different values along the x, y, and z axes of the real-world coordinate system. Using position data from both the controllers and Mediapipe, we calculate the correction coefficients P_X, P_Y, and P_Z for the x, y, and z directions, respectively, as follows:

$$P_X = \frac{CL_X - CR_X}{IL_X - IR_X} \tag{3}$$

$$P_Y = \frac{CL_Y - CR_Y}{IL_Y - IR_Y} \tag{4}$$

$$P_Z = \frac{CL_Z - CR_Z}{IL_Z - IR_Z} \tag{5}$$

For instance, for the z-axis, CL_Z and CR_Z are the z-coordinates of the left and right controllers, respectively, and IL_Z and IR_Z are the z-coordinates of the left and right wrists from Mediapipe. The same calculation method applies to the x and y axes, as shown in formulas (3) and (4).

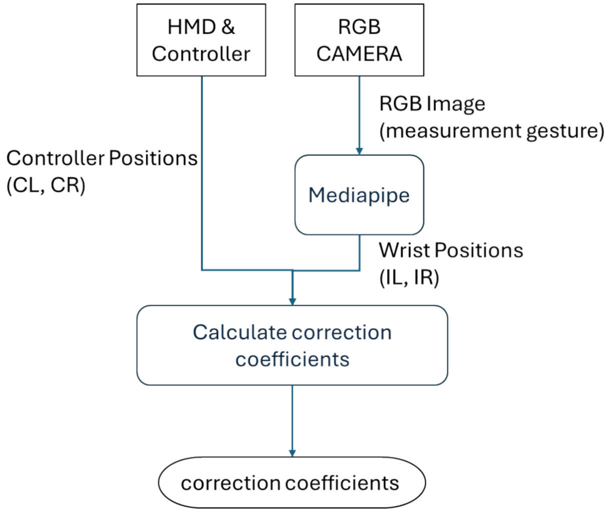


Fig. 2. Correction Coefficients Calculation.

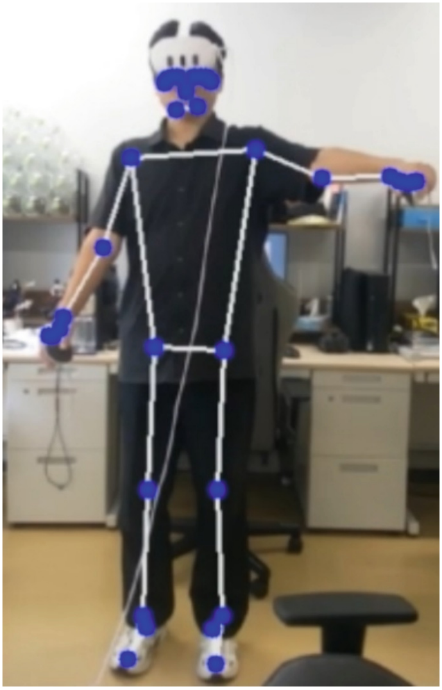


Fig. 3. Measurement gesture.

3.3 Avatar Control Process

The avatar control process is illustrated in Fig. 4. In the avatar control process, we first apply filtering to the 3D positions from Mediapipe to reduce jitter. Next, we calculate the direction vectors from the filtered 3D positions. Finally, we use the HMD, controllers, and direction vectors to control the avatar's gestures.

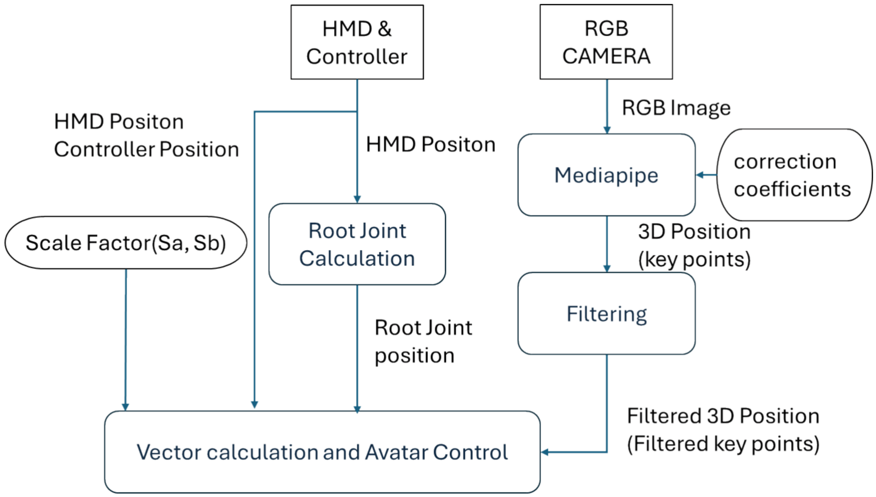


Fig. 4. Avatar control process

Filtering. 2D camera-based human gesture recognition often suffers from significant jitters in the results, which can notably impact the gaming experience. To address this, we apply filtering in our method. The filtering process is applied to the adjusted Mediapipe coordinates I'_X, I'_Y, I'_Z, which will be introduced in Sect. 3.3.

First, we used a Moving Average Filter (MAF) to reduce the jitter. Although it helps and makes the avatar's body move more smoothly compared to when MAF is not used, the jitter remains more significant than expected.

The Extended Kalman Filter (EKF) is often used in human gesture recognition filtering. Compared to the MAF and Kalman Filter, EKF can handle nonlinear motion.

However, compared to EKF, the Moving Least Squares (MLS) does not require a prediction model, meaning that MLS is more robust to unpredictable human gestures.

Root Joint. Before avatar gesture control, the avatar's root joint should be defined. Each vector requires a parent object as the start and a child object as the end. The top-level parent object is the root joint.

In Unity, avatar models typically use hips as the root of the bone structure. However, since the HMD provides the most accurate 3D position and the avatar's head needs to align with the HMD's position. In Jalapati [13], a Multi-Parent Constraint was used to ensure that the avatar's head aligns with the HMD's position. They set the head to be the root of the skeletal hierarchy and determine the rotation and translation for the rest of the body.

However, setting the avatar's head as the root of the skeletal hierarchy presents a problem: the rotation and translation of the avatar follow a forward kinematics (FK) method, which causes precision to decrease from the root to the extremities. This means that precision will significantly decrease by the time it reaches the feet. Additionally, the Multi-Parent Constraint does not perform well in VR scenarios due to the unpredictable nature of the user's head movement.

Therefore, it makes sense for Unity to set the hips as the root of the avatar, as they are the central point of the body. This minimizes errors in the extremities of the body. Therefore, we retain the hips as the root. Furthermore, setting the head as the root of skeletal hierarchy is not the only method to ensure the avatar's head aligns with the HMD's position. We chose to keep the avatar's head aligned with the VR HMD's position in Unity. The body position calculation is as follows:

$$P_b' = P_{HMD} - P_h + P_b \tag{6}$$

where, P_{HMD} is the HMD's position in Unity, P_h is the avatar's head position, P_b is the body position from the previous frame and P_b' is the body's position in the current frame. By doing so, we can keep the avatar's head in the same position with HMD with the root joint to be the hips.

Vector Calculation and Avatar Control. The 33 key points that Mediapipe can detect are illustrated in Fig. 5. While Mediapipe outputs 3D coordinates for human key points that can serve as a basis for controlling a character model, directly using these coordinates in pixel units introduces complex 3D coordinate transformation challenges. For in-stance, mapping a single pixel in Mediapipe to real-world distances presents a significant challenge. Furthermore, perspective projection causes changes along the z-axis to affect the relationship between pixels and real-world measurements on the x and y axes. However, the directional vectors between key points remain unaffected by these trans-formations. Consequently, we utilize the directional vectors of human key points to control the avatar.

Mediapipe's output is used as the base for these body parts, and we apply the following correction functions to adjust the Mediapipe output key points' coordinates:

$$I_X' = I_X P_X \tag{7}$$

$$I_Y' = I_Y P_Y \tag{8}$$

$$I_Z' = I_Z P_Z \tag{9}$$

where, I_X, I_Y, I_Z are unadjusted Mediapipe output key points' coordinates, while I'_X, I'_Y, I'_Z are the adjusted. The P_X, P_Y, and P_Z are the correction coefficients calculated in Sect. 3.2. The adjusted coordinates are then used to calculate direction vectors, for example, right arm vector V_A is:

$$V_A = \left(ERI_X' - ARI_X', ERI_Y' - ARI_Y', ERI_Z' - ARI_Z'\right) \tag{10}$$

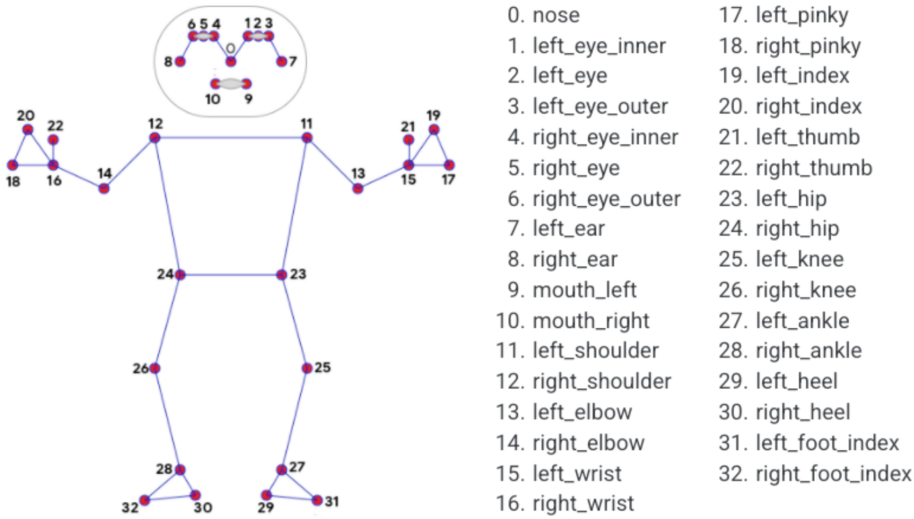


Fig. 5. Mediapipe key points

where, ARI'_X, ARI'_Y, ARI'_Z and ERI'_X, ERI'_Y, ERI'_Z are the adjusted coordinates of the right arm and elbow, respectively. We use V_A to control the right arm direction and determine the position of the right elbow. The positions of AR, ER and V_A are illustrated in Fig. 6. The same method can be used to calculate the direction vectors for the left arm, right arm, left elbow, right hip, right knee, left hip, and left knee.

Calculating the direction vectors of the spine and hips is more complex compared to the limbs. Among the key points provided by Mediapipe, no pair directly represents the direction of the human chest. Therefore, we first calculate the vector from the right shoulder to the right hip VR_{sh}, and from the left shoulder to the left hip VL_{sh}, and take the average vector to represent the vector from the chest to the center of the hips V_{sh}. Then, we calculate the direction vector from the left shoulder to the right shoulder V_{ss} and perform a cross-product operation to obtain the direction vector of the chest V_b:

$$V_b = V_{ss} \times V_{sh} \tag{11}$$

However, the direction vector of the chest does not determine whether the left shoulder is higher than the right shoulder or vice versa. Therefore, when controlling the chest direction, we again use the direction vector from the left shoulder to the right shoulder V_{ss} to control the position of the left and right shoulders.

In Unity, for each object, there is a local coordinate system. For spine, as illustrated in Fig. 7, the z-axis represents the forward vector, the y-axis represents the up vector, and the x-axis represents the right vector. Therefore, the spine control method for the avatar is defined as follows: The z-axis direction vector of the spine aligns with the vector V_b, and the x-axis direction vector of the spine aligns with the direction of vector V_{ss}.

Similarly, the direction vector for the waist is:

$$V_h = V_{hh} \times V_{sh} \tag{12}$$

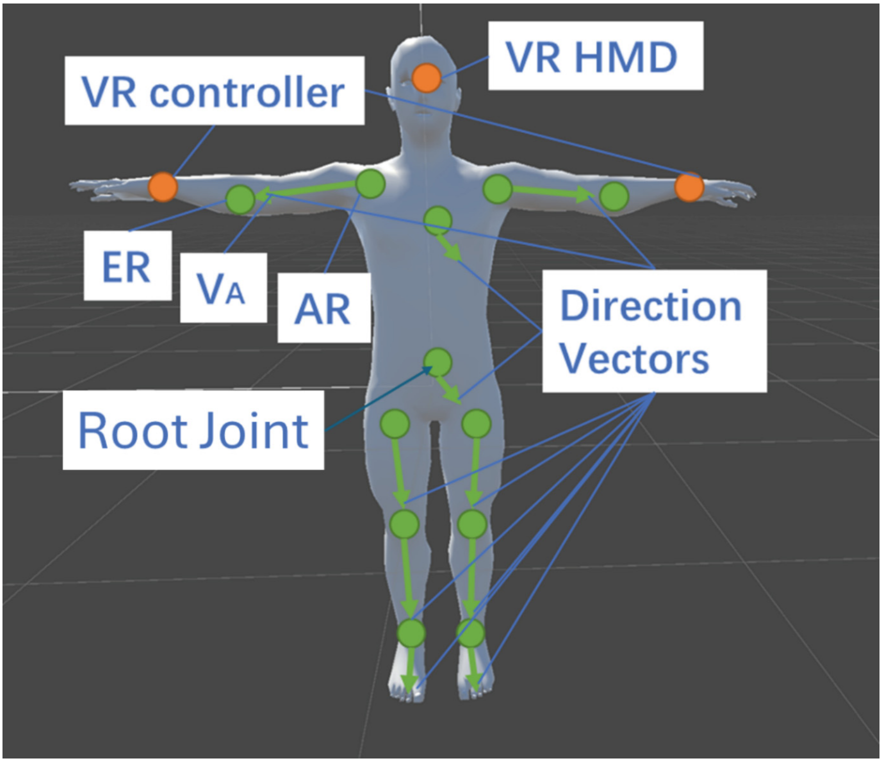


Fig. 6. Sensor fusion with Mediapipe

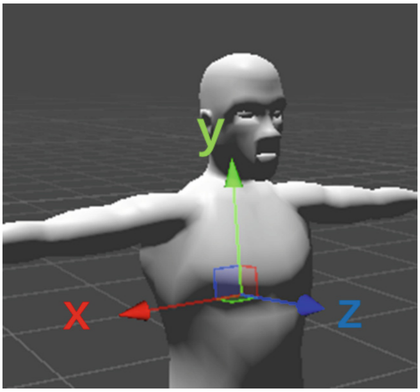


Fig. 7. Spine local coordinate system

where V_{hh} is the vector from the left hip to the right hip.

The z-axis direction vector of the waist aligns with the vector V_h, and the x-axis direction vector of the spine aligns with the direction of vector V_{hh}.

For body parts not tracked by the controllers and HMD, including the spine, upper arms, elbows, upper legs, knees, and ankles, we use the direction vectors to control them. We align the avatar body parts' direction vectors, illustrated in 0, with the calculated direction vectors. Starting from the root joint, direction vectors are used to control the body parts connected to it. For instance, direction vector V_h controls the direction of the hip, which then affects the positions of the spine and upper legs, as they are child objects of the hips. These vectors adjust the rotation of each body part, progressively aligning the avatar's pose with the user's gesture. This process continues with each subsequent body part, such as the arms and knees, being controlled by their respective direction vectors to further refine the pose. As a result, the avatar's gesture will match the user's gesture. To ensure the avatar's head matches the HMD's position, we adjust the root joint's position based on the HMD's position and avatar head's position, enhancing VR immersion.

4 Evaluation

4.1 Visual Comparison

To evaluate our method, we compare it with the Kinect method and the pure Mediapipe method in Unity simultaneously.

Using the passthrough feature of the Quest 3, we directly compared the results with the real scene, which is the ground truth. As shown in Fig. 8, the white avatar is controlled by our method, the orange shadow by Mediapipe, and the green shadow by Azure Kinect. Among these methods, ours is the closest to the real arms, demonstrating that our method enhances VR immersion and improves the accuracy of Mediapipe.

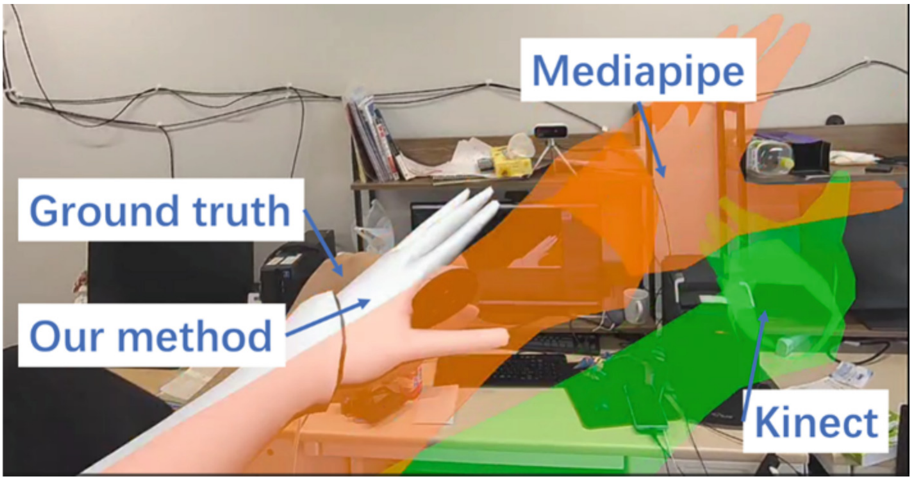


Fig. 8. Visual accuracy comparison

4.2 Quantitative Comparison

Although passthrough can provide a visual comparison and show the results of the methods, we need a quantitative comparison to analyze the methods.

We run our method, the Mediapipe method, and the Kinect method in Unity simultaneously. We also recorded the camera scene simultaneously and made camera measurements as the ground truth. The camera measurements use two cameras to capture the front and left-side views of the human body's movements. The cameras provide images for each moment, with no occlusion in the camera's view. We measured each joint's local angle and position in a coordinate system with the head as the origin. The results were obtained by averaging the measurements taken over three separate measurements. We chose several typical gestures to evaluate our method, including walking, sitting, bending, jumping, waving, and raising hands. For each gesture, we took 20-s videos of typical gestures for evaluation.

We use Unity to output the results of each frame from the methods mentioned earlier and compare the results with the ground truth obtained from the camera measurement. The body parts' positions and their angles will be recorded by Unity every 0.1 s. Then, we calculate the Mean Per Joint Positional Error (MPJPE) and Mean Per Joint Angle Error (MPJAE) for the methods. In this paper, for MPJPE, we calculate the body parts, which include the spine, arms, elbows, wrists, hips, knees, and ankles. For MPJAE, the body parts include the spine, arms, elbows, hips, and knees.

To calculate the MPJPE and MPJAE, we used the camera measurement results and the Unity results. In VR, the head's position in Unity is the same as in reality. Therefore, we set it as the origin and define a coordinate system. We calculate the distance from the joint position to the corresponding joint in the ground truth to compute the MPJPE. For MPJAE, we calculated each joint's local angle and subtracted it from the corresponding joint's local angle in the ground truth.

5 Results

We compared our method with Mediapipe and Kinect by computing the MPJPE of gestures including walking, sitting, bending, jumping, waving, and raising hands. The result is shown in Table 1:

Table 1. MPJPE (mm) of different methods

Methods	Walk	Sit	Bend	Jump	Wave	Raise	Avg
Ours(unscaled)	53.04	97.47	55.65	104.85	42.15	43.28	66.07
Ours(scaled)	**34.76**	**65.26**	**40.07**	**50.58**	**35.27**	**40.81**	**44.46**
Kinect	62.06	138.54	89.01	105.99	68.71	63.88	88.03
Mediapipe	181.09	181.76	189.45	235.79	112.31	136.68	172.85

Table 2. MPJAE (°) of different methods

Methods	Walk	Sit	Bend	Jump	Wave	Raise	Avg
Ours(unscaled)	8.71	17.45	10.42	18.45	8.59	8.65	12.04
Ours(scaled)	**5.91**	**11.35**	**6.97**	**9.32**	**6.63**	**7.69**	**7.98**
Kinect	12.52	27.01	14.62	16.92	15.43	16.28	17.13
Mediapipe	30.18	36.76	32.76	49.47	29.98	34.11	35.54

We also calculated the MPJAE of the gestures mentioned above, and the result is shown in 0 (Table 2):

Through sensor fusion with Mediapipe, our method demonstrates a notable improvement. Additionally, the use of controllers and IK effectively corrects arm angle errors, thereby enhancing accuracy. The unscaled avatar controlled by our method appears unable to follow the arm's IK accurately, as the length of the avatar's arm differs from that of the user's arm. This causes the avatar's arm joint angles to differ from the ground truth. Additionally, although the leg joint angles are correct, the difference in leg length causes the distance between the avatar's legs and the ground truth to increase.

In our measurements, we found that although Mediapipe can recognize 3D human gestures, it is weak in feeling z-axis, which represents the depth for the camera. This causes the avatar controlled by Mediapipe to lean forward overall. This suggests that we can use a ratio to correct the z-axis value. It is a simple method but can significantly improve the accuracy of Mediapipe's results with minimal computational burden. This allows both real-time operation and precision to be considered in our method.

Compared to Kinect, our method provides superior hand tracking, though it is less effective at leg tracking. However, Kinect is more cumbersome and expensive, while our method only requires a standard camera. In our method, the hand positions are reflected on the avatar with a 10 ms computational latency, while the latency for other body parts is 43 ms. Although the latency of the elbows and arms is 43 ms, the hands' latency allows the elbows and arms to have a 10 ms computational latency, as the IK for the arms can determine the positions of the elbows and arms once the hand positions are decided. However, as mentioned in Sect. 1, IK alone is not able to provide the real position of the elbows. Therefore, the elbow position with 10 ms latency is less precise than the elbow position with 43 ms latency. Compared to our method, Kinect experiences a delay of 55 ms. The latency in Kinect causes its output to lag behind the ground truth, which increases the error in Kinect's results. In contrast, the lower latency of our method improves overall precision.

6 Conclusion

In this paper, we propose a method to enhance full-body motion capture for VR using a standard RGB camera. To ensure real-time gesture tracking, we initially employ Mediapipe for full-body motion capture. Subsequently, we apply avatar scale adjustment, sensor fusion, IMU-based adjustments, and MLS filtering to improve both accuracy and smoothness. Compared to Kinect-based full-body motion capture, our method offers lower foundational and computational costs while achieving superior hand tracking performance.

References

1. Fujimori, Y., Ohmura, Y., Harada, T., et al.: Wearable motion capture suit with full-body tactile sensors. In: 2009 IEEE International Conference on Robotics and Automation, pp. 3186–3193. IEEE (2009)
2. Fitzgerald, D., Foody, J., Kelly, D., et al.: Development of a wearable motion capture suit and virtual reality biofeedback system for the instruction and analysis of sports rehabilitation exercises. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4870–4874. IEEE (2007)
3. Kačerová, I., Kubr, J., Hořejší, P., et al.: Ergonomic design of a workplace using virtual reality and a motion capture suit. Appl. Sci. **12**(4), 2150 (2022)
4. Shin, S., Li, Z., Halilaj, E.: Markerless motion tracking with noisy video and IMU data. IEEE Trans. Biomed. Eng. **70**(11), 3082–3092 (2023)
5. Pan, S., Ma, Q., Yi, X., et al.: Fusing monocular images and sparse IMU signals for real-time human motion capture. In: SIGGRAPH Asia 2023 Conference Papers, pp. 1–11 (2023)
6. Tölgyessy, M., Dekan, M., Chovanec, Ľ, et al.: Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. Sensors **21**(2), 413 (2021)
7. Chatzitofis, A., Zarpalas, D., Daras, P., et al.: DeMoCap: low-cost marker-based motion capture. Int. J. Comput. Vis. **129**(12), 3338–3366 (2021)
8. Scataglini, S., Abts, E., Van Bocxlaer, C., et al.: Accuracy, validity, and reliability of markerless camera-based 3D motion capture systems versus marker-based 3D motion capture systems in gait analysis: a systematic review and meta-analysis. Sensors **24**(11), 3686 (2024)
9. Horsak, B., Prock, K., Krondorfer, P., et al.: Inter-trial variability is higher in 3D markerless compared to marker-based motion capture: Implications for data post-processing and analysis. J. Biomech. **166**, 112049 (2024)
10. Zhu, W., Ma, X., Liu, Z., et al.: Motionbert: a unified perspective on learning human motion representations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 15085–15099 (2023)
11. Kim, J.W., Choi, J.Y., Ha, E.J., et al.: Human pose estimation using mediapipe pose and optimization method based on a humanoid model. Appl. Sci. **13**(4), 2700 (2023)
12. Waldron, K.J., Schmiedeler, J.: Kinematics. Springer Handbook of Robotics, pp. 11–36 (2016)
13. Jalapati, P., Naraparaju, S., Yao, P., et al.: Integrating sensor fusion with pose estimation for simulating human interactions in virtual reality. In: International Conference on Human-Computer Interaction, pp. 74–87. Springer Nature Switzerland, Cham (2022)

Virtual Environments for Learning, Training and Professional Development

Exploring How Augmented Reality Display Features Affect Training System Performance

Gerd Bruder[1(✉)], Ryan Schubert[1], Michael P. Browne[2], Austin Erickson[3], Zubin Choudhary[1], Matt Gottsacker[1], Hiroshi Furuya[1], and Gregory Welch[1]

[1] University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 08544, USA
{bruder,ryan.schubert,zubin.choudhary,matthew.gottsacker,hiroshi.furuya,welch}@ucf.edu

[2] Vision Products, LLC, Campbell, CA, USA
m.browne@visionproducts.llc

[3] KBR Inc., Wright-Patterson AFB, Houston, OH, USA
austin.erickson@us.kbr.com

Abstract. Augmented Reality (AR) head-worn display (HWD) technologies for warfighters have seen various advances over the last decade that make them attractive for simulations, training, and operations. In particular, optical see-through (OST) AR displays are becoming more used on the battlefield as they do not reduce warfighters' visual acuity of the real world. Unfortunately, these displays are still limited in terms of the field of view (FOV) and luminance of the display, the latter of which competes with the luminance in the warfighter's environment. The objective of this work is to evaluate how these two AR HWD factors impact participants' spatial task performance and perception. Specifically, this paper presents an experiment, performed inside a novel hybrid experimental space in which both the FOV and the luminance contrast of the HWD were varied compared to the simulated environment. Participants performed a spatial task involving simulated humans arranged in the 360-degree space around them, augmented with red or blue team member tags on the AR display. The results show that a FOV of 45 degrees or wider as well as a luminance contrast of 0.1 or higher were required for participants to reach a task performance that matched or exceeded that which could be reached without the use of AR. Implications for warfighter AR HWD systems are discussed, especially as they pertain to a warfighter's willingness to trust an AR HWD system and how that trust impacts performance.

Keywords: Augmented Reality · Training · Display Factors

1 Introduction

Augmented reality (AR) [42] systems come in a diverse array of configurations, each offering distinct display capabilities [3,4]. While it is intuitive that these

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 79–95, 2025.
https://doi.org/10.1007/978-3-031-93715-6_6

display characteristics influence users' spatial abilities and overall performance, empirical evidence in the scope of training and operational scenarios remains scarce. Particularly important factors are the field of view (FOV) and the contrast of AR imagery shown on optical see-through (OST) head-worn displays (HWDs), which have been argued to affect situational awareness, navigation, and related cognitive processes [18,25,40]. However, the research community's understanding of both objective and subjective effects stemming from these display configurations remains limited. Specifically, questions persist regarding user reliance on and trust in AR systems, as well as the implications on overall performance.

This paper examines two AR display factors and their implications for user trust and performance. The investigation centers on the following aspects [25]:

- **Field of View**: We tested the effects of horizontal FOVs in AR using the Vision Products SA-147/S HWD, which has one of the widest horizontal FOVs currently available, allowing simulated FOVs from 15° to an expansive 143°. By scrutinizing FOV variations, the study investigates how users' spatial task performance is influenced and whether trust in the AR system is affected by the FOV.
- **Visual Contrast**: Within the scope of this work, visual contrast pertains to the luminance disparities between the AR imagery displayed by an OST HWD and the physical environment. These contrast variations will likely impact human perception and cognition. Investigating the interplay between visual contrast and user experience is crucial, as it may shape subjective perceptions of reliability and trust in the AR system.

We investigated how far these AR display factors affect user performance in a search and selection task that involves searching for and selecting targets in a 360° range around the user, similar to a 360° shooting gallery. We present an experiment (N = 20), which evaluates the interactions between the two AR display factors FOV and visual contrast. The results show effects of the AR display factors on users' task performance as well as on users' sense of reliance and trust in the AR system. In particular, the results show objective and subjective benefits of wider FOVs and higher contrasts, and that neither is capable of fully compensating for limitations in the other factor. This research contributes to a deeper understanding of AR display configurations, bridging the gap between theoretical considerations and practical implications [6,25]. By elucidating the effects of FOV and visual contrast, as well as their mutual connections, this study paves the way for informed design decisions and optimized AR experiences.

2 Background

This section provides background information on trust and behaviors, and the two AR display factors.

Fig. 1. Conceptual illustration of AR tags registered to real people, identifying them as friend or foe.

2.1 Trust in and Reliance on AR Spatial Cues

AR technologies hold significant promise in their capacity to display spatial cues directly within the user's FOV. AR cues can offer information in a visually accessible manner, surpassing the limitations of cues derived solely from the physical environment. Unlike real-world cues, which often remain small, concealed, and challenging to access, AR cues can be designed to be visible, salient, and intuitively comprehensible to improve user performance in a range of spatial tasks [43]. A noteworthy example are AR wayfinding cues, which can abstract and simplify otherwise complex navigational tasks for users. Another example are AR tags (see Fig. 1) that can be seamlessly registered with real-world objects, supporting object search and identification, especially in cluttered environments or time-pressure situations. In such cases, as technically redundant information is available to users, they may choose to rely only on the AR cues, only on the real-world cues, or a combination of both to increase their confidence in their decisions, though integrating cues from both channels has the drawback of potentially slowing users down.

Various factors play a role in a user's interaction with an AR system during a given task. These factors include the system's reliability, the user's attitudes toward such systems, the workload associated with the task, the user's confidence in their ability to perform the task, and their trust in the system [33]. In this context, system reliability-specifically, the reliability of AR cues-differs from a user's reliance on that system. Reliability refers to the system's actual capability, such as the percentage of correct actions performed by an alarm system. Reliance, on the other hand, reflects a user's decision to depend on the system for task execution. Trust, a related but distinct concept, draws from interpersonal trust-the trust a person has in another individual [29]. There is currently no agreed-upon definition of "trust" in the context of AR systems assisting users in spatial tasks. However, it is generally accepted as a latent concept. Lee and See proposed that trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [26]. A user's trust in and reliance on an AR system are influenced by various factors related to the system, environmental conditions, and the specific task context. Related work found that AR users' reliance and trust often correlate with perceived changes in performance efficacy [30,35,43]. If a user's negative impression of an AR system is unfounded, it may result in underreliance or undertrust that may cause unnecessary drops in performance [7,27,33,39], while overreliance or overtrust may result in overuse and also potentially decreased performance [31,32].

2.2 AR Head-Worn Display Factors

Field of View. Human peripheral vision serves as a valuable source of information [8]. For example, Jones et al. highlighted the importance of presenting information in peripheral vision for calibrating movement within an environment [22]. Limiting AR cues to a narrow region of a user's FOV can introduce perceptual ambiguities or restrictions for various visual tasks [25]. Additionally, the narrow FOV of an AR display can lead to cognitive challenges, such as misjudging object speeds [36] or increasing cognitive load [5]. Further, studies conducted in immersive virtual environments have revealed performance and behavioral differences related to FOVs. For instance, Covelli et al. investigated the effects of FOV on pilot performance during training, observing distinct head movement patterns and visual scan patterns associated with a narrow FOV [9]. Kishishita et al. explored secondary visual search tasks and found task improvements for FOVs up to 100°, with diminishing benefits beyond 130° [24]. Trepkowski et al. reported increased search performance for text and symbols with wider FOVs, resulting in fewer overlooked search targets [41]. In some cases, wide FOVs have been shown to enhance training performance [34]. For instance, Arthur demonstrated that a narrow FOV led to performance degradation during maze navigation and visual search [1].

Visual Contrast. In OST AR HWDs, aside from static or active dimming techniques, the display relies on an additive light model. This means that the display emits light that is added to the light received by the user's eyes from the real world [15,17]. In other words, to create an AR image with the correct illuminance and color, the display aims to introduce additional light on top of the ambient light [17,18]. However, a consequence of this approach is that when the physical background is well-lit-such as in direct sunlight outdoors-the AR imagery displayed by OST AR devices can lose contrast [4,14]. Users tend to perform worse on AR tasks in brighter environmental conditions because the AR imagery becomes washed out and lacks contrast [13]. For instance, Kim et al. demonstrated this effect when reading text under indoor lighting conditions ranging from 10 to 300 lux [23]. Gattullo et al. observed a similar phenomenon when reading text with lighting between 1,000 and 4,000 lux, which corresponds to bright indoor industrial settings or dim overcast outdoor lighting [19]. Debernardis et al. also reported comparable trends related to text readability, although they did not provide specific illuminance measurements [11].

2.3 Hybrid Projection Plus AR Setups

Hybrid setups that involve both immersive projection technologies and AR HWDs are very rare, and have not been leveraged to their full potential for experimentation and training yet. In fact, no other setup was identified that uses the approach that was taken in this work to study the effects of AR HWDs on user performance and perception under controlled laboratory conditions. There are other examples of setups that use an installation like a Cave Automatic Virtual Environment (CAVE) [10] in the sense of simulating an environment for participants or trainees to perform activities within, such as the Wide Area Virtual Environment (WAVE) at the Uniformed Services University, which is used to train medical teams on how to perform their tasks under simulated battlefield conditions [28]. However, none of these setups appear to have their participants or trainees wear AR displays at the same time while they are performing these tasks. As AR displays are increasingly finding their way into warfighter training and operations, it is important to not only simulate an immersive training environment with CAVE-like installations [21,37,38] but also to be able to accurately simulate how AR displays worn by warfighters would perform within such a realistic environment [20].

2.4 Current AR Display Tradeoffs

Currently available AR displays for warfighters and in the consumer market come in a wide range of configurations, each putting an emphasis on different features at the cost of other aspects. For instance, commercial OST AR HWDs like the Microsoft HoloLens 2 or the Magic Leap 2 provide a reasonable horizontal FOV of approximately 43° or 45°, respectively, but at the cost of a low luminance display that causes AR overlays to look increasingly transparent (i.e., lower contrast) the brighter the ambient light is in the real-world environment in which warfighters operate. As found by Erickson et al., even under cloudy outdoor lighting conditions at around 1,000 lux, displays like the HoloLens 2 can barely provide 0.2 contrast, going way down to 0.01 contrast or less beyond 10,000 lux in sunny outdoor conditions [14].

3 Experiment

This section describes the experiment that investigates the AR display features FOV and contrast. This experiment was approved by the Institutional Review Board (IRB) of the University of Central Florida (UCF) and the United States Air Force Human Research Protection Program (USAF HRPP).

3.1 Participants

Following initial pilot tests, the effect size of the anticipated strong effects was estimated. Based on a power analysis using G*Power 3 [16], 20 participants were recruited from UCF's community for this experiment. Demographic information was collected as in the ACM Demographics Questionnaire [2]. The participant group consisted of 13 males and 7 females, with ages ranging from 18 to 42 (M = 24.8, SD = 6.3). None of the participants reported any visual, motor, or cognitive disabilities. All participants had normal or corrected-to-normal vision and no history of visual or vestibular disorders, such as color blindness, night blindness, dyschromatopsia, or balance issues. Participants included both students and non-student members of UCF's community who responded to open calls for participation. Monetary compensation was provided for their involvement. The experiment took participants around one hour to complete.

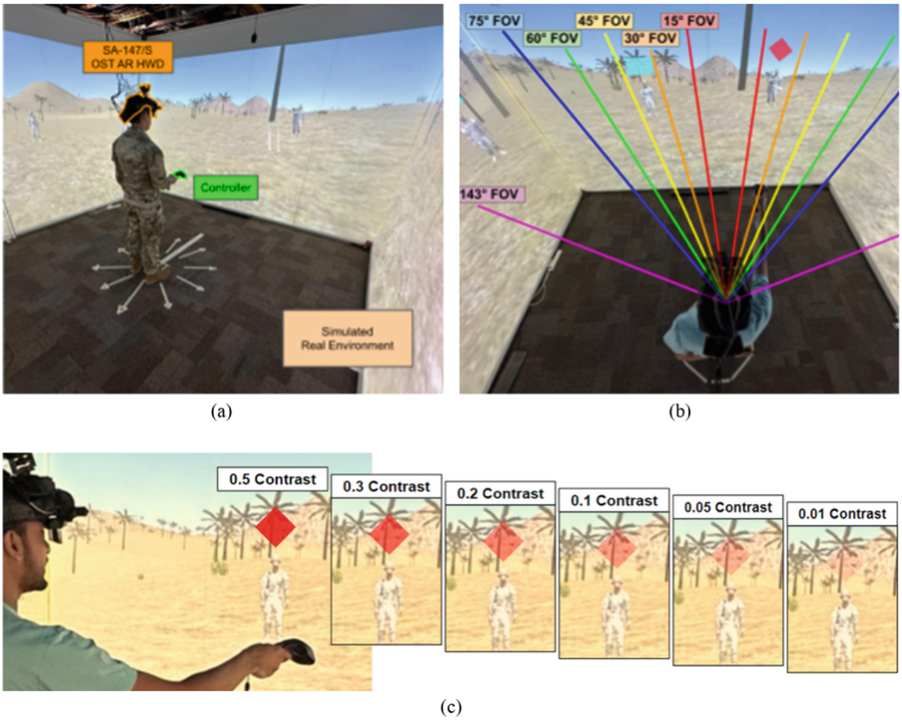


Fig. 2. Experimental setup and stimuli: (a) participant completing the experiment in a CAVE-like interaction space simulating an outdoor desert environment while wearing a Vision Products SA-147/S AR HWD, holding a controller in their dominant hand. The simulated environment shows human figures arranged in a circle around the center of the simulation space, while red/blue AR tags were presented on the AR HWD that floated over the simulated humans' heads. (b) and (c) show illustrations of the different tested horizontal FOVs and visual contrasts affecting the visibility of the AR tags that were overlaid over the simulated environment via the AR HWD. (Color figure online)

3.2 Material

The study took place in a hybrid interaction space, based on a CAVE-like installation, shown in Fig. 2. This space measures 4 m by 4 m. Each of the four walls was fully covered by imagery from NEC U321H ultra-short throw projectors, providing a resolution of 1080p per wall and a total resolution of 7680 × 1080 pixels. Within this setup, participants wore a Vision Products SA-147/S OST AR HWD. The SA-147/S utilizes four OLED microdisplays, offering a resolution of 3840 × 1200 pixels per eye. Its vertical FOV spans 33°, while the horizontal FOV reaches 143° (with a binocular overlap of 53°). To track participants' head movements, a Vive Tracker 3.0 was attached to the HWD. Additionally, two SteamVR Base Station 2.0 units were positioned diagonally in the upper corners of the interaction space. A single BOXX APEXX X3 desktop computer, equipped with two Nvidia Quadro RTX 6000 graphics cards, drove the experimenter interface, all four projected wall displays, and the two video channels for the SA-147/S. Participants interacted using a Vive Pro controller, which provided tracked point-and-click input. Experiment control and rendering were handled using Unity (version 2021.3.2) with SteamVR integration, capturing real-time head and controller poses as well as trigger presses.

The simulated virtual environment featured a desert landscape. Ten simulated human figures stood in a spaced-out circle, positioned ten meters away from the participant at regular 36° intervals (see Fig. 2). These simulated humans wore subtly colored arm bands-five in red and five in blue-to distinguish two groups (red team vs. blue team). The color order was randomized. The projected scene on the walls adjusted perspective based on the user's tracked head position, while the AR tags presented on the OST HWD were positioned to align with the simulation.

AR Stimuli. For the experimental conditions outlined below, the horizontal FOV of the AR HWD was varied in the Unity rendering environment, which meant that AR overlays were only visible within a portion of the total horizontal FOV supported by SA-147/S HWD, while the vertical FOV and overlap remained fixed in all conditions. Further, the visual contrast (c) of these AR overlays was varied, where visual contrast is defined as the illuminance ratio between the foreground illuminance (lf) and the background illuminance (lb) using Eq. 1.

$$c = l_f/(l_f + 2l_b) \in [0, 1] \tag{1}$$

This equation was proposed by Erickson et al. [12] and is derived from the Michelson contrast equation. In this experiment, the background illuminance in the CAVE-like environment was measured as 12 lux, and the foreground illuminance of the AR overlays on the HWD were varied up to 24 lux, producing a maximal contrast of 0.5.

3.3 Procedure

Upon arrival, participants were first given a physical copy of the informed consent document for the study to read, which they and the experimenter signed to confirm their consent to take part. The experimenter then guided participants to the experimental environment and verbally obtained consent to place the SA-147/S HWD on their head. The professional AR HWD was then adjusted for each participant, ensuring accurate eye positions, interpupillary distances (IPD), tracker alignment, and boresight calibration. The experimenter then explained the task to the participants, who were then allowed to practice the task as many times as they needed to make sure that they understood the instructions and were proficient and comfortable completing the task. The experiment itself consisted of baseline tasks before and after a series of tasks with different display factors. Following the completion of the search-and-selection tasks, the AR HWD was removed, and participants answered a post-questionnaire about the experienced conditions and a demographics questionnaire, were debriefed, and given monetary compensation.

3.4 Methods

This experiment used a partial factorial within-subjects design with the following display factors and baselines:

- *Fields of View:* The six horizontal FOVs of 15°, 30°, 45°, 60°, 75°, and 143° (the widest horizontal FOV supported by the SA-147/S AR HWD) were stimulated and each tested for the lowest (0.01) and highest (0.5) considered visual contrast. This study's focus was to sample the FOV range from very narrow cockpit-mounted displays, the two dominant consumer/research OST AR displays (Microsoft HoloLens 2 and Magic Leap 2), up to approximately what the Integrated Visual Augmentation System (IVAS) offered; the maximum FOV of 143° ‘ was included to show any effects at the extreme.
- *Visual Contrasts:* The six visual contrasts of 0.01, 0.05, 0.1, 0.2, 0.3, and 0.5 (the highest contrast supported by the SA-147/S AR HWD in the experimental environment) were simulated and each tested for the narrowest (15°) and widest (143°) considered FOV.
- *Baseline (pre/post):* Two baseline tests were included (before and after the experimental trials), in which participants wore the AR HWD, but it was turned off, i.e., participants had to base their decisions entirely on the environment. These baselines were included mainly as a sanity check regarding learning effects during the study.

The pre/post baseline conditions were tested in fixed order, but the order of the other tested conditions was randomized. Each condition was tested twice, once while rotating clockwise and once in counterclockwise direction to avoid participants becoming entangled in the cables that were suspended from the ceiling (see Fig. 2). In total, each participant completed 52 trials, for a total of 1040 trials for all participants.

3.5 Search-and-Selection Task

During the experiment, participants engaged in a selection task within the environment-similar to a 360° shooting gallery. Their goal was to swiftly scan the entire surroundings and identify all simulated humans marked in red, while avoiding those marked in blue. Throughout most of the trials, the simulated humans were distinguishable by both a colored arm band and an AR tag floating above their heads. The AR tag took the form of either a blue rectangle or a red diamond, following basic NATO symbology, presented via the SA-147/S HWD. The design of the stimuli and task was inspired by applications that utilize AR spatial cues registered with real-world objects to enhance users' spatial tasks. Although these AR cues technically duplicate information available from the environment, the benefit lies in their clear visibility, salience, and ease of interpretation.

Each trial involved a full 360° sweep, and participants were timed for speed and instructed to maintain accuracy. Trials began with a click on a "start" button and concluded with a click on an "end" button. Participants completed each trial twice-once clockwise and once counterclockwise. The characteristics of the AR imagery presented to users varied across experimental conditions according to the tested FOVs and contrasts. Baseline trials occurred at the beginning and end of the experiment, during which no AR tags were present and participants relied solely on the colored arm bands, as if they were not wearing an AR display.

3.6 Measures

Objective Data. The duration it took for participants to perform a complete 360° sweep of the environment was recorded during each trial. Specifically, the elapsed time from when they initiated the trial by clicking the "start" button to when they signaled completion by clicking the "end" button was recorded. Additionally, the number of simulated humans marked in red that the participants overlooked (false negative) and the number of simulated humans marked in blue that they selected (false positive) was logged.

Subjective Data. Subjective responses from the participants were collected through questionnaires that they completed on a laptop after completing the trials. Participants were asked to rate their perception of the different experimental conditions with respect to the following 7-point scales using single item questionnaires:

- *Trust:* On a scale from 1 (Not Trustworthy) to 7 (Very Trustworthy), how much would you trust the AR system?
- *Reliance:* On a scale from 1 (cannot rely) to 7 (very reliable), how much reliance on the AR system would you have?

4 Results

The responses for this partial-factorial design were analyzed with repeated-measures analyses of variance (RM-ANOVAs) and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. Normality was confirmed with Shapiro-Wilk tests at the 5% level and QQ plots. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity was not supported.

No significant differences were found between the clockwise and counterclockwise trials as well as the pre and post baselines, so the responses were pooled with respect to the analysis. Overall, less than 5% selection errors (false positives and false negatives) were observed, and no significant differences in errors between any of the conditions were found, resulting in the decision to focus on reporting of the timing data and subjective responses.

4.1 Objective Data

The descriptive statistics for the elapsed trial times are shown in Fig. 3, and the statistical test results are shown in Table 1.

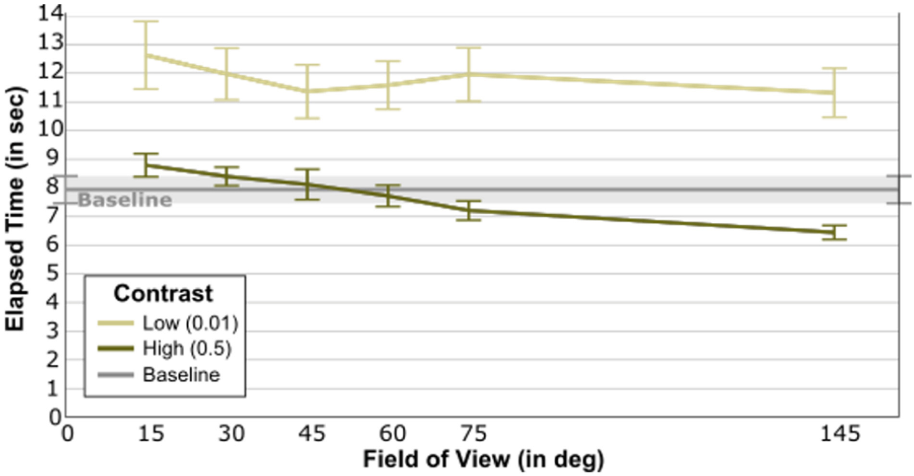
For the full-factorial analysis of the six FOVs with respect to the two visual contrasts (lowest and highest), the results show significant main effects for the FOV and visual contrasts on elapsed time. Specifically, the results indicate that elapsed times were significantly higher (i.e., worse) for the lowest contrast compared to the highest contrast. Moreover, elapsed times were significantly higher (i.e., worse) for the 15° FOV compared to the 143° FOV.

For the full-factorial analysis of the six visual contrasts with respect to the two FOVs (narrowest and widest), the results show significant main effects for the visual contrasts and FOVs on elapsed time. Specifically, the results show that elapsed times were significantly higher (i.e., worse) for the 15° FOV than the 143° FOV. Moreover, elapsed times were significantly higher (i.e., worse) for the lowest contrast (0.01) than all higher tested contrasts, and for the second-lowest contrast (0.05) than all higher tested contrasts.

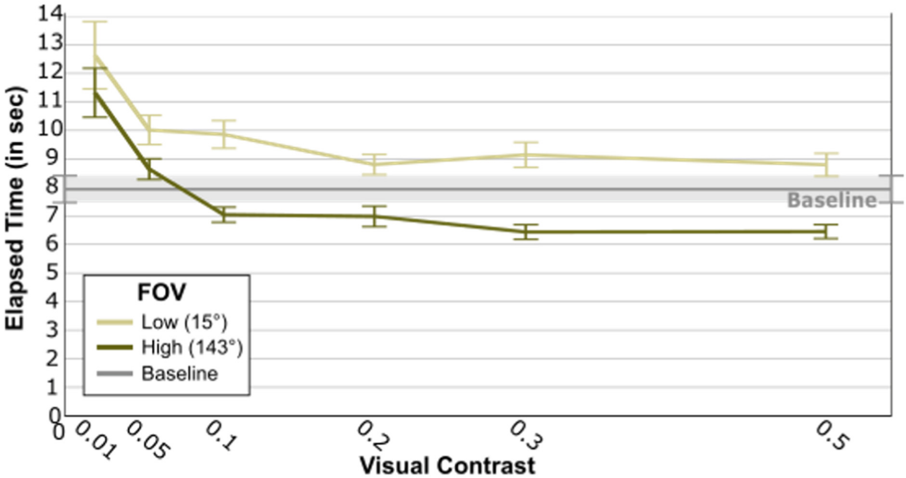
4.2 Subjective Data

The descriptive statistics for the subjective responses are shown in Fig. 4, and the statistical test results are shown in Table 1.

The results further show that with respect to the six FOVs, no significant main effect on Trust and Reliance in the single-item questionnaires was found. However, a significant main effect of the six contrast levels on trust and reliance was observed. On trust and reliance, all pairs were significant, except (0.1, 0.05), and (0.05, 0.01) contrast levels. Additionally, a significant interaction effect between the six FOVs and six contrasts on trust and reliance was found. On trust, the following two significant FOV pairs for a 0.3 contrast were found: (75° $>$ 45°), and (75° $>$ 15°).



(a)



(b)

Fig. 3. Elapsed time results structured by the (a) fields of view and (b) contrast trials and the baseline condition. Lower is better. The error bars show the standard error.

5 Discussion

The objective performance data that was collected in this experiment clearly shows that any potential benefit of AR HWDs for spatial task performance in this task are negated below a 0.1 contrast. At such low contrasts, completing the task without the use of the AR tags (see baseline in Fig. 3b) actually resulted in faster task completion than if participants tried to make sense of the barely visible AR tags that were supposed to make the task easier and faster to complete.

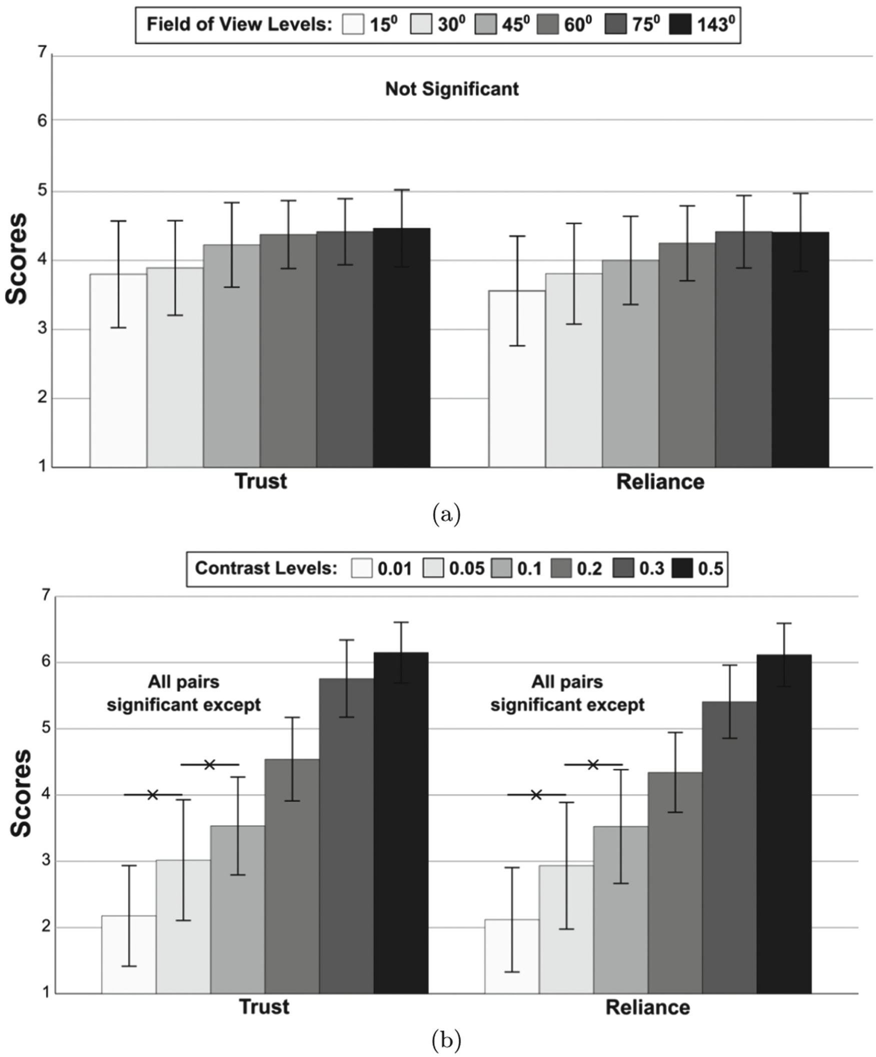


Fig. 4. Subjective data results for the six levels of (a) FOV and (b) contrast. For the different tested FOVs, no pairwise comparisons were significant. For the different tested visual contrasts, the horizontal lines with crosses in the middle indicate those pairs that were not significant ($p>0.05$). All other pairs were significant. Higher is better. The error bars show the standard error.

At the same time, the results also show that the FOV of the AR display is an important factor for task performance. Specifically, the results show that trying to rely on AR tags with a horizontal FOV below 45° slowed them down compared to not relying on AR at all (see baseline in Fig. 3a). Moreover, Fig. 3a shows that not even the performance benefit of the widest tested FOV (143°) was able to

Table 1. Statistical test results.

Measures	Factors	df_G	df_E	***F***	***p***	η_p^2
Elapsed Time	Contrast (2 levels)	1	19	27.27	**< 0.001**	0.59
	Field of View (6 levels)	5	95	4.54	**< 0.001**	0.19
	Contrast (2 levels) * Field of View (6 levels)	5	95	1.36	0.25	0.07
Elapsed Time	Field of View (2 levels)	1	19	50.41	**< 0.001**	0.73
	Contrast (6 levels)	1.55	29.49	23.63	**< 0.001**	0.55
	Field of View (2 levels) * Contrast (6 levels)	2.59	49.29	1.87	0.15	0.09
Trust	Field of View (6 levels)	1.29	23.73	3.28	0.075	0.15
	Contrast (6 levels)	2.45	46.52	41.95	**< 0.001**	0.69
	Field of View (6 levels) * Contrast (6 levels)	4.49	85.37	41.12	**0.003**	0.18
Reliance	Field of View (6 levels)	1.2	23.06	5.49	**0.023**	0.22
	Contrast (6 levels)	2.64	50.24	40.09	**< 0.001**	0.68
	Field of View (6 levels) * Contrast (6 levels)	4.09	77.71	2.87	**0.027**	0.13

compensate for the performance impairment of the lowest tested contrast (0.01), and Fig. 3b shows that not even the performance benefit of the highest tested contrast (0.5) was able to compensate for the performance impairment of the narrowest tested FOV (15°). In these cases, none of them were even close to the performance that could be gained if AR had not been used. The results show that for the effective and efficient use of AR HWDs by warfighters it is imperative for these displays to reach and exceed minimal thresholds for both contrast and FOV, or otherwise they may end up being less than valuable and helpful to warfighters for spatial tasks.

Having a well-calibrated sense of reliance and trust in an AR system is important for warfighters as it can have significant implications for their task performance as discussed above. As mentioned in the Background section, reliance and trust are related concepts, indicating how users feel about a system and whether or how they would use it. The subjective feedback and estimates that were collected in this experiment (see Fig. 4) indicate that participants judged contrast to have the largest effect on their sense of trust and reliance in the AR HWD, while the different FOVs only showed a non-significant trend with respect to these subjective ratings. This makes sense as low contrast means that the AR tags are harder to make out, making participants question what AR information they are seeing, while a narrow FOV mainly just limits the visual region in which AR information is shown but does not affect how reliable or trustworthy that information appears when it is presented.

While the subjective ratings indicate that participants generally understood and judged that "higher is better" for both the AR display contrast and FOV, looking at the magnitudes of the estimated scores shows some interesting effects. For both reliance and trust scores a 1 to 7 scale was used, higher is better, with a score of 4 indicating borderline ratings. In other words, AR systems that are striving to be perceived as helpful and useful by users should exceed borderline ratings. Just going off the magnitudes of the scores shown in Fig. 4, a trend appears that participants' ratings matched or exceeded borderline scores for horizontal FOVs of 45° or wider and contrasts of 0.2 or higher. Note that their subjective estimates were pretty on point with respect to the objective performance data for the FOV threshold at 45° shown in Fig. 3a at which performance matches or exceeds baseline levels, but Fig. 3b reveals that participants subjectively underestimated the performance benefits they gained even for lower contrasts than 0.2. Specifically, the objective results show clear performance benefits of a 0.1 contrast over the baseline, despite the subjective ratings indicating that participants had low reliance and trust in the AR system for this contrast. In other words, while objective performance may increase from a 0.1 contrast, participants may only be consciously aware of these benefits starting from a 0.2 contrast. Overall, AR HWD manufacturers should strive to provide horizontal FOVs of 45° or wider and contrasts of 0.2 or higher to be perceived as reliable and trustworthy.

6 Conclusion

This paper presented a user study (N = 20) investigating the effects of the horizontal FOV and contrast of AR OST HWDs on objective task performance and subjective ratings of reliance and trust in the AR system. Our results show both objective and subjective thresholds below which the task performance with the AR system either objectively drops or is perceived to drop below the performance that could be reached without the use of AR. Our results have implications for the design and use of AR systems, such as the apparent 0.1 contrast floor for task performance, the significance of horizontal FOV (e.g., importance of having $\geq 45°$ horizontal FOV), the necessity of meeting minimal thresholds for *both* contrast and FOV rather than focusing on one and ignoring the other, and how some AR factors may lead to a disconnect between users' perceived trust or reliance and actual task performance. Future work could include replicating these results in physical locations in the field or exploring AR display factors across different types of tasks.

Acknowledgments. This material includes work supported in part by Vision Products LLC via US Air Force Research Laboratory Award Number FA864922P1038, the Office of Naval Research under Award Numbers N00014-21-1-2578 and N00014-21-1-2882 (Dr. Peter Squire, Code 34), and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arthur, K.W.: Effects of field of view on performance with head-mounted displays. The University of North Carolina at Chapel Hill (2000)
2. Association for Computing Machinery: ACM Demographic Questionnaire (2025). https://community.acm.org/demographics/

3. Azuma, R.T.: A survey of augmented reality. Teleoperators and Virtual Environments, Presence (1997)
4. Azuma, R.T.: The challenge of making augmented reality work outdoors. Mixed Real. Merg. Real Virt. Worlds **1**, 379–390 (1999)
5. Baumeister, J.: Cognitive cost of using augmented reality displays. IEEE Trans. Visual Comput. Graphics **23**, 2378–2388 (2017)
6. Billinghurst, M.: Grand challenges for augmented reality. Front. Virt. Real. **2**, 578080 (2021)
7. Brunyé, T.T., Moran, J.M., Houck, L.A., Taylor, H.A., Mahoney, C.R.: Registration errors in beacon-based navigation guidance systems: influences on path efficiency and user reliance. Int. J. Hum Comput Stud. **96**, 1–11 (2016)
8. Cannon Jr, M.W.: Recent advances in understanding peripheral vision. In: Human Factors Society Annual Meeting, vol. 30, pp. 601–603 (1986)
9. Covelli, J.M., Rolland, J.P., Proctor, M., Kincaid, J.P., Hancock, P.: Field of view effects on pilot performance in flight. Int. J. Aviat. Psychol. **20**, 197–219 (2010)
10. Cruz-Neira, C., Sandin, D.J., DeFanti, T.A.: Surround-screen projection-based virtual reality: the design and implementation of the CAVE. Association for Computing Machinery (2023)
11. Debernardis, S., Fiorentino, M., Gattullo, M., Monno, G., Uva, A.E.: Text readability in head-worn displays: color and style optimization in video versus optical see-through devices. IEEE Trans. Visual Comput. Graphics **20**(1), 125–139 (2013)
12. Erickson, A., Bruder, G., Welch, G.F.: Adapting michelson contrast for use with optical see-through displays. In: IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 409–410 (2022)
13. Erickson, A., Kim, K., Bruder, G., Welch, G.: A review of visual perception research in optical see-through augmented reality. In: International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments (2022)
14. Erickson, A., Kim, K., Bruder, G., Welch, G.F.: Exploring the limitations of environment lighting on optical see-through head-mounted displays. In: ACM Symposium on Spatial User Interaction, pp. 1–8 (2020)
15. Erickson, A., Kim, K., Lambert, A., Bruder, G., Browne, M.P., Welch, G.: An extended analysis on the benefits of dark mode user interfaces in optical see-through head-mounted displays. ACM Trans. Appl. Percept. **18**(3), 22 (2021)
16. Faul, F., Erdfelder, E., Lang, A.G., Buchner, A.: G* power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods **39**(2), 175–191 (2007)
17. Gabbard, J.L., Swan, J.E., Hix, D.: The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. Presence **15**, 16–32 (2006)
18. Gabbard, J.L., Swan, J.E., Zarger, A.: Color blending in outdoor optical see-through ar: the effect of real-world backgrounds on user interface color. IEEE Virt. Reality. 157–158 (2013)
19. Gattullo, M., Uva, A.E., Fiorentino, M., Monno, G.: Effect of text outline and contrast polarity on ar text readability in industrial lighting. IEEE Trans. Visual Comput. Graphics **21**(5), 638–651 (2014)
20. Gottsacker, M., et al.: Investigating the relationships between user behaviors and tracking factors on task performance and trust in augmented reality. Elsevier Comput. Graph. **123**, 1–14 (2024)

21. Gottsacker, M., Norouzi, N., Schubert, R., Guido-Sanz, F., Bruder, G., Welch, G.F.: Effects of environmental noise levels on patient handoff communication in a mixed reality simulation. In: ACM Symposium on Virtual Reality Software and Technology (VRST), pp. 1–10 (2022)
22. Jones, J.A., Swan, J.E., Singh, G., Ellis, S.R.: Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In: ACM Symposium on Applied Perception in Graphics and Visualization, pp. 29–36 (2011)
23. Kim, K., Erickson, A., Lambert, A., Bruder, G., Welch, G.: Effects of dark mode on visual fatigue and acuity in optical see-through head-mounted displays. In: ACM Symposium on Spatial User Interaction, pp. 1–9 (2019)
24. Kishishita, N., Kiyokawa, K., Orlosky, J., Mashita, T., Takemura, H., Kruijff, E.: Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 177–186 (2014)
25. Kruijff, E., Swan, J.E., Feiner, S.: Perceptual issues in augmented reality revisited. In: IEEE International Symposium on Mixed and Augmented Reality, pp. 3–12 (2010)
26. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors **46**(1), 50–80 (2004)
27. Lee, J., Rheem, H., Lee, J.D., Szczerba, J.F., Rajavenkatanarayanan, A., Mathieu, R.: Sharing vehicle situation awareness reduces driver-initiated overrides in urban environments. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 67, pp. 1178–1183. SAGE Publications Sage CA, Los Angeles, CA (2023)
28. Liu, A., Acosta, E., Cope, J., Henry, V., Reyes, F., Bradascio, J., Meek, W.: The wide area virtual environment: a new paradigm for medical team training. In: Augmented cognition: users and contexts: 12th International Conference, AC 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, 15–20 July 2018, Proceedings, Part II, pp. 293–304. Springer (2018)
29. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Acad. Manag. Rev. **20**(3), 709–734 (1995)
30. Merlo, J.L.: Effect of reliability on cue effectiveness and display signaling. University of Illinois at Urbana-Champaign (1999)
31. Mosier, K.L., Skitka, L.J., Heers, S., Burdick, M.: Automation bias: decision making and performance in high-tech cockpits. Int. J. Aviat. Psychol. **8**, 47–63 (1998)
32. van den Oever, F., Sætrevik, B., Fjeld, M., Nordby, K.: A virtual reality experiment shows that augmented reality can improve collaboration in ship navigation (2023)
33. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Factors **39**, 230–253 (1997)
34. Ragan, E., Bowman, D., Kopper, R., Stinson, C., Scerbo, S., McMahan, R.: Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. IEEE Trans. Visual Comput. Graphics **21**(7), 794–807 (2015)
35. Raikwar, A., et al.: Beyond the wizard of oz: negative effects of imperfect machine learning to examine the impact of reliability of augmented reality cues on visual search performance. IEEE Trans. Vis. Comput. Graph. (2024)
36. Sabelman, E.E., Lam, R.: The real-life dangers of augmented reality. IEEE Spectr. **52**(7), 48–53 (2015)
37. Schubert, R., Bruder, G., Tanaka, A., Guido-Sanz, F., Welch, G.F.: Mixed reality technology capabilities for combat-casualty handoff training. In: Chen, J., Fragomeni, G. (eds.) HCII 2021. LNCS, vol. 12770, pp. 695–711. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77599-5_47

38. Schubert, R., Welch, G., Daher, S., Raij, A.: HuSIS: a dedicated space for studying human interactions. IEEE Comput. Graphics Appl. **36**(6), 26–36 (2016)
39. Sorkin, R.D.: Why are people turning off our alarms? J. Acoust. Soc. Am. **84**, 1107–1108 (1988)
40. Steinicke, F., Bruder, G., Kuhl, S., Willemsen, P., Lappe, M., Hinrichs, K.: Natural perspective projections for head-mounted displays. IEEE Trans. Visual Comput. Graphics **17**(7), 888–899 (2011)
41. Trepkowski, C., Eibich, D., Maiero, J., Marquardt, A., Kruijff, E., Feiner, S.: The effect of narrow field of view and information density on visual search performance in augmented reality. In: IEEE Virtual Reality, pp. 575–584 (2019)
42. Welch, G.F., Bruder, G., Squire, P., Schubert, R.: Anticipating widespread augmented reality: insights from the 2018 AR Visioning Workshop. Technical Report 786, University of Central Florida and Office of Naval Research (2019)
43. Yeh, M., Wickens, C.D.: Display signaling in augmented reality: effects of cue reliability and image realism on attention allocation and trust calibration. Hum. Factors **43**(3), 355–365 (2001)

Beyond Videoconferencing: How Collaborative Tools Make Virtual Design Reviews Work

Francisco Garcia Rivera[1](✉), Asreen Rostami[2,3], Huizhong Cao[4], Dan Högberg[1], and Maurice Lamb[1]

[1] University of Skövde, Skövde, Sweden
francisco.garcia.rivera@his.se
[2] RISE Research Institutes of Sweden, Gothenburg, Sweden
[3] Stockholm University, Stockholm, Sweden
[4] Chalmers University of Technology, Gothenburg, Sweden

Abstract. This study examines how specific collaborative features for independent viewpoint control, collaborative pointing, sketching, and manikin representations support design reviews in a virtual environment. Participants conducted design reviews using the collaborative design tool Gravity Sketch while engaging with these features. Results show that friction situations were minimal, with viewpoint control eliminating the need for perspective adjustments. Collaborative pointing and sketching effectively clarified design modifications, while manikin representations aided discussions on ergonomics. The findings highlight how the evaluated features facilitate communication in remote design reviews.

Keywords: Design Reviews · Remote Collaboration · Product Development

1 Introduction

Design reviews (DRs) are important checkpoints in the product design and development process. During a DR, engineers, designers, and other stakeholders evaluate various aspects of product design proposals, including performance, functionality, and manufacturability at various stages [1, 2]. The primary goal of DRs is to identify potential design flaws or issues before advancing to subsequent phases [3], thereby enhancing overall product quality and reducing the risk of downstream problems. Additionally, DRs serve as collaborative platforms where stakeholders with diverse roles and interests work together with engineers and designers to refine design approaches that meet agreed-upon requirements [4]. Although DRs have traditionally been conducted in physical settings, they are increasingly transitioning to remote and hybrid formats due to the rise of distributed work environments [5]. While these new formats enhance accessibility and scheduling flexibility, they also introduce challenges, particularly in sustaining the continuity of discussions. Factors such as technical difficulties reduced non-verbal cues, and the potential for disengagement and delays in remote distributed settings can hinder the natural flow of conversation, making it more difficult to maintain effective collaboration.

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 96–112, 2025.
https://doi.org/10.1007/978-3-031-93715-6_7

A key aspect of DRs is visual support, which plays a crucial role in driving discussions and decision-making [6]. Visual elements, including drawings, 3D models, and physical prototypes, provide participants with essential reference points and something to "*look at*" during the DR. However, in remote DRs and hybrid settings, effectively integrating and interacting with these visual supports requires new approaches to ensure discussions remain productive and well-structured. When participants are remotely located, the most common method for providing visual support involves using computer-aided design (CAD) software in combination with screen-sharing in videoconferencing tools [7]. However, this approach often disrupts the flow of communication [8]. Usually, a single user (the presenter) controls the shared viewpoint, while other participants provide verbal instructions to engage with the content. This constraint can hinder collaboration because the presenter must interpret instructions and manually adjust the shared view, leading to potential delays and misunderstandings, especially when requested views are complex or need attention to complicated details. In this context, misunderstandings and delays create friction situations, disrupting the continuity of interaction and impeding smooth collaboration.

Friction situations are interruptions in workflow that arise from technical or procedural barriers during remote DRs. Through a previous study, we identified four types of recurring friction situations that disrupt collaboration and decision-making in distributed DRS [8]:

- Requesting specific viewpoints: The need for specific perspectives on the design, requiring the presenter to adjust the view manually.
- Indicating specific elements: The challenge of pointing to or highlighting specific design features without direct interaction.
- Expressing changing design ideas: Difficulty in conveying modifications and adjustments efficiently.
- Evaluating physical ergonomics: The inability to assess human interaction with the proposed design due to limited visualization tools.

Because friction situations disrupt remote DRs and because remote DRs have become common, there is a need to identify methods for reducing friction situations in remote DRs. A commonly proposed solution is implementing extended reality (XR) technologies, which can improve certain aspects of collaboration in DRs, such as real-time spatial awareness, intuitive interaction with 3D models, a better understanding of design change ideas, and enhanced remote collaboration [7, 9–11]. When conducting DRs with XR technologies, DRs occur in collaborative virtual environments (CVEs), digital spaces where users can gather and interact. CVEs can be accessed through a range of devices, from standard laptops to immersive augmented reality (AR) and virtual reality (VR) systems [12, 13]. As such, unlike traditional screen-sharing solutions, which assume 2D screen-based interactions adapted to 3D CAD tools, CVEs are designed assuming 3D navigation and interaction by default. Thus, while XR devices offer advantages by facilitating natural interactions with 3D content, much of the improvement in collaboration in DRs may come from the CVE itself and not the specific technology used to access the CVE. Unlike traditional screen-sharing in video conferencing applications, a CVE provides participants with a greater sense of agency and autonomy, allowing them to explore and manipulate visual content independently rather than relying on the presenter's-controlled viewpoint.

Building upon our previous findings [8], this study investigates the dynamics of friction situations within a CVE during DRs. Specifically, we examine a commercially available CVE design tool not explicitly made for DRs, Gravity Sketch, to investigate how users leverage various collaborative features that are potentially important to mitigate or eliminate friction situations in DRs, including:

- Advanced Viewpoint Control: This feature allows users to navigate the 3D model independently or to follow another participant's perspective. Additionally, users can request that others follow their views. In the Gravity Sketch implementation, the decision to control one's viewpoint or to follow another's is left to the user rather than being centrally managed, and users can only have one viewpoint at a time. Furthermore, while VR users cannot continuously follow others, they can move to a specific location in space to align with another participant's view. Users can also request others to follow their viewpoint, and this can be accepted or rejected by others in the CVE.
- Collaborative Pointing Tools: This feature enables users to point out elements or specific locations within the CVE so everyone's attention can be drawn to the same location or elements. In Gravity Sketch, a pointer indicates a specific location in space, though it does not highlight individual elements. The pointer is available to users on 2D devices, while VR users can either use a laser pointer or extend their hand or controller towards an object, others in the virtual space can follow their gesture.
- Manikin Representations: Manikins facilitate ergonomics evaluations of product design proposals by integrating digital human figures. Gravity Sketch offers male and female manikins whose postures can be adjusted (in VR only). Although the models do not reflect anthropometric diversity, they can be uniformly rescaled, meaning that all body measurements adjust proportionally. There are also 3D assets of specific body parts, including the torso and head.
- Sketching Tools: These tools support real-time drawing directly within the CVE. VR users can sketch in 3D and have various options for quick sketching, while 2D users can only do basic sketches in a 2D place.

While remote DRs offer increased accessibility and flexibility, they also introduce friction situations that can disrupt collaboration and impede decision-making. Shifting to a CVE can provide a viable alternative to conducting DRs by facilitating more natural, interactive engagement with 3D content. Our previous research identified the aforementioned friction situations and these key features that have the potential to reduce them or eliminate them. In this study, we investigate how the implementation of these features affects the occurrence and resolution of friction during DRs.

2 Method

This study utilized Gravity Sketch[1], an XR-native design tool, as the platform for conducting DRs. Gravity Sketch integrates key collaborative features and supports simultaneous use of XR devices (e.g., Meta Quest 3) and 2D screens, making it a suitable choice

[1] Gravity Sketch Limited, London, UK: https://gravitysketch.com.

for this experiment. Its capabilities allowed us to observe how participants exploited its features to overcome friction situations [8].

In Gravity Sketch, the CVE (referred to as "CollabRooms" by Gravity Sketch) is accessible via invitation on both XR devices and 2D devices. Users on 2D devices must install the *Gravity Sketch ScreenCollab* software to join. Within the CVE, XR users are represented as headsets with two controllers, as shown in Fig. 1, while 2D users are represented as small squares with a line underneath, as shown by the red squares in Fig. 2. For this experiment, one participant used a Meta Quest 3, and two participants joined using laptops connected to an external display.

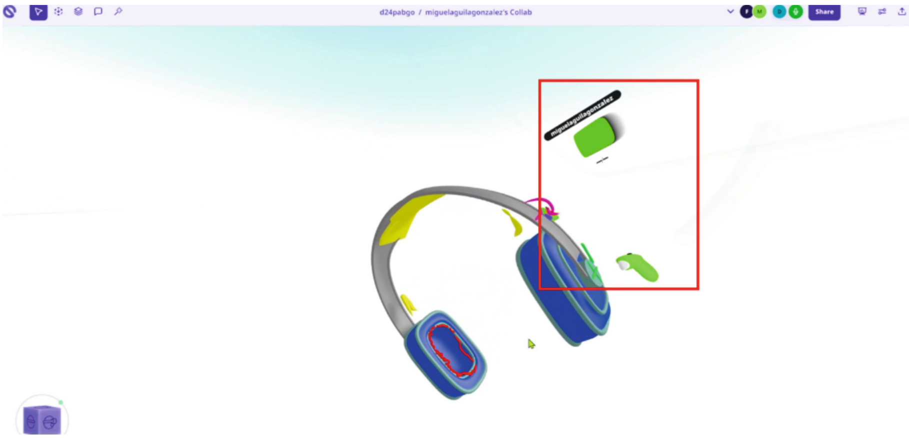


Fig. 1. 2D participant view of the CVE. The XR participant is represented inside of the red square. (Color figure online)

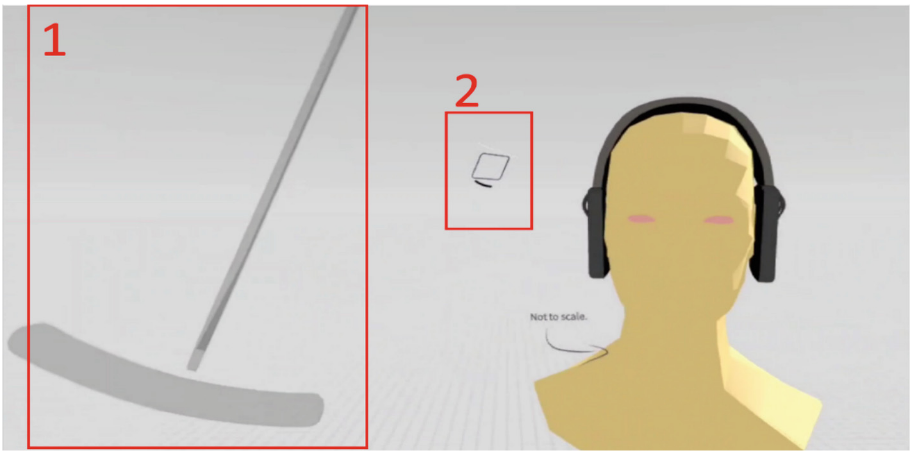


Fig. 2. XR participant view of the CVE. The 2D participants are represented inside of the red squares. (Color figure online)

2.1 Participants

A common challenge in XR experiments is recruiting participants with both hardware knowledge and software proficiency. In this study, participants were final-year BSc Product Design Engineering students enrolled in the *Digital Tools in Product Development* course at University of Skövde. As part of their coursework, the authors provided Gravity Sketch training, which began two months before the experiment. Each group of four students was assigned a Meta Quest 3 headset, requiring them to share and take turns using it. Students were trained in Meta Quest 3 and ScreenCollab for Gravity Sketch to ensure active participation. Students were encouraged to use *ScreenCollab* throughout the course for collaboration while alternating headset use. Additionally, students practiced DR-style collaborative tasks in the CVE, gaining hands-on experience in using the software for teamwork.

By the time of the experiment, all participants were proficient in Gravity Sketch, enabling them to fully utilize its features during the DRs. As part of their coursework, they were tasked with designing over-ear headphones, providing a relevant context for the DRs. Out of 18 students in the course, nine students (five males and four females, aged 20–45) volunteered for the experiment, forming three groups of three. Participating students were informed that participation in the study would not affect their course grade and written consent was obtained. Students were informed that they could end the study task at any time, though all groups completed the task. To maintain familiarity with the design and ensure established team dynamics, groups remained the same as in coursework, ensuring prior experience in CVE collaboration.

Each DR focused on reviewing the group's headphone design, aiming to collaboratively identify around five key action items for improvement. The role of the XR participant was assigned through voluntary selection within each group.

2.2 Procedure

Each session began with an introduction, during which the three participants were briefed by the first author on the purpose of the study and the task they were about to undertake. The experiment took place with everyone in the same physical room, seated, as shown in Fig. 3 One participant in XR drove the conversation, and two others sat each with laptops connected to an external display. The 2D participants were facing each other while the XR participant stood next to them. Participants could not see each other screens.

Following the briefing, participants were introduced to their task: conducting a DR of their headphone designs realized as 3D models within their course. They were instructed to evaluate the design from various perspectives, focusing on key elements such as the comfort of the headband, the positioning of the ear cups, the adjustability of components, and the distribution of pressure across the head and ears. They were also reminded of the collaborative features available in Gravity Sketch, both in XR and in the ScreenCollab version, especially the possibility of following someone's view, the collaborative pointing tool, the manikin representation, and the sketching tool. They were not specifically told which features they should use but were reminded that the features were available.

A researcher was present in the physical room (sitting between the XR participant and the 2D participants in one of the corners to not enter the XR safe area), observing the discussions but not directly intervening in the selection of action items. While the researcher did not intervene in the discussions or influence the specific action items, they guided the conversation when suggestions were too general. This was necessary because overly broad action items often lacked references to specific shapes, locations, components, or perspectives—factors that typically contribute to friction situations. For instance, if a participant suggested, "*make the headphones more comfortable*" or "*make the headphones foldable*". The action items had to have a narrow focus that could be actionable (i.e., clear instructions on what to do). One accepted action item could be, "*Implement an axis of rotation at the end of the headband so that the headphones can be folded around the longitudinal axis when wearing them around the neck*", or "*Add foam material to the locations marked in red in the headband to distribute weight across the skull better*". It was also necessary that participants understood how the action item had to be implemented.

After the session, participants were given additional information about the study objectives, and there was a debriefing for questions and concerns.

2.3 Data Collection

The data collected during the DR sessions comprised multi-source recordings to allow for a comprehensive analysis of participant interactions and software feature usage after the sessions. Data was collected through screen recordings which included audio. OBS Studio[2] was used for screen recording on the laptops. The Meta Quest 3 viewpoint was recorded using the Meta Horizon companion app for smartphones.

Once all recordings were collected, they were imported into DaVinci Resolve software for synchronization. The three video streams were aligned (using the common audio data since all participants were in the same room) to visualize the session from all perspectives, as shown in Fig. 3. Furthermore, the audio recordings were transcribed using a locally hosted instance of OpenAI's Whisper model for the analysis. The combined video and audio datasets were the primary data used to analyze the DR session and code the friction situations.

2.4 Data Analysis Method

The data analysis aimed to determine whether the friction situations identified in previous research [8] emerged during these DRs and, if so, to understand how participants used specific collaborative features to overcome the friction situations to ensure an efficient design review process. In this context, friction situations are defined as disruptions or barriers to smooth communication and workflow that arise from technological or procedural constraints during a DR. To provide a structured analysis, we focused on four main categories of friction challenges:

- Requesting Specific Viewpoints: This friction occurs when participants need to view particular perspectives of the 3D model to understand better or discuss design elements. In traditional remote DRs, a central presenter typically controls the viewpoint, meaning that participants must request adjustments to see the 3D model from

[2] OBS Studio, https://obsproject.com/.

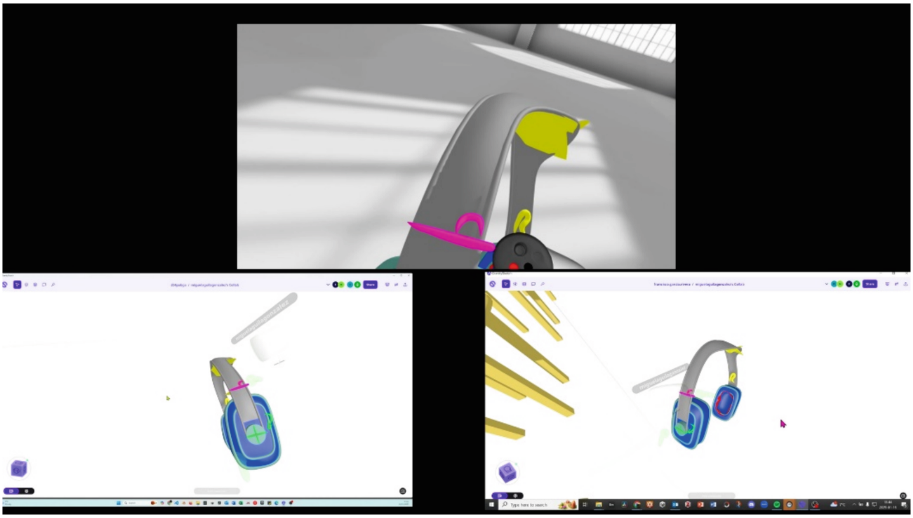


Fig. 3. Synchronized view for analysis. The top panel is the view from the XR participant. The bottom panels are the views from the 2D participants.

the desired location in the CVE. Such requests can lead to *delays* and *interruptions*, especially when the desired viewpoints are complex, multiple participants request different perspectives simultaneously, or the shared control mechanism limits immediate access.

- Indicating Specific Elements: This category refers to situations where participants encounter difficulties pointing out or referring to particular components within the 3D model. In complex designs, elements can be obscured or appear similar to surrounding parts, leading to *misunderstandings* about which component is under discussion. The friction situation can be worsened when the indication is done verbally or without adequate visual support, thus hindering the natural flow of the discussion.
- Expressing Design Change Ideas**:** Friction in this category arises when a participant proposes modifications to the design, but the idea is not immediately clear to the rest of the team. The challenges here come from the inherent difficulty of conveying spatial and design nuances through verbal descriptions alone. When participants *struggle to communicate* their design change ideas effectively, additional tools, including real-time sketching or screenshots, may be needed to bridge the gap, although their use can sometimes introduce further complexity or delay.
- Evaluating Ergonomics: This friction situation involves difficulties in assessing how well a design accommodates human use. Ergonomics evaluations require understanding spatial relationships and human interaction with the design. These assessments can be challenging on 2D screens or when using simplified digital manikins, leading to *prolonged discussions* as participants attempt to reach a consensus on issues such as comfort, reachability, and usability. Limited level of accuracy provided by the manikin representations, e.g. in terms of representing anthropometric details or diversity within the targeted user population, can worsen these challenges.

To understand how these friction situations arose or were avoided, we examined the use of four collaborative features within Gravity Sketch described earlier: advanced viewpoint control, collaborative pointing tools, manikin representations, and sketching tools. A coding protocol was developed to track friction situations and use of the target collaborative features in Gravity sketch. This included recording the following information for identified friction situations:

- Time Stamp: The onset and offset times (using the [mm:ss] format).
- Friction Category: One of the four types: Requesting Specific Viewpoints, Indicating Specific Elements, Expressing Design Change Ideas, or Evaluating Ergonomics.
- Brief Description: A summary of the situation to provide context, including what the user was trying to communicate and how the other participants understood (or misunderstood).
- Collaborative Feature Utilized: Documentation of any tool use (e.g., advanced viewpoint control, collaborative pointing, manikin representations, or sketching tools), employed to address or mitigate the friction.

The data analysis process started by visualizing the videos several times to get familiar with the data. After that, the videos were watched in more detail to transcribe the friction situations according to the template. Every time a friction situation was identified, the video was paused, and the specific timestamps limiting the friction event were documented. All screen recordings were visible simultaneously, allowing for an analysis of what each participant was doing when the friction occurred and why it happened. If there was a misunderstanding, the recorded footage provided insights into what one participant intended to convey versus how others interpreted it. These friction instances were reviewed multiple times. Each section was replayed to confirm its category, provide a detailed description, and identify features used during the friction situations. The first author conducted all the coding. Additionally, the footage was reviewed to understand tool usage during smooth interactions, whether the features were part of the workflow or used only in cases of misunderstanding, and to clarify the purposes for which the different features were employed. Although these instances were not documented in the template and were gathered in the form of unstructured notes, this review provided insight into how features were used proactively, beyond simply resolving friction situations.

3 Results

Each DR session lasted approximately 15 min. A total of 10 friction situations were observed, primarily related to expressing design change ideas, indicating specific design elements, and evaluating ergonomics. Two sessions recorded three friction situations each, while the third session had four. The recorded friction situations are presented in Table 1. These friction situations were relatively rare and when they did occur, they were quickly resolved using the available collaborative features.

The collaborative pointing tool was the most frequently used feature for resolving friction situations (see Table 2). It was particularly effective when participants struggled to indicate specific elements or express design changes. Additionally, it played a crucial role in confirming proposed modifications, enabling users to quickly validate

Table 1. Recorded friction situations.

Friction Situation	Total Instances	Effectiveness of Collaborative Features
Expressing design changes	5	Quickly resolved in most cases using collaborative pointing and VR sketching
Indicating elements	3	Occurred mainly when participants forgot to use the pointing tool; quickly corrected when used
Evaluating ergonomics	2	Manikin representations helped clarify usability concerns immediately
Requesting viewpoints	0	No issues reported, indicating viewpoint control tools worked as intended

whether they were referring to the same component or spatial information. In five separate instances, collaborative pointing effectively clarified intended changes, swiftly resolving misunderstandings. The sketching tools were used three times to address friction situations, primarily when participants needed to illustrate modifications that were difficult to describe verbally. Manikin representations were employed twice, exclusively in ergonomics evaluations. While independent viewpoint control was frequently utilized throughout the sessions, especially when identifying new action items, it was not directly tied to resolving friction situations. Table 2 provides a breakdown of how these features contributed to overcoming friction situations.

Table 2. Feature used to overcome friction situations. Note that the total usage per feature only refers to total usage to solve friction situations, the features (especially independent viewpoint) were constantly used during the DRs.

Collaborative Feature	Total Usage (to resolve friction)	Primary Role
Collaborative pointing	5	Helped clarify intended changes and indicate specific design elements
Sketching tools	3	Used for visually explaining complex design changes, preventing miscommunication
Manikin representations	2	Addressed ergonomics concerns, reducing ambiguity in usability discussions
Independent viewpoint	0	Not tied to friction resolution but enabled participants to explore models autonomously

The following section provided further detail on how each collaborative feature was used.

3.1 Collaborative Pointing

Proactive vs Reactive Use. This tool was used both proactively and reactively. When used proactively, participants directed attention to specific elements or directions, preventing friction situations from arising in the first place. In these cases, the tool was an intuitive means of communication, seamlessly integrating into the workflow without causing disruptions. In contrast, when used reactively, collaborative pointing solved misunderstandings. Participants employed the pointing tool to visually clarify their statements if verbal explanations did not convey an idea effectively. This helped resolve confusion and allowed for smoother communication.

Friction Situation Addressed. The friction situations addressed are explained in Table 3.

Table 3. Friction situations addressed by collaborative pointing.

Friction Situation	Example
Indicating Specific Elements: Collaborative pointing was mostly used to refer to the specific element being discussed, especially in cluttered zones with multiple components	During one session, a participant stated, "*We should make this button smaller…*" while discussing a control panel with four buttons. Initially, they hovered over the button with their mouse without using the pointing tool, which led to the other two participants responding, "*I don't know what button you are talking about.*" Once the pointing tool was employed, the intended button was immediately identified, and the discussion moved forward without further confusion
Expressing Design Changes: Collaborative pointing was also used to illustrate design changes, particularly those involving movement or rotation	In one instance, a participant said, "*We should make the headphones foldable like this…*" while pointing in a specific direction. This allowed the group to visualize the intended adjustment and confirm agreement before proceeding with further refinements. In most cases, this usage did not introduce friction. However, if the design idea was too complex, other features (mostly sketching) were used in combination to clarify the concept

3.2 Sketching Tools

Proactive vs Reactive Use. XR participants initiated multiple sketches throughout the sessions, often drawing freehand annotations, highlights, or rough modifications to support their verbal descriptions. These sketches were mostly created spontaneously, and in many cases, XR participants would make a quick sketch and then erase it once the point was communicated.

2D participants used sketching reactively, mostly in response to confusion arising during discussions. When a design change idea was unclear, 2D participants would attempt to use the sketching tool to clarify their thoughts. However, this usage was far less frequent than among XR participants, as 2D participants typically relied more on verbal explanations and pointing. Moreover, there were instances where the 2D participants began to sketch something, but the XR participant took over sketching based on the vague explanations that the 2D participants provided. Sometimes, rough feedback was needed to refine the sketch, but the XR participants mostly did the sketch.

Friction Situations Addressed. The friction situations addressed are explained in Table 4.

Table 4. Friction situations addressed by sketching tools.

Friction Situation	Example
Expressing Design Changes: Sketching tools were primarily used to illustrate modifications that were difficult to describe verbally	A 2D participant was discussing the placement of padding in the frame of the headphones. However, the other participants did not properly understand the proposed change This led to a prolonged exchange where the other participants misunderstood how the padding would be distributed. The 2D participant began using the sketching tool to clarify their idea and the XR participant also began to draw a rough outline of the intended placement and shape. The 2D participant deferred to the XR participant. Although some feedback was needed to capture the 2D participant's idea, the sketch was fully done by the XR participant
Indicating Specific Elements (XR Participants Only): While the sketching tool was used several times as a pointing tool, one friction situation was recorded when the feature was used to indicate specific elements	During a discussion concerning the structural integrity of a hinge for folding the headphones, an XR participant attempted to verbally indicate a specific weak point but was met with uncertainty from the team. To clarify, they circled the joint using the sketching tool

3.3 Manikin Representation

Proactive vs Reactive Use. Manikins were not proactively used; their use was exclusively reactive when ergonomics concerns could not be solved with verbal information alone. Because of the nature of the product, participants only used partial human representations (in this case the head) as opposed to the full manikin.

Friction Situations Addressed. The friction situations addressed by manikin representations are addressed in Table 5.

Table 5. Friction situations addressed by manikin representation.

Friction Situation	Example
Evaluating Ergonomics: Manikin representations were effective in resolving usability, mostly related to the fit on the ears or the curvature of the frame around the head	A 2D participant raised concerns about whether the headband of the headphones would distribute pressure evenly. The other participants did not share this concern, leading to uncertainty about the design. Moreover, the others were unsure if a proposal of additional padding material on the sides would impact overall fit. A manikin representation was introduced to simulate a human head, allowing the group to observe contact points and confirm that adjustments were necessary

3.4 Viewpoint Control

Proactive vs Reactive Use. All of the viewpoint control usage was proactive, as participants freely adjusted their perspectives without requiring intervention. Some 2D participants relied heavily on the follow view feature, almost always following someone else's viewpoint. In contrast, others rarely, if ever, followed anyone, preferring to navigate independently. There were instances where a conversation was ongoing, but a (participant who was not involved in the discussion) might have been focusing elsewhere, possibly not paying attention. These moments did not cause misunderstandings but indicated variations in engagement strategies.

4 Discussion

The study's findings demonstrate that CVE can significantly enhance remote DRs for both XR and 2D screen-based users by reducing communication breakdowns and workflow disruptions. However, beyond their immediate functionality, a broader examination of their implications on collaborative design, participant behavior, and system design principles offers valuable insights for future development.

4.1 Rethinking Remote Collaboration: Beyond Passive Viewing to Active Participation

Traditional remote DRs often limit participants to passive viewing of shared screens, requiring verbal exchanges to navigate and discuss designs. In contrast, CVEs redefine remote participation by enabling direct engagement with 3D content, enabling active participation rather than passive observation. The ability to freely navigate, point, sketch, and manipulate objects shifts the role of participants from spectators to co-creators. Active engagement in CVEs is not limited to XR participants with many features of CVEs benefiting active engagement for 2D screen-based users as well. Future collaborative system designs, such as the one explored here, can focus on facilitating active, multi-user engagement, ensuring that all participants can contribute equitably rather than relying on a single presenter to dictate the discussion.

4.2 Balancing Autonomy and Synchronization in Collaborative Workflows

Another key takeaway is the importance of flexibility in viewpoint control, allowing participants to choose between independent exploration and synchronized viewing. However, the occasional disengagement of participants when navigating independently suggests that unstructured autonomy can sometimes lead to fragmented discussions. One way to address this would be the design of CVEs that integrate adaptive synchronization features, where viewpoints can be temporarily anchored or guided during critical discussions without restricting participants' ability to navigate freely when needed. For instance, drawing inspiration from the creative practices of artists who guide the attention of XR users toward specific tasks or virtual objects [14], a soft-guidance system could be developed to provide non-intrusive cues. This system would subtly direct users' focus toward relevant discussion points while still allowing for autonomy when appropriate. Such an approach could enhance collaboration by ensuring that critical aspects of the design review receive collective attention without imposing rigid viewpoint restrictions.

4.3 Intuitive Tool Discoverability and Proactive Assistance

While the study found that collaborative tools like pointing and sketching effectively resolved friction situations, their inconsistent usage suggests that participants may overlook available tools in dynamic discussions. As the modeling of objects in CVEs and their digital representation becomes more complex—such as mechanical assemblies with interdependent moving parts—the technical challenge of rendering and delivering these environments to XR devices in real time increases. Ensuring that participants can manipulate objects seamlessly while maintaining a coherent sense of movement and causality becomes particularly critical in XR, where the temporal alignment of actions differs from traditional 2D interfaces. Unlike users interacting with virtual objects on a 2D screen, XR participants may experience delays or preemptions when interacting with the same object. These discrepancies arise due to the need for real-time spatial rendering, affecting both responsiveness and the perceived continuity of actions.

However, while latency is a challenge in both 2D and XR in CVE, XR introduces additional complexities due to its spatial and interactive nature. In XR, slight delays in rendering or synchronization can create a disconnect between user input and system response, leading to interaction breakdowns. These timing discrepancies are particularly relevant when multiple users interact with the same digital object in a CVE, as maintaining a shared, temporally consistent experience requires precise real-time updates [15]. Thus, usability challenges arise not just from ensuring tool availability but also from making tools discoverable and seamlessly integrated into workflows. Addressing these challenges requires strategies such as context-aware tool suggestions to prompt users when verbal descriptions cause ambiguity, shortcut-based activations like gesture-based or multimodal interactions to reduce cognitive effort, and subtle system nudges that highlight underutilized tools without disrupting the workflow.

4.4 Equitable Collaboration Between XR Users and 2D Users

The observation that XR participants predominantly performed sketching tasks suggests that CVEs need to address, or at least make explicit, role imbalances between device types. While XR provides a more immersive interaction mode, non-XR users must be equipped with equally intuitive input methods to contribute effectively. For example, Gugenheimer et al. [16] have previously highlighted both the opportunities and challenges of designing a shared, collaborative mixed-reality game for XR users and 2D users. Potential solutions for DR could include enhanced 2D sketching interfaces, such as stylus or touchscreen support, to enable 2D users to contribute more fluidly; cross-platform consistency to ensure that tool accessibility and interaction remain device-agnostic rather than XR-centric; and shared control dynamics, allowing 2D and XR users to co-edit sketches simultaneously, fostering truly collaborative workflows instead of hierarchical interaction patterns.

On the other hand, the reactive use of manikin representations suggests that ergonomics considerations are often overlooked until explicitly needed, highlighting a broader issue in design workflows. Usability and human-centered factors tend to emerge late in the design process rather than being proactively integrated from the outset.

One way to address this is by embedding ergonomics evaluation prompts at relevant points in the discussion, encouraging teams to consider human factors earlier. Additionally, real-time ergonomics simulation overlays can allow designers to visualize ergonomics constraints dynamically rather than treating them as a separate step. Another approach involves scenario-based analysis tools, where usability tests within the CVE simulate real-world interactions based on predefined ergonomics benchmarks, ensuring that ergonomics considerations become an integral part of the design review process rather than an afterthought.

4.5 Limitations and Future Work

While this study provides valuable insights into the effectiveness of CVEs in DRs, some limitations should be acknowledged. One consideration is the relatively small number of participants. While more participants would be valuable, the use of recently trained final-year product design students with substantial experience using Gravity Sketch, as well as direct familiarity with the product design they were reviewing, allows for observation of realistic behaviors of trained professionals.

Another limitation is that the study focused on a single phase of the design process, specifically an early-stage design review of over-ear headphones. This phase involved identifying key improvements and potential modifications rather than refining final design details. While this is a relevant and necessary stage of product development, later phases (where designs are more constrained by manufacturing and engineering considerations) may present different friction situations and collaboration needs. Future research could explore how CVEs support more advanced design stages, where precision and technical feasibility play a larger role in decision-making.

Additionally, the study was conducted exclusively within Gravity Sketch. While this platform is one of the most advanced available for collaborative design in XR, it represents only one possible implementation of a CVE. At present, few alternatives offer comparable feature sets, particularly in terms of supporting both VR and 2D collaboration. However, as XR collaboration tools continue to evolve, future studies could examine whether different CVE implementations introduce unique interaction patterns or if the results observed here are consistent across multiple platforms. Finally, the study lacks a comparison group to represent the traditional DRs without Gravity Sketch CVEs features. This could lead to less qualitative/quantitative comparison to specify which friction situations are eliminated the most. But with our previous study and references for traditional DRs, [8] we could still have a preliminary assessment of the CVEs with practical study and observations.

Despite these limitations, this study provides strong evidence that CVEs, when equipped with well-integrated collaborative tools, and utilized by trained users can be a viable alternative to carry out DRs.

5 Conclusion

This study highlights the effectiveness of CVEs in supporting remote DRs involving both XR and 2D screen-based users, demonstrating that interactive tools such as collaborative pointing and sketching facilitate communication and decision-making in DRs. One of the most notable findings was the absence of friction related to viewpoint control. Participants could navigate the 3D model independently or follow another's perspective when necessary, ensuring that discussions remained focused on the design rather than on the logistical challenges of adjusting viewpoints. This suggests that giving participants autonomy on how to manage their viewpoint, makes smooth collaboration in DRs regardless of their use of XR or 2D screen-based technologies.

Friction situations were rare. When they did occur, participants resolved them efficiently using the available collaborative features. Occasional lapses in tool usage suggest that user interface guidance and training could further support users in fully using the available features. The findings reinforce that digital collaboration tools can play an important role in DRs, providing an interactive workspace where design discussions can take place.

Additionally, while CVEs provide tools that help structure discussions, they do not eliminate the need for organizational measures to ensure that DRs run smoothly. Clear facilitation, well-defined objectives, and structured participation remain essential to conducting effective sessions. The availability of collaborative features can support communication, but their impact is ultimately shaped by how they are integrated into the DRs.

Acknowledgments. This research was funded by Swedish innovation agency Vinnova in the PLENUM project (Grant Number: 2022–01704).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Huet, G., Culley, S.J., McMahon, C.A., Fortin, C.: Making sense of engineering design review activities. AI EDAM. **21**, 243–266 (2007). https://doi.org/10.1017/S0890060407000261
2. Cash, P., Daalhuizen, J., Hekkert, P.: Evaluating the efficacy and effectiveness of design methods: a systematic review and assessment framework. Des. Stud. **88**, 101204 (2023). https://doi.org/10.1016/j.destud.2023.101204
3. Patil, H., Sirsikar, S., Gholap, N.: Product design and development: phases and approach. Int. J. Eng. Res. V6 (2017). https://doi.org/10.17577/IJERTV6IS070136
4. Adams, R.S., Cardella, M., Purzer, Ş: Analyzing design review conversations: connecting design knowing, being and coaching. Des. Stud. **45**, 1–8 (2016). https://doi.org/10.1016/j.destud.2016.03.001
5. Horvat, N., Martinec, T., Perišić, M.M., Škec, S.: Comparing design review outcomes in immersive and non-immersive collaborative virtual environments. Procedia CIRP. **109**, 173–178 (2022). https://doi.org/10.1016/j.procir.2022.05.232
6. Bochenek, G.M., Ragusa, J.M., Malone, L.C.: Integrating virtual 3-D display systems into product design reviews: some insights from empirical testing. Int. J. Technol. Manage. **21**, 340–352 (2001). https://doi.org/10.1504/IJTM.2001.002917
7. Burova, A., et al.: Distributed asymmetric virtual reality in industrial context: enhancing the collaboration of geographically dispersed teams in the pipeline of maintenance method development and technical documentation creation. Appl. Sci. **12**, 3728 (2022). https://doi.org/10.3390/app12083728
8. Rivera, F.G., Lamb, M., Högberg, D., Alenljung, B.: Friction situations in real-world remote design reviews when using CAD and videoconferencing tools. Empathic Comput. (2025). https://doi.org/10.70401/ec.2025.0001
9. Horvat, N., Kunnen, S., Štorga, M., Nagarajah, A., Škec, S.: Immersive virtual reality applications for design reviews: Systematic literature review and classification scheme for functionalities. Adv. Eng. Inform. **54**, 101760 (2022). https://doi.org/10.1016/j.aei.2022.101760
10. Freeman, I.J., Salmon, J.L., Coburn, J.Q.: CAD integration in virtual reality design reviews for improved engineering model interaction. In: Presented at the ASME 2016 International Mechanical Engineering Congress and Exposition February 8 (2017). https://doi.org/10.1115/IMECE2016-66948
11. Bordegoni, M., Caruso, G.: Mixed reality distributed platform for collaborative design review of automotive interiors. Virtual Phys. Prototyping. (2012)
12. Reski, N., Alissandrakis, A., Kerren, A.: An empirical evaluation of asymmetric synchronous collaboration combining immersive and non-immersive interfaces within the context of immersive analytics. Front. Virtual Real. **2**, 743445 (2022). https://doi.org/10.3389/frvir.2021.743445
13. Thoravi Kumaravel, B., Nguyen, C., Diverdi, S., Hartmann, B.: TransceiVR: Bridging asymmetrical communication between VR users and external collaborators. In: UIST 2020 - Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, pp. 182–195. Association for Computing Machinery, Inc (2020). https://doi.org/10.1145/3379337.3415827
14. 1Rostami, A., McMillan, D.: The normal natural troubles of virtual reality in mixed-reality performances. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing

Systems, pp. 1–22. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3491102.3502139

15. Rostami, A., Karlgren, K., McMillan, D.: Kintsugi VR: Designing with fractured objects. In: Proceedings of the 2022 ACM International Conference on Interactive Media Experiences, pp. 95–108. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3505284.3529966
16. Gugenheimer, J., Stemasov, E., Frommel, J., Rukzio, E.: ShareVR: Enabling Co-located experiences for virtual reality between HMD and Non-HMD users. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 4021–4033. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3025453.3025683

A Study of Comparison Between Real and Virtual Environment of Operation Experience of PVD Coating Machine

Yinghsiu Huang(✉)

Department of Industrial Design, National Kaohsiung Normal University, Kaohsiung City, Taiwan
yinghsiu@mail.nknu.edu.tw

Abstract. Physical Vapor Deposition (PVD) technology plays a crucial role in precision industries. Originating in the mid-20th century, PVD was first used for semiconductor and optical coatings. With rising demand, its applications expanded to surface treatments for mechanical parts, enhancing wear resistance, hardness, and corrosion resistance. Market research projects the global PVD market to grow at a 7.3% annual rate, reaching 4.2 billion USD by 2030. Coating machine operation requires expertise and precision, involving complex procedures like interface operation, rotating cage assembly, target replacement, and routine maintenance. Novices face increased risks of equipment damage due to inexperience. Traditional coating machine training relies on paper manuals and video instruction, which is a higher learning barrier for beginners and involves higher operational risks. VR technology offers a novel training approach, allowing operators to experience the coating machine operation in an immersive environment, familiarizing themselves with operational steps and the internal structure of the equipment. Based on this, the research question focuses on how VR technology can be utilized for PVD coating machine operation training and how to present the SOP for first-time operations in VR to enhance training efficiency and operational safety.

Keywords: Virtual Reality Training · Virtual Reality Experience · Industrial Equipment Simulation · PVD Technology · Immersive Learning

1 Introduction

The rapid advancement of Virtual Reality (VR) technology has led to its widespread adoption in industrial training. VR enables the simulation of complex and precise industrial equipment operation environments, allowing users to learn in a safe and controlled setting. This not only reduces the risks and costs associated with training on physical equipment but also provides an immersive learning experience, enabling operators to grasp essential machine operation skills more efficiently.

Many industries have already integrated VR into their training programs. For example, PIXO VR in the United States has developed VR simulations for low-voltage rescue operations and high-altitude safety training, demonstrating VR's value in high-risk work

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 113–129, 2025.
https://doi.org/10.1007/978-3-031-93715-6_8

environments [1]. Similarly, TSUKAT Studio, in collaboration with Hauni, a subsidiary of the Körber Group, created a multi-user industrial machinery training simulation for tobacco processing equipment, allowing trainees from different countries to participate in the same training sessions remotely [2].

Physical Vapor Deposition (PVD) technology plays a crucial role in precision industries. Originally developed in the mid-20th century, PVD was initially used for semiconductor and optical coatings. As demand grew, its applications expanded to surface treatments for mechanical components, improving their wear resistance, hardness, and corrosion resistance [3]. According to market research, the global PVD market is expected to grow at an annual rate of 7.3%, reaching USD 4.2 billion by 2030 [4]. The operation of PVD coating machines requires specialized expertise and precision, involving complex procedures such as interface operation, rotating cage assembly, target material replacement, and routine maintenance. Inexperienced operators face an increased risk of equipment damage, which poses challenges to training efficiency and workplace safety.

Traditional training for PVD coating machines relies heavily on printed manuals and instructional videos, which present a steep learning curve for beginners and increase operational risks. By contrast, VR technology introduces a novel training approach, enabling operators to experience the coating machine in a fully immersive environment. This allows them to familiarize themselves with operational procedures and internal machine structures in a risk-free setting. Based on these observations, this study aims to address the following research questions:1. How can VR technology be effectively utilized for PVD coating machine training? 2. How can VR-based training systems present Standard Operating Procedures (SOPs) for first-time operations to enhance training efficiency and operational safety?

The research objectives are as follows: 1. Develop a VR coating machine system using Unreal Engine and Meta Quest 3. 2.Enable professional coating machine operators to experience VR and assess aspects such as immersion and smoothness. 3.Provide concrete suggestions for the design and improvement of the VR coating machine system. This research collaborates with SURFTECH Technology Co., Ltd., a Taiwanese coating machine manufacturer. Using machine design schematics, operational workflows, and expert insights provided by SURFTECH, the VR coating machine system is developed with Unreal Engine, simulating real operations and offering three main features: (1) Machine Viewing: In VR, users can explore the coating machine's exterior and interior, open chamber doors, hide the casing, and view the framework in detail. (2) Coating Process Simulation: Users operate the machine in VR, following prompts for cage assembly, target replacement, and starting/stopping the coating process. (3) SOP Simulation: Users practice startup and shutdown procedures, following SOPs to master the equipment workflow through simulated training.

2 Related Works

This study explores three key areas of related research: (1) an overview of PVD coating machines in Taiwan and globally, (2) the development of virtual reality, and (3) the application of VR in large-scale industrial machinery.

2.1 PVD Coating Machines

Surftech Technology Co., Ltd., founded in 1993, has been dedicated to the research and development of PVD hard coating technology. The company continuously innovates and aligns with the latest international advancements. Globally, several major manufacturers specialize in PVD hard coating technology, including Platit in Switzerland (Fig. 1), Swiss-PVD in Switzerland, and Oerlikon Balzers in Finland. These companies are recognized as leaders in the PVD coating industry. Surftech Technology competes internationally by consistently introducing advanced PVD coating processes, providing optimized and industry-specific coating solutions for cutting tools, molds, mechanical components, and precision parts.

This research project aims to integrate virtual reality and interactive interfaces to develop a VR-based simulation of the STAR PLUS PVD coating machine, visualizing both its exterior and internal structure through a head-mounted VR display. By leveraging VR, Surftech Technology can showcase its equipment at international trade exhibitions without the need for physical transportation. This VR solution not only enables realistic equipment simulations but also presents the machine's operational workflow. This approach aligns with government policies promoting energy efficiency and carbon reduction by minimizing logistics-related costs and emissions. Additionally, it enhances global marketing efforts, allowing a worldwide audience to experience Taiwan's latest advancements in PVD coating technology. Through industry-academic collaboration, this project aims to strengthen Taiwan's competitiveness in research and manufacturing on an international scale.

Fig. 1. Platit from Switzerland.

2.2 Development of Virtual Reality

Virtual Reality (VR) is a technology that simulates real-world environments by generating a virtual space distinct from external reality. By leveraging three-dimensional (3D) technology, head-mounted displays (HMDs), and motion-sensing devices, VR provides users with immersive sensory experiences, including visual, auditory, and even tactile feedback. According to Burdea and Coiffet (2003) [5], VR is characterized by three fundamental attributes: immersion, interactivity, and imagination.

The concept of VR emerged in the early 1900s. Rather than being an entirely new field, VR is an interdisciplinary technology that redefines how people interact with digital environments [6]. In 1962, Morton Heilig [7] invented a device called Sensorama (Fig. 2, left), which simulated sensory experiences through visuals, sounds, smells, and vibrations. Around the same time, Ivan Sutherland [8] developed the first head-mounted display, laying the foundation for VR technology. However, due to technological limitations at the time, these early VR systems failed to deliver fully immersive experiences, resulting in limited academic and commercial interest.

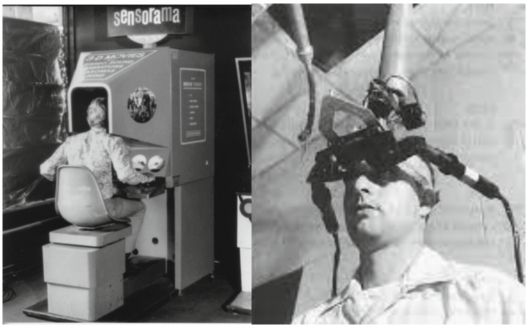


Fig. 2. Left: first VR "Sensorama"; Right: Ivan Sutherland developed the first head-mounted display.

From the late 1990s onward, the rapid advancement of computer technology significantly contributed to the growth of VR, particularly in gaming and entertainment. Notable developments include Sega's Mega Drive 32X and Nintendo's Virtual Boy. In the early 2000s, improvements in computing power and graphical processing led to the commercialization of VR-related products, such as Oculus VR and HTC Vive, which introduced head-mounted displays and hand controllers.

Over the past few decades, VR has expanded beyond entertainment and now plays a crucial role in education, science, and medicine [9]. Additionally, related technologies such as Augmented Reality (AR), Mixed Reality (MR), and Spatial Computing have become increasingly integrated, offering more immersive and realistic experiences. These advancements have led to several emerging trends and possibilities (Fig. 3):

1. More Natural and Intuitive Interaction Methods

   Current VR systems primarily rely on hand controllers for interaction. Future developments will focus on more natural and intuitive interfaces, such as gesture recognition, eye tracking, and voice commands, allowing for more fluid and immersive user experiences.
2. Integration with AI and IoT

Fig. 3. Meta Oculus VR.

VR is expected to integrate with Artificial Intelligence (AI) and the Internet of Things (IoT), enabling more personalized and adaptive experiences. For instance, AI-powered virtual tutors and assistants could enhance VR learning environments.

3. Enhanced Social and Collaborative Features

    The future of VR will emphasize social interactions and collaboration, enabling users to connect and share experiences in virtual spaces. This could drive innovations in virtual meetings, online social platforms, and remote teamwork, breaking geographical limitations.

4. Cloud-Based VR Solutions

    With the rise of 5G networks and cloud computing, cloud-based VR applications will become more prevalent. This will allow users to access high-quality VR experiences without the need for expensive hardware, improving accessibility and real-time data sharing.

5. Personalized and Customizable VR Experiences

    As VR technology evolves, users will have more control over their virtual experiences, customizing settings to match personal preferences. This will lead to more tailored and engaging VR content.

Between 2019 and 2023, Oculus VR (Facebook Reality Labs) has consistently advanced VR technology. According to International Data Corporation (IDC), Oculus held nearly 80% of the market share in 2022. In 2019, Oculus introduced Oculus Link, which enabled users to connect their Oculus Quest headset to a PC via USB for enhanced performance. In 2021, the company further expanded its ecosystem with Oculus Air Link, a wireless connectivity system that allows users to stream high-quality VR content over Wi-Fi.

2.3 Application of Virtual Reality in Large-Scale Machinery

Numerous international companies have adopted VR technology to simulate large-scale industrial machinery operations. These simulations enhance training efficiency and improve workplace safety by allowing trainees to practice in realistic, risk-free environments.

One notable example is PIXO VR, a U.S.-based company specializing in customized VR training solutions. PIXO VR provides hardware-to-software integrated training systems for high-risk occupational training scenarios, such as low-voltage rescue operations, forklift operation training, and high-altitude safety simulations. Their extensive experience in developing training programs for international enterprises underscores VR's growing role in industrial training and workplace safety (Fig. 4).

Fig. 4. PIXO.

Another key player, TSUKAT, an interactive design studio, has developed VR/AR-based training simulations for industrial machinery. In collaboration with HAUNI, a leading provider of tobacco processing technology and services, TSUKAT created a multi-user VR simulation for training workers on complex machinery. A unique feature of this project is its support for collaborative training, allowing multiple trainees and instructors to engage in VR-based training sessions simultaneously, even from different countries. The system also includes an observer application, enabling trainers to monitor and record entire training sessions for assessment and review (Fig. 5) [2].

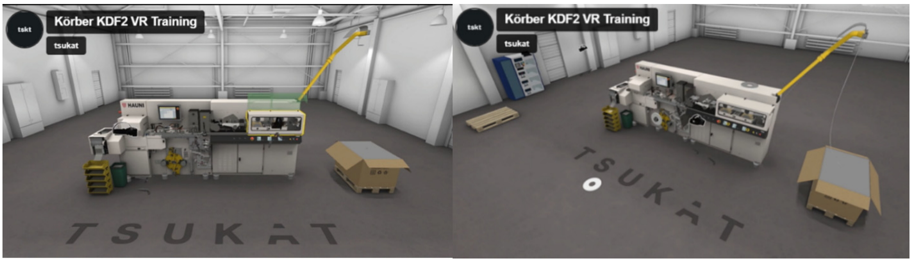


Fig. 5. TSUKAT and HAUNI VR/AR-based training simulations.

Workplace accidents remain a critical issue in industrial sectors. In Taiwan, the construction industry accounts for 45% of major occupational accidents, followed by the manufacturing sector at 25%. To improve workplace safety, traditional training programs rely heavily on written materials and classroom-based hazard awareness instruction, which often lack immersive engagement and fail to create lasting learning experiences. VR-based training solutions offer a more effective alternative, providing immersive hazard recognition and safety training through four key functions:

1. Workplace Experience Simulation – Helps trainees understand the working environment.
2. Accident Experience Simulation – Enables users to experience potential hazards (e.g., falls, electric shocks, fires).
3. Hazard Recognition and Emergency Response Training – Guides users in identifying workplace hazards and implementing proper responses.
4. Skill-Based Training – Provides practical, hands-on training for operating high-risk machinery.

Research has demonstrated that VR-based safety training is more effective than traditional methods for novice operators of heavy machinery. A study comparing VR and PC-based safety training for tower crane operators found that VR provided greater immersion, realism, and depth perception, leading to improved hazard recognition accuracy—particularly in identifying critical dangers such as cable entanglement. By allowing novice operators to train in a controlled virtual environment, experienced instructors can guide them through hazard identification, significantly reducing the risk of accidents on construction sites (Fig. 6) [10].

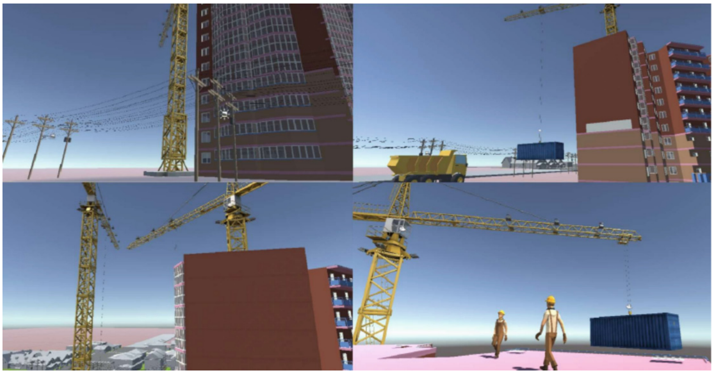


Fig. 6. Safety training for tower crane operators in VR.

VR technology has also been applied to industrial maintenance training. Simweb Technology Co., Ltd., for instance, has developed VR-based factory automation and electrical engineering training systems. In 2017, the company introduced a VR production monitoring module, enabling trainees to learn how to oversee production operations and control manufacturing processes within a VR environment. Additionally, the system includes a maintenance module, allowing trainees to practice equipment repair and system modifications through realistic VR simulations (Fig. 7).

2.4 Summary

VR technology has matured significantly in both software and hardware, with widespread applications in gaming, spatial simulations, industrial machinery training, and maintenance procedures. Among these applications, full-scale spatial simulation of industrial equipment has proven particularly effective.

Fig. 7. Practice equipment repair and system modifications through VR simulations.

This industry-academic collaboration aims to enhance the visibility of domestically manufactured equipment in the global market. Surftech Technology's products have already been adopted in the United States, Canada, Vietnam, China, and Myanmar. Feedback from these international clients demonstrates that Surftech's PVD coating equipment meets high technical standards, competing with leading manufacturers in Europe and the United States.

The STAR PLUS vacuum coating system developed in this project operates entirely within a vacuum chamber, unlike traditional electroplating methods that require large quantities of water as a reaction medium. Vacuum coating consumes only minimal water for cooling and does not produce wastewater emissions, making it a sustainable and environmentally friendly technology. Additionally, the newly developed magnetron sputtering target system is more energy-efficient than traditional targets, generating higher metal ion output at the same power consumption, thereby supporting energy conservation and carbon reduction efforts.

Another key advantage of VR is reducing exhibition costs. Transporting a physical coating machine to international trade shows is both expensive and time-consuming. By utilizing VR to showcase the STAR PLUS coating machine, potential buyers can explore its structure and operation process in a virtual environment. Furthermore, VR allows the demonstration of various operational modes, such as: Opening the coating chamber, Changing the target material, Performing target replacement simulations, Practicing maintenance procedures…etc. By replacing traditional on-site demonstrations with VR-based training and presentations, this project aims to significantly reduce logistics expenses while enhancing customer engagement and understanding of the technology.

3 Research Methodology and Procedure

To achieve the objectives of this study, the research is divided into two major components: Development of the VR-based PVD Coating Machine Simulation System, and Implementation of the Media Engagement Questionnaire (MEQ).

3.1 Development of the VR-Based PVD Coating Machine Simulation System

This study was conducted in collaboration with Surftech Technology Co., Ltd., a Taiwanese manufacturer specializing in PVD coating machines. Based on the company's technical drawings, operational workflows, and expert recommendations, we developed a VR-based PVD coating machine training system using Unreal Engine. This system simulates the real-world operation of a PVD coating machine in a virtual environment.

The VR system interface features a circular navigation menu (Fig. 8) that allows users to select from three primary functions: 1. Machine Visualization Mode – Users can inspect the exterior and internal structure of the PVD coating machine in a VR environment. The system allows users to open machine doors, remove protective covers, and reveal internal components for closer examination. 2. Operational Process Simulation – Users are guided step by step through key operational procedures, including rotating cage assembly, target material replacement, and alarm troubleshooting. 3. Standard Operating Procedures (SOP) Training – Users must follow a structured SOP for machine startup and shutdown, ensuring they learn the proper procedures before operating a real machine.

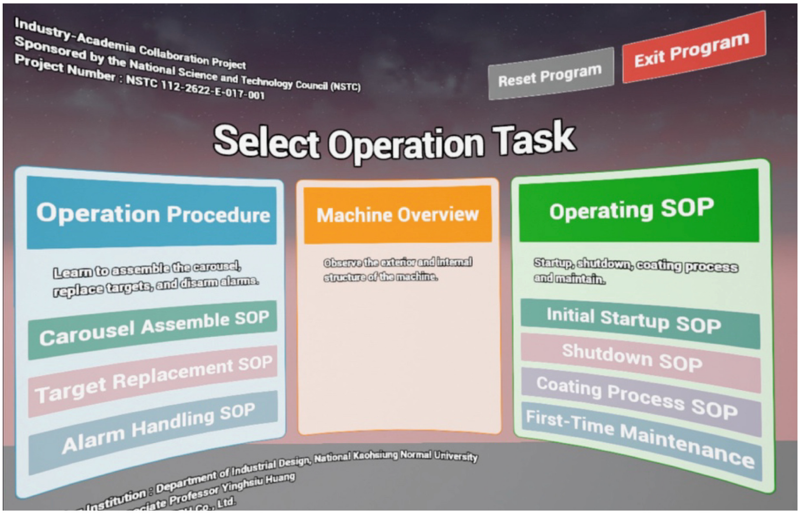


Fig. 8. A circular navigation menu in VR.

Machine Visualization Mode. In this mode, users can explore the exterior of the coating machine within a VR environment. By walking around the machine and interacting with various components, users can open hatches and doors to examine the internal structure. The system also includes auditory feedback, enhancing the sense of realism. Additionally, a side panel allows users to toggle the visibility of the outer shell and internal framework, enabling a detailed exploration of the machine's design (Fig. 9).

Operational Process Simulation. This module includes step-by-step simulations of three key operational procedures: Rotating Cage Assembly, Target Material Replacement, and Alarm Troubleshooting.

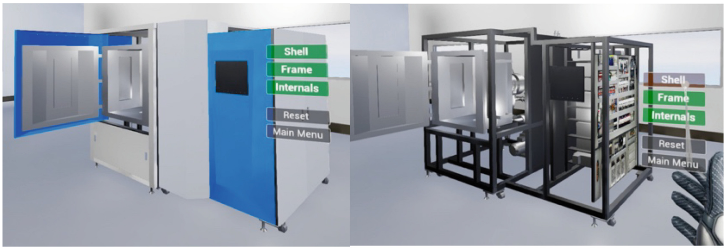


Fig. 9. Machine Visualization Mode.

Rotating Cage Assembly Simulation. In this simulation (Fig. 10), an orange transport cart holds the core circular platform of the rotating cage. A semi-transparent blue guidance model helps users correctly position each component. Users must follow a predefined sequence to assemble the cage properly. Each correctly placed component is automatically locked into place, ensuring users learn the correct assembly process for holding workpieces during PVD coating.

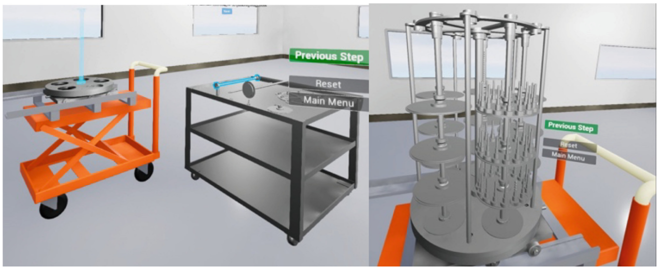


Fig. 10. Rotating Cage Assembly Simulation

Target Material Replacement Simulation. Replacing the PVD target material involves handling several key components, including: the target itself, two flexible rubber rings, and a thick sealing ring. During the VR simulation, users are guided through the proper cleaning and assembly steps. After completing the preparation, users must install the target onto the vacuum chamber, following real-world procedures, including:1. Securing the target in place; 2. Connecting the gas pipeline; 3. Attaching the power supply and ion source (Fig. 11).

Alarm Troubleshooting Simulation. In this module, users respond to error messages displayed on the machine's interface. If an issue arises, the system highlights the corresponding warning light, prompting the user to investigate the cause of the malfunction. The VR simulation provides troubleshooting guidance, allowing users to correct errors step by step. Once the issue is resolved, the warning light turns green, indicating the machine is operational again (Fig. 12). Users can also review historical error logs, helping them diagnose recurring machine issues.

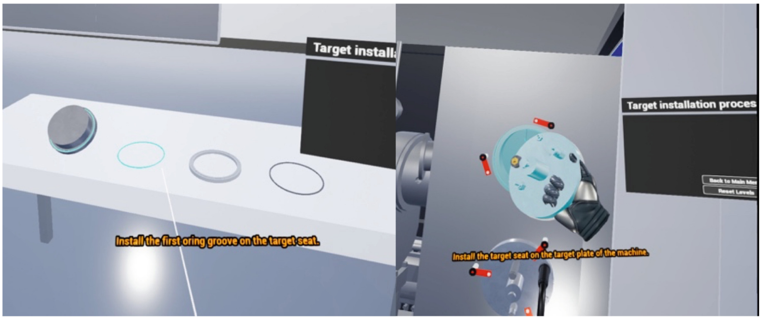


Fig. 11. Target Material Replacement Simulation.

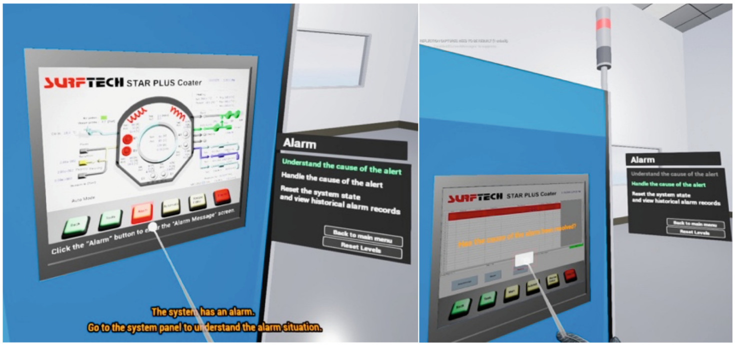


Fig. 12. Alarm Troubleshooting Simulation.

Standard Operating Procedures (SOP) Training. The SOP simulation focuses on the initial startup and shutdown procedures of the PVD coating machine. When customers purchase Surftech's coating machines—which cost over $400,000 USD per unit—the first startup is a critical moment. Errors in operation could result in severe damage to the machine. Traditionally, Surftech's technicians provide on-site training or instructional video tutorials to guide new users through this process. To enhance training effectiveness and reduce risks, this study developed a VR-based SOP training system, allowing users to practice machine startup and shutdown procedures in a controlled environment.

Machine Start-up SOP. For first-time users, operating a PVD coating machine can be intimidating. In traditional training, Surftech's technical staff provides on-site instruction, or users must rely on video tutorials. The VR startup SOP simulation simplifies this process by offering: Step-by-step visual guidance through an on-screen interface; Ground-level navigation arrows and flashing indicators to highlight important machine components (Fig. 13); Sequential activation instructions, leading users through power-up sequences and system checks. By following this structured VR-based startup procedure, users gain confidence and familiarity with the machine before operating real equipment.

Fig. 13. Machine Start-up SOP.

Machine Shutdown SOP.

Fig. 14. Machine Shutdown SOP

The shutdown procedure is the reverse of the startup process, but it follows the same structured guidance system. Users are led through the necessary steps to power down the machine safely, ensuring that all systems are properly deactivated (Fig. 14). The VR-based shutdown simulation reduces the likelihood of user errors, protecting the equipment from damage.

3.2 Media Engagement Questionnaire (MEQ) Design

This study employs the Media Engagement Questionnaire (MEQ). The system was tested using Meta's Oculus Quest 3, and participants included employees of Surftech Technology, who are highly familiar with real-world PVD coating machines. Their feedback provides valuable insights into the differences between VR training and hands-on machine operation.

The MEQ is based on the Game Engagement Questionnaire (GEQ) developed by Brockmyer et al. [11]. The original GEQ consists of 19 items, each rated on a 7-point Likert scale (Table 1). This study modifies the GEQ into a VR-specific MEQ, focusing on five key dimensions (Table 2):

Table 1. Game Engagement Questionnaire (GEQ).

No.	Question Items
1	I lost track of time
2	Things seemed to happen naturally
3	I felt different from my usual self
4	I felt scared
5	The game felt realistic
6	If someone called me while playing, I wouldn't hear them
7	I felt very excited
8	Time seemed to stop while I was playing
9	I was completely focused on the game
10	When someone talked to me, I couldn't respond
11	I couldn't tell if I was tired
12	Playing the game became a habit
13	My thoughts were racing
14	While playing, I lost awareness of my surroundings
15	I didn't need to think much about how to play the game
16	Playing the game made me feel calm
17	I played longer than I had expected
18	I was deeply immersed in the game
19	I felt like I couldn't stop playing

Additionally, in the GEQ (Game Engagement Questionnaire) proposed by Poels et al. [12], there are seven different dimensions: immersion, flow, competence, tension, challenge, positive affect, and negative affect. Based on the attributes of these questions, this study adjusted them into five dimensions in Table 2.

4 MEQ Data Processing and Analysis

In this study, employees of Surftech Technology Co., Ltd. Were invited to experience the VR-based PVD coating machine simulation. Immediately after the experience, they completed the MEQ (Media Engagement Questionnaire) as described in Sect. 3.2. Since Surftech's employees are familiar with the physical coating machine's structure, operational workflow, and SOP, their feedback provides valuable insights into how the VR training system compares to real-world machine operation.

A total of 37 employees participated, including 30 males and 7 females. The analysis consists of two parts: Reliability Analysis – to assess the internal consistency of the

Table 2. 5 dimensions GEQ (Game Engagement Questionnaire) by this study.

Category	Question Items
a. immersion	a1. I was fully engaged in the VR experience a2. Clearly understood the SOP within VR a3. I could easily follow the VR training steps
b. flow	b1. I enjoyed the VR training process b2. The VR experience made me lose track of time b3. I was highly focused on the VR training
c. competence	c1. After the VR training, I could operate the PVD machine c2. I felt I had learned the machine operation steps c3. I could freely explore the VR training system
d. positive effect	d1. I found the VR experience engaging d2. The VR training felt realistic d3. Learning to operate the PVD machine in VR was a unique experience
e. negative affect	e1. The VR experience negatively affected my mood e2. I found the VR training boring e3. The VR experience lacked realism

questionnaire, and Independent Sample t-Test – to examine whether there are statistically significant differences among the questionnaire items.

4.1 Reliability Analysis

In the MEQ reliability analysis, Cronbach's alpha was used to measure the internal consistency of the questionnaire. According to Wu (1984), a Cronbach's alpha between 0.50 and 0.70 is considered an acceptable reliability standard, while a range of 0.70 to 0.90 indicates high reliability, which is the standard for academic research. DeVellis (1991) also suggested that 0.70 is the minimum acceptable reliability threshold.

As shown in Table 3, the overall reliability of the MEQ in this study reached 0.851, indicating a high level of reliability. Specifically, the subcategories immersion (0.890), competence (0.894), and positive affect (0.756) all demonstrated strong reliability, meeting the standard for academic research. However, flow (0.511) and negative affect (0.514) only achieved the basic reliability threshold (Table 4).

4.2 One-Simple T-Test

A one-sample t-test was conducted using a test value of 4, with $p < 0.05$ indicating statistical significance. Among the 15 questionnaire items, 8 items showed high statistical significance ($p < 0.001$, marked in gray in Table 4). The items with negative t-values suggest a strongly negative response, while positive t-values indicate a strongly positive experience. The most significant results (in descending order of absolute t-value) were:

e1: "The VR experience negatively affected my mood" ($t = -9.234$, $p < 0.001$)

e2: "I found the VR training boring" ($t = -9.139$, $p < 0.001$)

Table 3. VR MEQ reliability analysis.

Category	Question Items	Cronbach's α
a. immersion	a1. I was fully engaged in the VR experience a2. Clearly understood the SOP within VR a3. I could easily follow the VR training steps	.890
b. flow	b1. I enjoyed the VR training process b2. The VR experience made me lose track of time b3. I was highly focused on the VR training	.511
c. competence	c1. After the VR training, I could operate the PVD machine c2. I felt I had learned the machine operation steps c3. I could freely explore the VR training system	.894
d. positive effect	d1. I found the VR experience engaging d2. The VR training felt realistic d3. Learning to operate the PVD machine in VR was a unique experience	.756
e. negative affect	e1. The VR experience negatively affected my mood e2. I found the VR training boring e3. The VR experience lacked realism	.514
	Overall Cronbach's Alpha	.851

Table 4. VR One-simple t-test.

Category	Question Items	t	Two-tailed
a. immersion	a1. I was fully engaged in the VR experience a2. Clearly understood the SOP within VR a3. I could easily follow the VR training steps	2.144 3.571* 3.181*	.039 .001 .003
b. flow	b1. I enjoyed the VR training process b2. The VR experience made me lose track of time b3. I was highly focused on the VR training	3.045* 6.827** 4.286**	.004 <.001 <.001
c. competence	c1. After the VR training, I could operate the PVD machine c2. I felt I had learned the machine operation steps c3. I could freely explore the VR training system	3.045* 4.623** 3.129*	.004 <.001 .003
d. positive effect	d1. I found the VR experience engaging d2. The VR training felt realistic d3. Learning to operate the PVD machine through VR was a unique experience	4.879** –3.111* 7.488**	<.001 .004 <.001
e. negative affect	e1. The VR experience negatively affected my mood e2. I found the VR training boring e3. The VR experience lacked realism	–9.234** –9.139** 4.537**	<.001 <.001 <.001

* * p <0.001; * p <0.05

d3: "Learning to operate the PVD machine through VR was a unique experience" ($t = 7.488$, $p < 0.001$)

b2: "The VR experience made me lose track of time" ($t = 6.827$, $p < 0.001$)

d1: "I found the VR experience engaging" ($t = 4.879$, $p < 0.001$)
c2: "I felt I had learned the machine operation steps" ($t = 4.623$, $p < 0.001$)
e3: "The VR experience lacked realism" ($t = 4.537$, $p < 0.001$)
b3: "I was highly focused on the VR training" ($t = 4.286$, $p < 0.001$)

These findings indicate that Surftech employees generally had a positive experience with the VR coating machine simulation. They did not feel bored or negatively affected, and they found the experience engaging, immersive, and effective for learning.

Interestingly, e3 ("The VR experience lacked realism") also reached $p < 0.001$, suggesting that Surftech employees, who have real-world experience with the coating machine, found the VR simulation less realistic—which is an expected outcome given their familiarity with the actual machine.

Additionally, 6 items reached $p < 0.05$ significance:
a2: "I clearly understood the SOP within VR" ($t = 3.571$, $p = 0.001$)
a3: "I could easily follow the VR training steps" ($t = 3.181$, $p = 0.003$)
c3: "I could freely explore the VR training system" ($t = 3.129$, $p = 0.003$)
d2: "The VR training felt realistic" ($t = -3.111$, $p = 0.004$)
b1: "I enjoyed the VR experience" ($t = 3.045$, $p = 0.004$)
c1: "After the VR training, I could operate the PVD machine" ($t = 3.045$, $p = 0.004$)

Only a1 ("I was fully engaged in the VR experience") did not reach statistical significance ($t = 2.144$, $p = 0.039$). This suggests that, while users were generally engaged, they did not feel fully immersed at a significant level.

5 Conclusions and Recommendations

Following the VR-based PVD coating machine training, employees from Surftech Technology Co., Ltd. Completed the MEQ questionnaire, providing direct feedback on their VR experience. The analysis using a one-sample t-test revealed that participants generally had a positive perception of the VR simulation, with significant results at $p < 0.001$ for overall engagement and at $p < 0.05$ for SOP-related learning effectiveness. These findings indicate that the VR training platform developed in this study provides a meaningful and effective learning experience.

Notably, since all participants were Surftech employees with real-world experience operating PVD coating machines, their feedback suggests that while the VR training was engaging and useful, it lacked a certain level of realism compared to actual machine operations. This result was expected, as experienced operators are more attuned to the physical nuances and tactile feedback of real equipment.

To further evaluate the effectiveness of VR-based training, the next phase of this research will focus on users with no prior experience operating PVD coating machines. Conducting the same VR simulation with novice users will allow for a comparison with Surftech employees, providing insight into whether those without hands-on experience perceive the VR environment as more realistic due to their lack of familiarity with actual machine operations.

Discussions with Surftech Technology also highlighted key aspects for improving the VR simulation, particularly in the areas of alarm handling and component debugging. Future enhancements will focus on refining the VR-based troubleshooting experience, enabling users to respond to system alerts and error diagnostics in a more interactive and realistic manner. In addition, incorporating training for fault detection and component maintenance will further strengthen the simulation's ability to prepare users for real-world repair and maintenance tasks.

Acknowledgments. This study was supported by the National Science and Technology Council (NSTC) under the Industry-Academia Collaboration Project NSTC 112–2622-E-017–001. The research team extends its gratitude to the graduate students Tzu-Hsuan Chen, Meng-Shiuan Tsai, and Hsin-Chiao Chan from the Graduate Institute of Indus-trial Design at National Kaohsiung Normal University for their contributions to this project.

References

1. PIXO VR: Virtual reality training (2021). https://pixovr.com/virtual-reality-training/
2. Tsukat: Körber KDF2 VR training: Tsukat portfolio (2024). https://tsukat.bestclients.com.ua/portfolio/korber-kdf2-vr-training/
3. Baptista, A., Silva, F., Porteiro, J., Míguez, J., Pinto, G.: Sputtering physical vapour deposition (PVD) coatings: a critical review on process improvement and market trend demands. Coatings, **8**(11), 402 (2018). https://doi.org/10.3390/coatings8110402
4. Lucintel. Physical Vapor Deposition (PVD) Market Report: Trends, Forecast and Competitive Analysis 2024–2030. Global Information, Inc. (GII) (2023). https://www.giiresearch.com/report/luci1387481-physical-vapor-deposition-market-report-trends.html
5. Burdea, G., Coiffet, P.: Virtual Reality Technology. John Wiley &Sons Inc (2003)
6. Machover, C., Tice, S.: Virtual reality. Art. IEEE Computer Graphics and Applications (1994)
7. Heilig, M.: The Sensorama: One of the First Functioning Efforts in Virtual Reality (1962). https://www.historyofinformation.com/detail.php?id=2785
8. Ivan Sutherland, I.: First Virtual Reality Head Mounted Display System (1968). https://www.historyofinformation.com/detail.php?id=861
9. de Strulle, A.: Differentiation of the Causal Characteristics and Influences of Virtual Reality and the Effects on Learning at a Science Exhibit Dissertations (2004)
10. Shringi, A., et al.: Efficiency of VR-Based Safety Training for Construction Equipment: Hazard Recognition in Heavy Machinery Operations (2022)
11. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N.: The development of the game engagement questionnaire: a measure of engagement in video game-playing. J. Exper. Soc. Psychol. **45**(4), 624–634 (2009). https://doi.org/10.1016/j.jesp.2009.02.016
12. Poels, K., De Kort, Y., Ijsselsteijn, W.: It is always a lot of fun! Exploring dimensions of digital game experience using focus group methodology. In: Proceedings of the 2007 Conference on Future Play, Futureplay 2007, Controlling order-effect bias. Toronto, ON, Canada (2007)

Integrating Virtual and Augmented Reality Into Public Education: Opportunities and Challenges in Language Learning

Tanja Kojić[1(✉)], Maurizio Vergari[1], Giulia-Marielena Benta[1], Joy Krupinski[1], Maximilian Warsinke[1], Sebastian Möller[1,3], and Jan-Niklas Voigt-Antons[2]

[1] Quality and Usability Lab, TU Berlin, Berlin, Germany
tanja.kojic@tu-berlin.de

[2] Immersive Reality Lab, Hamm-Lippstadt University of Applied Sciences, Lippstadt, Germany

[3] German Research Center for Artificial Intelligence (DFKI), Berlin, Germany

Abstract. Virtual Reality (VR) and Augmented Reality (AR) are emerging as transformative tools in education, offering new possibilities for engagement and immersion. This paper explores their potential in language learning within public education, focusing on their ability to enhance traditional schooling methods and address existing educational gaps. The integration of VR and AR in schools, however, is not without challenges, including usability, technical barriers, and the alignment of these technologies with existing curricula. Drawing on two empirical studies, this work investigates the opportunities and challenges of VR and AR-assisted language learning, proposing strategies for their effective implementation in the public sector. Based on two empirical studies, findings show that VR boosts motivation and immersion but has an unclear impact on vocabulary retention, with technical limitations and cognitive overload as challenges. AR enhances contextual learning and accessibility but faces usability constraints and limited personalization. To facilitate effective adoption, this paper recommends improving interface design, reducing cognitive load, increasing adaptability, and ensuring adequate infrastructure and teacher training. Overcoming these barriers will enable a more effective integration of immersive technologies in language education.

Keywords: Virtual Reality · Augmented Reality · Language Learning · Public Education · Immersive Technologies · Usability · Educational Technology

1 Introduction

Virtual Reality (VR) and Augmented Reality (AR) are becoming popular tools for learning languages. These technologies create interactive and immersive experiences, making learning more engaging. Traditional language learning often involves memorization and passive study, but VR and AR allow learners to

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 130–144, 2025.
https://doi.org/10.1007/978-3-031-93715-6_9

interact with realistic situations. These experiences can increase motivation and improve learning [4].

VR places learners in fully immersive environments where they can practice language skills in lifelike scenarios. Research shows that VR makes learning more exciting and engaging. However, its impact on language improvement, especially vocabulary learning, is not clear [9]. Some studies suggest that while learners enjoy VR, the complexity of using it can make learning harder. Also, the need for special devices limits its use in schools [10].

AR adds digital content to the real world, making learning more interactive. Unlike VR, AR works on regular mobile devices, making it easier to use. AR apps like Mondly AR help learners interact with digital elements in their surroundings, making learning more engaging and practical [2]. However, AR also has challenges, such as too much information at once, difficult interfaces, and lack of personalization. Making AR fit into school lessons while keeping students interested is still a challenge for developers and teachers [11].

Research shows that VR and AR both have advantages and limitations in language learning. VR helps learners feel present and involved, while AR makes learning easier and more relevant to real life [12]. However, personal factors such as past experience with VR, ability to learn languages, and the number of languages a person knows can affect how useful these tools are [8]. It is also important to design user-friendly interfaces and effective learning methods to make these technologies work well in education [5].

The aim of this paper is to analyze the benefits and challenges of VR and AR in language learning by synthesizing insights from empirical studies and usability assessments. Specifically, the paper explores the impact of VR and AR on learner engagement and comprehension, identifies usability challenges and technological limitations, and proposes strategies for improving the effectiveness of immersive language learning tools in educational settings.

2 Related Work

2.1 Augmented Reality in Language Learning

Augmented Reality has gained attention in language learning for its ability to merge virtual elements with real-world environments, fostering interactive and engaging experiences. Studies highlight AR's potential to improve comprehension and retention through multimodal interaction, particularly in vocabulary acquisition, pronunciation, and reading comprehension. Despite these advantages, its integration into educational frameworks remains limited, necessitating further research on long-term effectiveness and adaptability.

Research indicates that AR enhances learning outcomes by providing personalized instruction, interactive feedback, and gamification elements. The ability to manipulate virtual objects and engage in real-time interactions strengthens motivation and reinforces learning. Features such as adaptive content, which adjusts based on user proficiency, allow for a more tailored educational experience. Gamification, including rewards and interactive challenges, has also been shown to sustain learner engagement over time.

Although AR has been successfully implemented in vocabulary-building and pronunciation exercises, its application to higher-order language skills remains underexplored. Writing, grammar acquisition, and discourse-based learning still require significant development in AR-based methodologies.

Despite its potential, several barriers hinder the widespread adoption of AR in language education. Technical limitations, such as high development costs, hardware constraints, and software compatibility, impact accessibility. Additionally, the cognitive demands of AR environments, where multiple sensory inputs must be processed simultaneously, can lead to information overload, reducing learning efficiency. Another challenge lies in usability, where poorly designed interfaces and unclear navigation reduce the effectiveness of AR-based learning tools. While AR demonstrates strong potential in foundational language skills, its capacity to support advanced linguistic competencies, including writing and critical thinking, remains an area requiring further investigation.

Research suggests that usability plays a crucial role in determining the success of AR-based language learning. Factors such as intuitive navigation, interactive storytelling, and real-time feedback mechanisms contribute significantly to engagement and learning retention. Adaptive learning environments, where content adjusts based on user progress, have been shown to enhance user satisfaction and improve language acquisition [13]. Beyond usability, learner background and educational setting influence AR's effectiveness [1]. Prior exposure to AR, language proficiency levels, and learning context-whether self-directed or classroom-based-impact how users engage with and benefit from AR applications. Addressing these factors through customized experiences can improve accessibility and maximize AR's educational potential.

To overcome existing challenges, researchers propose advancements in AI-driven personalization, enabling content to dynamically adapt to individual learning styles and progress [7]. Enhanced gamification techniques, incorporating interactive elements and real-time feedback, can further sustain motivation and engagement. The development of cloud-based and cross-device platforms is also recommended to improve accessibility, ensuring seamless integration across different learning environments. Moreover, structured teacher training programs and pedagogical support will be essential in bridging the gap between AR technology and effective educational implementation. While AR holds significant promise, future studies should explore its role in more complex linguistic tasks, refine usability strategies, and develop scalable learning models for broader adoption.

2.2 Virtual Reality in Language Learning

Virtual Reality has emerged as an immersive alternative for language learning, offering real-time simulated interactions and contextualized learning experiences. Studies suggest that VR can enhance listening, pronunciation, and conversational fluency, but its effectiveness in reading comprehension, vocabulary retention, and writing remains debated [4].

Language acquisition is shaped by various cognitive and environmental factors. Age plays a significant role, with younger learners relying more on intuitive absorption, whereas adults depend on analytical reasoning and structured learning [3]. Language aptitude, particularly associative memory and phonological short-term memory, has been identified as a key predictor of language learning success [14]. Additionally, immersion remains a crucial component, as extended exposure to a language environment strengthens neural pathways associated with language processing. Another important factor is cross-linguistic transfer, where prior knowledge of a second language facilitates the acquisition of a third language, demonstrating the cognitive benefits of multilingual learning [8].

A systematic review examined VR-assisted language learning (VRALL) research from 2015 to 2018, finding that VR enhances engagement and motivation but lacks extensive studies on long-term learning outcomes [10]. More recent work, which analyzed studies from 2018 to 2022, noted a significant increase in VR adoption, particularly following the COVID-19 pandemic. While VR was found to improve learner confidence and immersion, concerns were raised about content authenticity and the cognitive load associated with virtual environments [6]. Research indicates that VR is particularly effective in developing pronunciation and conversational fluency, whereas vocabulary acquisition, reading comprehension, and writing proficiency show more mixed results. The lack of standardized learning models in VR-based education further complicates its integration into formal curricula.

This study builds on previous research, which investigated the effectiveness of the Mondly VR language-learning application. Their findings indicated that while VR increased engagement and immersion, it did not result in significantly higher language competence compared to mobile learning [9]. Expanding on these insights, the present study explores additional variables such as device variability and the role of third language acquisition in VR-assisted learning. By incorporating cognitive and pedagogical factors, this research aims to deepen understanding of VR's educational potential.

To enhance the efficacy of VR in language education, researchers emphasize the need for improved content authenticity, ensuring culturally relevant and realistic interactions. The integration of AI-driven adaptive learning models could provide personalized instruction, dynamically adjusting content based on learner progress. Furthermore, cross-platform compatibility between VR, mobile, and desktop applications would facilitate flexible learning pathways. The alignment of VR technologies with structured curricula and teacher training programs is also crucial to maximizing its educational benefits. While VR has demonstrated clear advantages in fostering immersion and engagement, future research must focus on evaluating its effectiveness across a wider range of linguistic skills, refining interactive methodologies, and ensuring scalability for diverse educational contexts.

2.3 Comparison of AR and VR in Language Learning

Both AR and VR offer different benefits for language learning, each supporting different parts of the learning process. AR adds digital content to real-world settings, making it useful for vocabulary learning and interactive practice. Gamification and real-world context help learners stay engaged, especially beginners. VR, on the other hand, creates immersive environments that help with pronunciation, listening comprehension, and conversation skills. By simulating real-life situations, VR allows learners to practice language in a natural way.

Despite these advantages, both technologies have challenges in education. Technical issues, usability problems, and the need for more adaptable content make it difficult to integrate them into schools. To make AR and VR more effective, future research should focus on improving user-friendly designs, expanding learning models beyond basic skills, and studying their long-term impact on language learning. Overcoming these challenges will help ensure that AR and VR are not just engaging but also truly beneficial for language education.

3 Methodology

This section describes the research approach used to investigate the role of AR and VR in language learning. Two separate empirical studies were conducted to assess the impact of these immersive technologies on different aspects of language acquisition. The first study examined the usability and learning experience in an AR-based language learning application, while the second study evaluated how VR immersion influences pronunciation, listening comprehension, and conversational practice. A mixed-method approach was used in both studies, combining quantitative learning assessments with qualitative feedback analysis to provide a comprehensive understanding of their effectiveness.

3.1 AR Study: Usability and Learning Experience

The AR study focused on evaluating the role of Mondly AR, a widely used augmented reality language learning application. The research assessed usability, learner engagement, and cognitive load while exploring how interactive AR elements influence vocabulary acquisition.

Participants and Study Design. Participants were recruited through academic networks and online advertisements, targeting individuals with varying levels of language proficiency. The study aimed to include both novice and experienced users of AR technology to understand its accessibility across different learner backgrounds. A controlled experimental design ensured that all participants interacted with the AR application under standardized conditions. Each participant engaged in a structured learning session, allowing for a consistent evaluation of usability and learning outcomes.

Application. The Mondly AR application was selected for its ability to overlay virtual language-learning elements onto real-world environments. The app offers interactive vocabulary exercises, pronunciation feedback, and conversational practice using speech recognition. Learners engage with 3D-rendered objects that represent words in the target language, allowing for contextualized language acquisition. The app also incorporates gamification elements, such as progress tracking and achievement rewards, to maintain learner motivation. Figure 1 illustrates a typical scenario, where users engage in real-time spoken interactions within a simulated environment. The study assessed the effectiveness of these features in supporting vocabulary retention while also identifying potential usability challenges. Particular attention was given to interface design, ease of navigation, and cognitive load, as AR applications require learners to process both virtual and real-world stimuli simultaneously.

Procedure. Each participant completed a three-phase structured session, beginning with a pre-test to assess their initial vocabulary knowledge. They then engaged with the Mondly AR application, interacting with augmented objects and practicing pronunciation through speech recognition exercises. The session concluded with a post-test measuring vocabulary retention and a usability questionnaire evaluating the overall learning experience.

Data Collection and Analysis. Data was collected through a combination of learning assessments, usability questionnaires, and qualitative feedback. The pre- and post-test scores provided a measure of vocabulary improvement, while usability ratings captured participants' perceptions of ease of use, responsiveness, and interface clarity. Qualitative responses were analyzed to gain insights into learner adaptation, cognitive demands, and overall satisfaction with AR-based language learning. Thematic analysis was used to identify recurring usability challenges and determine how well AR supports interactive language acquisition.

3.2 VR Study: Immersion and Language Acquisition

The VR study investigated the role of ImmerseMe VR, an immersive language-learning application designed for conversational practice and pronunciation training. The study explored how virtual environments influence learner engagement and whether VR enhances language learning beyond traditional methods.

Participants and Study Design. Similar to the AR study, participants were recruited through academic networks and online platforms, ensuring a diverse sample with varying degrees of VR experience. The study followed a between-subjects design, where one group completed language exercises in VR while a control group used the same application on a standard desktop interface. The inclusion of both VR and non-VR conditions allowed for a direct comparison between immersive and traditional learning methods, providing insights into the added value of virtual environments.

Application. The ImmerseMe VR application was selected due to its focus on simulated real-world conversations and interactive speech-based exercises. As shown in Fig. 2, learners engage with interactive word-learning exercises, reinforcing language retention through real-time feedback. Unlike traditional language-learning software, ImmerseMe VR places learners in culturally relevant environments, such as restaurants, airports, and markets, where they must navigate conversations naturally. The app's adaptive difficulty levels allow learners to progress based on their speech accuracy and fluency. The study examined how these immersive elements affected user engagement and whether they contributed to improvements in pronunciation and conversational confidence. It also assessed usability factors, particularly the ease of interaction within VR environments, responsiveness of the speech recognition system, and potential challenges such as motion discomfort.

Procedure. Each participant completed a structured learning session consisting of an introduction, pre-test, interaction phase, and post-test. At the start, they were introduced to the VR headset and the ImmerseMe application, followed by a pre-test evaluating their pronunciation and listening comprehension skills.

Participants were assigned to different experimental conditions based on the application, device, and target language. The study included four conditions for each application (Mondly and ImmerseMe), resulting in a total of eight experimental groups. Table 1 outlines the distribution of conditions.

Table 1. The experiment conditions

Application	Device	Language	Code
Mondly VR	VR Headset	Greek	C1
Mondly VR	VR Headset	Indonesian	C2
Mondly PC	Desktop	Greek	C3
Mondly PC	Desktop	Indonesian	C4
ImmerseMe VR	VR Headset	Greek	C5
ImmerseMe VR	VR Headset	Indonesian	C6
ImmerseMe PC	Desktop	Greek	C7
ImmerseMe PC	Desktop	Indonesian	C8

Following the pre-test, participants engaged with either Mondly VR or ImmerseMe VR using a virtual reality headset or completed the same exercises on a desktop version of the application. Those in the VR groups practiced interactive dialogues and pronunciation exercises within immersive scenarios, while participants in the PC-based conditions followed structured text and audio prompts.

After the interaction phase, all participants completed a post-test to measure pronunciation improvements and comprehension accuracy. Those in the

VR conditions additionally completed the Igroup Presence Questionnaire (IPQ) to assess their sense of presence in the virtual environment. A usability questionnaire was also administered to evaluate navigation, cognitive load, and the overall learning experience.

Data Collection and Analysis. A mixed-method approach was used to analyze learning performance, engagement, and usability factors. Pronunciation and comprehension scores from pre- and post-tests were compared across VR and non-VR conditions to determine the effectiveness of immersion in language learning. In addition to quantitative assessments, qualitative feedback was collected through open-ended user reflections, where participants described their experiences, challenges, and perceived benefits of VR-based learning. This feedback was analyzed to identify common themes, such as learner motivation, usability challenges, and cognitive workload. By comparing the VR and non-VR groups, the study provided insights into how immersion influences language retention and whether virtual scenarios enhance engagement and confidence in conversational practice.

Fig. 1. A screenshot of a conversational Mondly VR scenario.

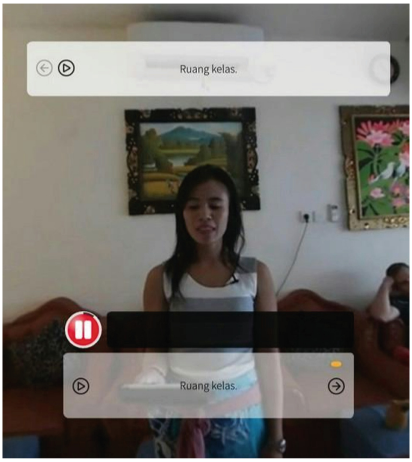


Fig. 2. A screenshot of the vocabulary module in ImmerseMe.

3.3 Ethical Considerations

Both studies were conducted in accordance with ethical research guidelines, ensuring participant safety, privacy, and informed consent. Prior to participation, all individuals received a detailed explanation of the study objectives, procedures, and potential risks. Participants voluntarily provided informed consent and were given the option to withdraw from the study at any stage without consequences. All collected data was anonymized, and no personally identifiable information was recorded. The research followed data protection regulations, ensuring secure storage and responsible handling of participant responses. By maintaining these ethical standards, the study ensured the reliability and validity of its findings while prioritizing participant well-being.

4 Results

This section presents the findings from the AR and VR studies, focusing on the impact of both technologies on vocabulary acquisition, pronunciation, usability, and user experience. The results are structured to allow for a clear comparison of both immersive learning technologies while also highlighting individual strengths and limitations. The analysis is based on both quantitative performance metrics and qualitative user feedback.

4.1 AR Study Results

The AR study aimed to examine how augmented reality supports vocabulary learning, particularly in terms of usability and learner engagement. The study included 45 participants, each engaging with Mondly AR for structured vocabulary training. To assess the effectiveness of AR-assisted language learning, a pre-test and post-test comparison was conducted, revealing a statistically significant improvement of 12.5% in vocabulary acquisition ($p = 0.03$). This suggests that AR-based learning can enhance language retention in the short term.

Participants were also asked to evaluate the usability of the Mondly AR application (restuls shown with Fig. 3, which received an average rating of 4.1 out of 5, indicating high user satisfaction with the interface design and navigation. Eighty percent of participants found the app easy to navigate, highlighting its accessibility for learners with varying levels of technological experience. However, fifty percent of users rated the visual appeal as only 2 out of 5, suggesting that while the interface was functional, improvements could be made to enhance the visual design.

A key aspect of the study was understanding the relationship between usability and learning performance. Correlation analysis revealed a positive relationship between usability ratings and vocabulary gains ($r = 0.42$, $p = 0.02$). This finding indicates that a well-designed interface and seamless user experience contribute to better learning outcomes. However, despite improvements in vocabulary retention, seventy percent of participants reported no significant increase in their confidence in speaking the target language after using the application. This suggests that while AR is effective for vocabulary acquisition, it may not sufficiently support broader language skills such as fluency and spontaneous speech production.

Another noteworthy aspect was the demand for greater customization in AR learning environments. Sixty-six point seven percent of users expressed a desire for adjustable settings that would allow them to tailor the difficulty level, interaction type, or pace of learning. This highlights the importance of personalization in educational applications to accommodate different learning preferences.

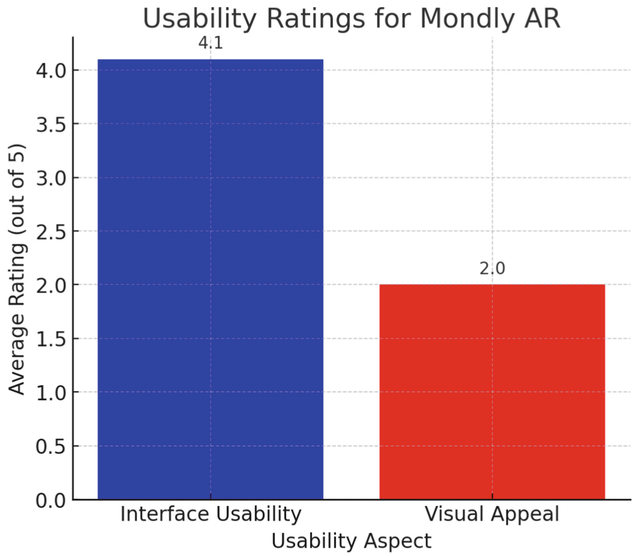


Fig. 3. Usability Ratings for Mondly AR, comparing interface usability and visual appeal.

The AR study demonstrated that augmented reality can be an effective tool for vocabulary acquisition, offering an engaging and interactive way to reinforce learning. However, limitations such as low visual appeal, limited long-term impact on confidence, and the lack of personalization suggest areas for future improvement.

4.2 VR Study Results

The VR study focused on pronunciation training, listening comprehension, and conversational practice, comparing the effectiveness of VR-based learning to non-VR methods. A total of 31 participants were divided into two conditions: a VR group that engaged with ImmerseMe VR and a non-VR control group that used a desktop version of the same application. The goal was to determine whether immersion in a virtual environment leads to higher learning gains compared to traditional screen-based learning.

Performance assessments showed that participants in the VR condition demonstrated an average improvement of 8.3 points (SD = 2.1) in their language test scores, whereas those in the non-VR condition achieved an average gain of 4.5 points (SD = 1.8). This indicates that VR-based language learning provides a measurable advantage over non-immersive methods, supporting previous research suggesting that immersion enhances retention and engagement.

To further investigate the role of immersion, the Igroup Presence Questionnaire (IPQ) was used to assess the perceived level of presence within the virtual environment, as shown with Fig. 4. The results for the VR group were as follows:

general presence score of 3.7, spatial presence score of 3.5, involvement score of 3.9, and experienced realism score of 3.6. These scores indicate a moderate-to-high level of presence, suggesting that participants felt relatively immersed in the virtual environment. Notably, involvement received the highest score, which may reflect the interactive nature of VR-based language tasks.

Session duration was also analyzed, with VR learners spending an average of 7.32 min per round (SD = 1.74 min), compared to 6.03 min (SD = 2.16 min) in non-VR conditions. This suggests that VR environments encourage longer and more engaged interactions compared to desktop-based learning.

Further analysis examined the correlation between presence and learning gains. A moderate positive correlation ($r = 0.45$, $p = 0.04$) was found between overall presence ratings and improvements in language scores. This suggests that the more immersed participants felt in the VR environment, the more they benefited from the learning experience.

Session duration varied between applications. Participants in Mondly VR spent an average of 7.54 min, while those using ImmerseMe VR engaged for 7.10 min per session. In contrast, non-VR users had significantly shorter interactions, with Mondly PC averaging 7.08 min and ImmerseMe PC averaging only 4.98 min. These findings indicate that VR encourages longer engagement, which may contribute to better learning retention.

Despite these positive findings, no significant correlation was found between prior VR experience and language improvement. This suggests that VR-based learning can be effective for both experienced and novice users, making it accessible to a wider range of learners. While VR appears to be a promising tool for language learning, future research should explore how different levels of interactivity and feedback mechanisms influence the long-term retention of language skills.

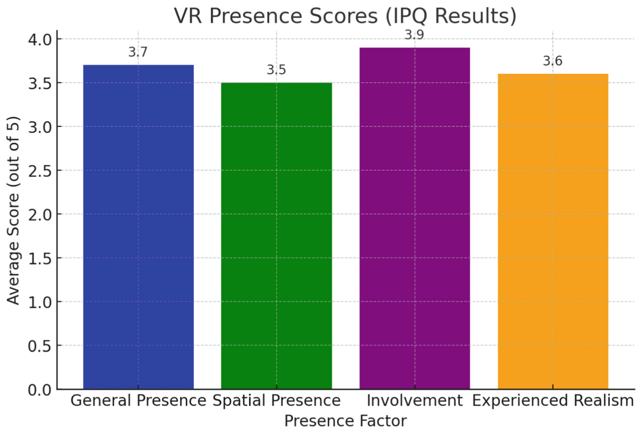


Fig. 4. VR Presence Scores (IPQ results), showing General Presence, Spatial Presence, Involvement, and Experienced Realism.

5 Discussion

This section discusses the key findings of the study, comparing the impact of AR and VR on language learning while considering their strengths, limitations, and alignment with previous research. The implications of these results for educational practice and future research are also explored.

5.1 Comparison of AR and VR Findings

The results indicate that both AR and VR enhance language learning, but they serve different purposes and impact learners in distinct ways. The AR study demonstrated that augmented reality is effective for structured vocabulary learning, supporting previous research that highlights AR's ability to integrate digital content into real-world settings and reinforce word associations through interactive engagement [2]. The usability ratings suggest that AR applications are generally accessible and intuitive, making them suitable for a wide range of learners. However, despite positive usability feedback, AR was found to have limited influence on spoken language confidence, which aligns with prior findings that AR-based learning tends to focus on object recognition and passive vocabulary recall rather than conversational fluency [9].

In contrast, VR was found to be more effective in supporting pronunciation and conversational practice, aligning with research that emphasizes the role of immersion and presence in second language acquisition [12]. Participants reported high levels of engagement and involvement, suggesting that VR fosters a more interactive and dynamic learning environment compared to traditional or screen-based learning methods. These findings are consistent with studies that highlight the role of spatial presence and realism in language immersion, where learners engage in simulated conversations that closely mimic real-world interactions [8].

A key distinction between the two technologies is how they influence learning engagement and cognitive load. AR applications provided structured learning experiences that were easier to navigate and adapt to, but they lacked customization features that could enhance long-term engagement. Previous studies have noted similar findings, where AR is effective for controlled learning tasks but does not inherently encourage spontaneous or adaptive learning behaviors [10]. In contrast, VR-based learning required longer session durations, which may indicate a higher cognitive demand associated with navigating virtual spaces, processing interactive dialogues, and maintaining presence [6]. While the increased engagement in VR can be beneficial, it may also lead to higher cognitive load, making it necessary to optimize session length and interaction complexity for different learner levels.

The differences in usability ratings also reflect broader challenges in immersive learning environments. AR was rated highly for usability and interface clarity, reinforcing the idea that AR applications, particularly those designed for mobile, benefit from a familiar interface and minimal hardware requirements [2].

However, its effectiveness in language retention and spontaneous language use remains limited. VR, on the other hand, was perceived as more engaging and immersive, but it also presented accessibility barriers due to hardware costs and learning curve. This contrast aligns with previous discussions on the trade-offs between immersion and accessibility in language technology [10].

Taken together, these findings suggest that AR and VR support different stages of language learning. AR is well-suited for beginners focusing on vocabulary acquisition and structured exercises, while VR is more beneficial for learners looking to practice conversational fluency in immersive contexts. This differentiation is crucial for educators and developers aiming to design adaptive and hybrid learning environments that integrate both technologies effectively.

5.2 Implications and Future Directions

The findings of this study have important implications for educational practice, technology development, and future research. The effectiveness of AR in vocabulary acquisition suggests that it can be integrated into classroom settings as a supplement to traditional language instruction. However, the low impact on conversational skills highlights the need for adaptive learning models that incorporate dialogue-based interaction, speech recognition, and contextualized language exercises. Future AR applications should also focus on improving interface aesthetics and expanding user control over learning paths to foster better engagement.

The strong engagement and presence reported in VR learning underscore the importance of immersion in language acquisition. However, VR also presents challenges such as higher cognitive load, extended session durations, and accessibility limitations due to hardware requirements.

To address these challenges, structured guidance, adaptive difficulty levels, and personalized feedback mechanisms should be incorporated into VR-based learning experiences [6]. Previous research suggests that shorter, goal-oriented VR sessions may be more effective in preventing cognitive overload while maintaining engagement [10].

A key insight from this study is that prior VR experience did not significantly influence learning outcomes, suggesting that VR learning can be accessible to a broad range of users, including those unfamiliar with immersive technology. However, the long-term effects of VR-based language learning remain underexplored. Future studies should examine whether repeated exposure to VR enhances language retention over time and whether combining AR and VR creates a more comprehensive learning experience [2].

Additionally, future research should explore the social dimensions of immersive language learning. Studies have shown that collaborative learning in VR settings can increase motivation and knowledge retention, yet the potential for multi-user VR language learning environments remains largely untapped [10]. Incorporating peer interaction and real-time feedback could significantly enhance the effectiveness of VR-based conversational training.

Another key consideration is infrastructure and accessibility. While VR offers strong engagement benefits, its adoption in public education remains limited by cost, space requirements, and motion sickness concerns. AR, being more accessible through standard mobile devices, provides a practical alternative for schools and educational institutions with limited budgets. Future research should investigate how to make VR learning more scalable and cost-effective, potentially through cloud-based VR solutions or simplified headset designs that reduce barriers to adoption.

Overall, this paper reinforces the idea that AR and VR each play a unique role in language learning. AR applications are effective for structured learning, particularly in vocabulary acquisition, whereas VR provides immersive, interactive experiences that enhance pronunciation and conversational fluency. Future research should focus on hybrid learning models that integrate AR for structured learning and VR for real-world interaction, maximizing the benefits of both technologies. By refining interface design, improving personalization, and optimizing session structures, immersive learning technologies can be developed into powerful tools that support diverse language learners across different learning environments.

6 Conclusion

Language learning benefits from both AR and VR, as each technology brings unique advantages while also presenting certain challenges. AR offers interactive and engaging learning experiences but is constrained by usability challenges and a lack of structured content. VR fosters high levels of engagement and immersion, yet its impact on language acquisition remains inconclusive. Future research should prioritize refining usability, integrating adaptive learning frameworks, and aligning immersive technologies with pedagogical best practices to maximize their educational potential.

References

1. Álvarez-Marín, A., Paredes-Velasco, M., Velázquez-Iturbide, J.Á., Palma-Chilla, L.: Insights into usability, academic outcomes, and emotional responses in an AR-interactive learning environment. Interact. Learn. Environ. 1–15 (2024)
2. Cai, S., Liu, E.: Augmented reality in language learning: improving engagement and comprehension. Int. J. Educ. Res. **112**, 102093 (2022). https://doi.org/10.1016/j.ijer.2022.102093
3. DeKeyser, R.M.: The robustness of critical period effects in second language acquisition. Stud. Second. Lang. Acquis. **22**(4), 499–533 (2000)
4. Deng, L., Yu, S.: The impact of virtual reality on language learning: a systematic review. Educ. Technol. Soc. **25**(3), 45–58 (2022)
5. Hardison, D., Son, J.: Usability and pedagogical design in immersive language learning environments. Lang. Learn. Technol. **16**(3), 39–55 (2012)
6. Hua, X., Wang, J.: Virtual reality and augmented reality in language education: advances and challenges. Comput. Educ. **185**, 104512 (2023). https://doi.org/10.1016/j.compedu.2023.104512

7. Imran, M., Almusharraf, N., Ahmed, S., Mansoor, M.I.: Personalization of e-learning: future trends, opportunities, and challenges. Int. J. Interact. Mob. Technol. **18**(10) (2024)
8. Kang, Y., Shin, D.: The role of prior language experience in virtual reality language learning. Second. Lang. Res. **37**(4), 567–589 (2021). https://doi.org/10.1177/0267658321998871
9. Nicolaidou, I., Antoniou, P., Constantinou, R.: Exploring virtual reality as a medium for second language acquisition: engagement and learning outcomes. J. Lang. Learn. Technol. **25**(2), 30–48 (2021)
10. Parmaxi, A.: Virtual reality in language learning: a systematic review and future prospects. Educ. Tech. Res. Dev. **68**, 1509–1531 (2020). https://doi.org/10.1007/s11423-020-09761-7
11. Pillai, R., Johnson, T.: Measuring the effects of augmented reality on cognitive load in language learning. Interact. Learn. Environ. **21**(6), 615–633 (2013). https://doi.org/10.1080/10494820.2013.832112
12. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: Factor analytic insights. Pres. Teleoperators Virt. Environ. **10**(3), 266–281 (2001)
13. Wedyan, M., Falah, J., Elshaweesh, O., Alfalah, S.F., Alazab, M.: Augmented reality-based English language learning: importance and state of the art. Electronics **11**(17), 2692 (2022)
14. Wen, T.H., Miao, Y., Blunsom, P., Young, S.: Latent intention dialogue models. In: International Conference on Machine Learning, pp. 3732–3741. PMLR (2017)

Enhancing Three-Dimensional Rendering Skills Through Virtual Reality: A Case Study of a Virtual Photography Studio

Ling Lee and Ming-Huang Lin(✉)

National Taiwan University of Science and Technology, Taipei, Taiwan
linludwig@mail.ntust.edu.tw

Abstract. This study aimed to test whether students could effectively apply lighting knowledge to 3D rendering software through VR virtual photography studio. The study was inspired by the authors' observation that design students are often unable to apply lighting techniques in 3D rendering processes. Therefore, we adopted an experimental approach to compare learning outcomes in a photography lighting course conducted in two environments: VR photography studio, and computer-based 3D rendering program.

15 participants were divided into an experimental group (VR photography studio running in Meta Quest 3) and a control group (Keyshot on Windows 10). The VR studio was created using an existing VR sandbox game programmed in the Unity framework to simulate a real-world photography studio. The studio interface was designed to be user-friendly. Experimental records comprised pretest and posttest results, time taken for the tests, and participants' feedback.

On the basis of the experimental results, we reached three conclusions: (1) Even in a foundational course, integrating photography lighting knowledge into 3D rendering processes can improve students' efficiency in product rendering and the quality of the rendered products. (2) Combining the advantages of spatial immersion and data-driven lighting adjustments, the VR photography studio integrates physical photography and 3D rendering. This integration facilitates the smooth application of lighting principles in real-world settings, and the engaging nature of VR enhances students' willingness to learn. (3) Further development of 3D rendering software to enable direct lighting manipulation in a VR environment can greatly enhance the precision of rendering tasks.

Keywords: Virtual reality · product photography · 3D rendering · design education · technology education

1 Introduction

Advancements in virtual reality (VR) technology have greatly increased its applications in various fields. In design and architecture, VR can assist designers in not only three-dimensional (3D) modeling and prototyping for shape exploration but also in early-stage ideation and collaborative design. In design education, VR effectively enhances

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 145–162, 2025.
https://doi.org/10.1007/978-3-031-93715-6_10

student learning outcomes and practical skills (Ibrahim et al. 2021). Due to the immersive and interactive nature of VR, it is attractive as an educational tool. However, research focusing on the integration of VR specifically in photography education and 3D rendering remains scarce (Fuchs & Grimm 2023). Most studies have concentrated on conventional photography education or general virtual photography systems; a thorough investigation into the applicability of VR for product photography and 3D rendering education is lacking.

The motivation for this study arose from observing the challenges design students face when applying photography knowledge in the 3D rendering process. Despite understanding fundamental photography theory, design students often struggle to transfer and apply real-world lighting techniques in 3D rendering software. This difficulty is mainly attributable to a lack of spatial awareness and practical experience. Existing 3D rendering tools, such as Keyshot, can quickly generate high-quality images; however, their two-dimensional (2D) screen-based interfaces often cause confusion and frustration during the lighting process, hindering users' ability to improve the realism and aesthetic quality of their work.

To address this problem, this research team developed a virtual photography studio and designed a product photography lighting course. An experimental approach was adopted to compare the effectiveness of two learning environments: one involving an immersive and interactive VR-based virtual photography studio and the other using a conventional computer-based 3D rendering program. This study aimed to investigate the potential of VR technology in bridging real-world photography and digital rendering knowledge and to explore the effect of VR on enhancing student learning outcomes.

2 Literature Review

2.1 Virtual Reality in Education and Training

VR is an immersive technology that generates 3D environments which simulate real or imagined experiences. The concept of "presence" in psychology refers to an individual's perception of a specific environment (Cypress & Caboral-Stevens 2022). Presence in VR is achieved by providing users with interactive experiences that mimic the real world. By replicating visual, auditory, and tactile sensory experiences, VR offers a high level of immersion (Guo 2022; Loureiro et al. 2020), creating a sense of realism that blurs the boundaries between the real and virtual worlds.

In the 2000s, the widespread use of personal computers and the advancements in computer graphics, processing power, and display technology led to major breakthroughs and the broader adoption of VR in various fields, including applications in aerospace training, medical fields, and architecture and engineering education. After 2010, technological developments and the introduction of affordable consumer-grade VR headsets, such as the Oculus Rift, HTC Vive, and Sony PlayStation VR, have considerably reduced the barrier to entry of VR application. In architecture and design education, VR headsets have enabled students to explore virtual buildings at a 1:1 scale and simulate construction processes (Ibrahim et al. 2021).

VR was first applied in design education in the 1990s. Macpherson and Keppell (1998) explored the potential of VR-based visual simulations to help students understand complex 3D structures in architectural and spatial design. In the 2000s, VR applications expanded into the field of product design. Li (2009) proposed using VR technology for product design training, highlighting its value in model construction and simulation tasks during the later stages of product design. Subsequently, VR applications in product design gradually shifted toward the early ideation phases. In the 2010s, VR was applied to improve design efficiency and educational outcomes in both product and architectural design. Kamińska et al. (2019) discussed the use of VR in engineering and product design education. They particularly emphasized its role in accelerating the design iteration process. In addition, VR has been applied in collaborative design. Pellas et al. (2020) noted that VR facilitates collaborative learning and design education by enhancing design reviews and real-time feedback mechanisms, thereby improving the efficiency of distributed design teams. Wang et al. (2018) integrated VR with building information modelling in architectural education and reported practical cases in which engineers, clients, architects, and contractors collaborated in VR-based virtual construction environments (common data environments). Several studies have indicated that VR offers opportunities to immersively experience spatial relationships. Interaction in virtual spaces tends to be motivational and thereby encourage users to explore virtual worlds (Dörner et al. 2022; Makransky & Petersen 2021).

2.2 Photorealistic 3D Rendering

Photorealistic 3D rendering is a computer-generated technique that uses algorithms, such as ray tracing and global illumination, to simulate real-world physical phenomena, including illumination, materials, perspectives, and interactions between light and objects. This process creates digital images that are nearly indistinguishable from photographs. Photorealistic 3D rendering is widely applied in architectural visualization and helps designers present highly realistic architectural renderings. The technique is also extensively used in film production, architectural visualization, product design, and game development (Abidin et al. 2003; Jong et al. 2007; Seidler 2019). Photorealistic 3D rendering includes several core steps.

1. Modelling/Import: Creating 3D objects from scratch or importing existing models into the rendering environment. These models define the main elements and structure of the scene.
2. Staging/Compositing/Framing: Arranging the layout of objects in the scene and guiding the viewer's perspective through the camera's angle, distance, and composition.
3. Lighting: Selecting light sources (key light, fill light, and rim light) and adjusting their attributes, such as brightness, colour, direction, and shadow effects, to simulate real-world lighting conditions. Proper lighting setups can create atmosphere and enhance depth.
4. Surface Texturing: Adding physical and visual properties to objects in the scene, including colour, reflectivity, transparency, and surface details, to simulate a variety of realistic materials, such as metal, wood, or glass. Accurately configuring material properties ensures that, during subsequent shading, light interacts realistically with object surfaces, producing effects such as smooth reflections or diffuse scattering.

5. Shading: This process simulates the behaviour of light interacting with the material properties of a model's surface, including reflection, refraction, and diffusion. In contrast to lighting, which focuses on providing illumination and creating atmosphere, shading involves the interaction between light and object surfaces, highlighting material details and textures to ensure that the surface appears realistic and tactile (Abidin et al. 2003).

Similarities Between Photorealistic 3D Rendering and Photography. Abidin et al. (2003) stated that both techniques rely on spatial perception and camera techniques to simulate or capture realistic visual effects. These similarities include the following:

- Depth of Field: This technique simulates the blurring of objects inside and outside the focal area. Blurring the background or foreground helps direct the viewer's attention to the subject and creates a sense of depth in the image.
- Motion Blur: This effect captures the blurring of moving objects, providing a sense of motion and reducing the stiffness or static nature of the image.
- Basic Design Understanding: Effective composition relies on principles such as contrast, repetition, alignment, proximity, and the Rule of Thirds. These principles guide the arrangement of elements within a frame, ensuring visual depth, balance, and appeal while drawing the viewer's gaze to the focal point of the scene.
- Lighting Arrangement and Shadows: Proper lighting setup is crucial for achieving realism. The three-point lighting technique—comprising key light, fill light, and rim light—ensures that the subject has adequate brightness and separation from the background. Different types of light sources, such as point light, directional light, and ambient light, control the effects of shadows, reflections, and highlights. Generated by both artificial and natural light, shadows are typically categorized into cast shadows and drop shadows. The combination of these elements determines the realism of the scene.

Achieving photorealistic rendering requires both technical skills and a solid understanding of photographic principles. Without a firm foundation in lighting techniques and composition rules, even users proficient with 3D rendering tools may not create compelling and artistically impactful work.

2.3 Virtual Photography Studio

To master photography skills, students must gain ample practical shooting experience as well as acquire sufficient knowledge of camera operation and lighting techniques. Although the technical aspects of cameras and their effects on image design can be learned theoretically, developing an understanding of composition, visual aesthetics, and expressive ability requires practical education and personal experience (Fu & Zhang 2021).

During the practical stages of coursework, a common problem for students is the lack of adequate hardware and space. Photography classes often involve group activities using a single camera, and a few students familiar with camera operation dominate the task while others participate passively. This dynamic reduces some students' learning motivation and efficiency. Furthermore, the learning process lacks immediate feedback; students cannot promptly review and evaluate their work while shooting, which hampers their progress and skill development. The learning experience is also insufficiently immersive or personalized. In physical environments, students often lack the freedom to experiment with different perspectives or setups. Moreover, spatial and environmental limitations prevent the recreation of specific scenarios, such as those involving particular lighting conditions or background settings (Fuchs & Grimm 2023).

Research on VR and photography learning has gradually increased in recent years. Fu and Zhang (2021) analysed the characteristics of VR technology and proposed its application in university photography education. They demonstrated how virtual spaces can improve students' practical skills. However, the study was only preliminary, and systematic evaluations of student learning outcomes were not fully developed. Fuchs and Grimm (2023) developed a VR-based photography studio that simulates a real-world studio environment. The system supported real-time adjustments of camera parameters, such as focal length and depth of field, and incorporated guidance, motivational, and feedback mechanisms to enhance students' technical skills and creativity. Despite the system's potential, it is still in the development phase, and further investigation is required to refine the system and expand its applications. Kobayashi and Nagao (2023) proposed a VR-based photography training system that allows users to photograph static subjects in a virtual space. The system used machine learning to assess the aesthetic quality, composition, and colour of portrait photos and offer corresponding composition suggestions to improve photographers' skills. However, the study focused on portrait photography and the proposed system required further validation of its evaluation accuracy and practical effectiveness. Juan et al. (2023) developed a VR photography application for panoramic images to evaluate users' short-term spatial memory and provide customizable learning environments. The application was mainly used for spatial memory testing and had limited utility for improving photography skills.

This literature review indicates that scholars have not specifically explored the use of VR in product photography or in learning lighting and composition techniques for product photography.

3 Research Method

The research objective was to determine whether students can effectively transfer lighting knowledge acquired in a VR virtual photography studio to the operation of 3D rendering software. Accordingly, the same lighting course content was used in both a VR virtual photography studio and a computer-based 3D rendering environment, and an experimental approach was employed to compare the learning outcomes.

3.1 VR Virtual Photography Studio Setup

First, four 3D mouse models were created in the Keyshot rendering software to serve as test objects for evaluating student performance before and after the course (Fig. 1). To simulate real-world studio conditions, a VR virtual photography studio was developed based on an existing VR sandbox game with the Unity framework. The virtual studio included three spotlights and eight objects, which corresponded to eight different course modules. Additionally, the studio setup featured several elements, such as cameras, studio stands, lighting equipment, and seamless background paper (Fig. 2). A user-friendly interface was designed with visual and numerical adjustment sliders (Fig. 3), enabling users to interact with the spotlights by moving them freely in the virtual space to adjust their position, angle, colour, intensity, and coverage area. The same set of objects and similar lighting conditions were replicated in the Keyshot software on a laptop for the computer-based learning environment.

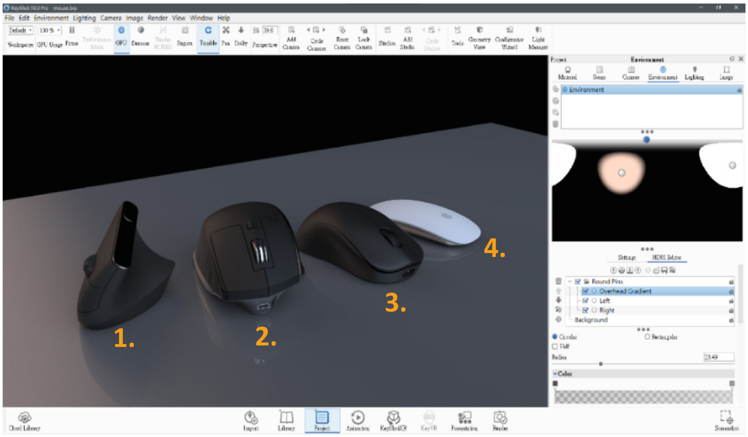


Fig. 1. Four 3D mouse models used to evaluate student performance before and after the course.

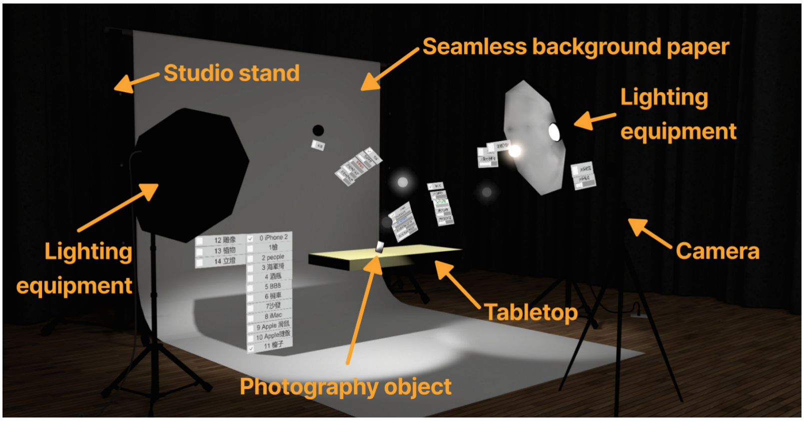


Fig. 2. VR photography studio.

3.2 Participants

A total of 15 participants aged 18 to 24 years were recruited from undergraduate and graduate programs. All participants had an interest in photography or 3D rendering but

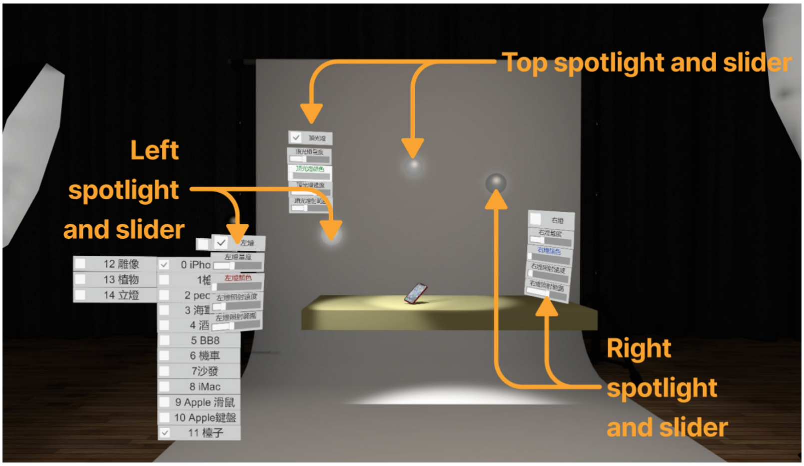


Fig. 3. Interface of the VR virtual photography studio for operational controls.

varying levels of experience. Experience levels were assessed through a questionnaire. The participants were divided into an experimental group (8 participants) and a control group (7 participants) such that the participants in each group had diverse experience levels with the goal of minimizing bias. The experimental group used a VR headset (Meta Quest 3) and a custom-designed VR virtual photography studio as their learning tool (Fig. 4). The participants conducted their tasks by freely moving within the VR environment. The control group used a Windows 10 computer and the Keyshot 3D rendering software (Fig. 5). All the subjects signed an informed consent form approved by the Institutional Review Board, and all ethical concerns related to the participation of human subjects in the study were addressed.

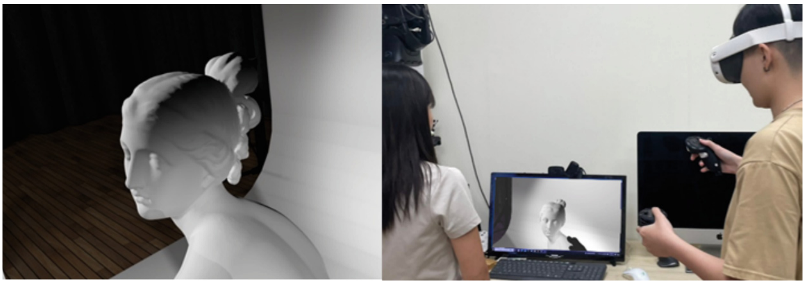


Fig. 4. Experiential learning scenario of the experimental group[1].

[1] https://youtu.be/0TeckTUozqw.

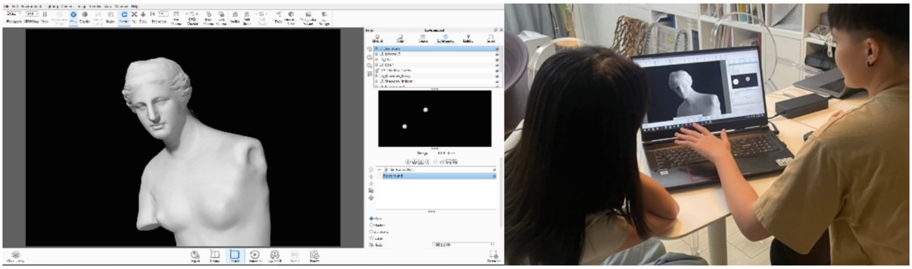


Fig. 5. Experiential learning scenario of the control group[2].

3.3 Experimental Design

The experimental process comprised three main stages: (1) Background survey and operation training, (2) photography lighting course and experimental tasks, and (3) feedback questionnaire and in-depth interviews. A mixed-methods approach was adopted to collect both quantitative and qualitative data. The experimental data included pretest and posttest results, the time taken for each test, and participants' feedback. The experimental process was documented with screenshots, audio recordings, and video recordings, and data were analysed separately for the experimental and control groups. The pretest and posttest results were evaluated by three experienced 3D rendering users and one design faculty member. They assessed the lighting details, background composition, and overall aesthetic quality of the rendered images before and after the course. The three stages are described as follows:

Background Survey and Operation Training. The researchers explained the purpose and procedures of the experiment and asked participants to complete a background survey to gather information on their experience with photography, rendering, and VR. Subsequently, participants were trained in the basic operations of either the VR virtual photography studio or Keyshot, including tasks such as moving the viewpoint and manipulating objects.

Photography Lighting Course and Experimental Tasks. This stage was divided into three parts: pretest, course learning, and posttest. (1) Pretest: All participants (both the experimental and control groups) used Keyshot to adjust lighting parameters—such as position, intensity, and colour—on four 3D mouse models provided for the experiment (Fig. 6). Screenshots of the results were captured (Fig. 6, left), and the researchers recorded the time required to complete the task. (2) Course learning with different tools: The researchers delivered a course covering eight fundamental photography techniques () using the appropriate teaching materials for the experimental and control groups. After a demonstration in either the VR virtual photography studio or Keyshot, participants were asked to reapply the lighting techniques and take screenshots. The researchers recorded the time needed to complete each task. (3) Posttest: Both groups used the same tool (Keyshot on Windows 10) for a posttest in which they applied lighting to the 3D mouse models displayed on the screen and took screenshots to document their progress (Fig. 6, right). The researchers recorded the time required to complete the task.

[2] https://youtu.be/ZseHCW072KQ.

Feedback Questionnaire and In-Depth Interviews. Participants completed a feedback questionnaire and participated in in-depth interviews. The interviews were recorded and transcribed for analysis.

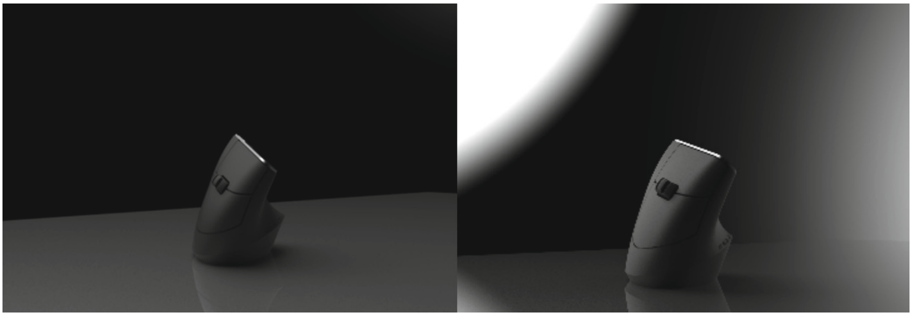


Fig. 6. Screenshots created by Participant D11 for Mouse 1 before (left) and after (right) the completing the course.

3.4 Content of the Photography Lighting Course

The course was designed in reference to two photography lighting books (Hunter et al. 2007; Tamauchi et al. 2012) and university-level product photography curricula. The course covered fundamental product photography lighting knowledge, including light placement, angles, colour temperature, brightness, distance adjustments, shadows, framing angles, depth perception, and material representation. Both the experimental and control groups learned the same course content but used different tools. The course was divided into six fundamental modules and two advanced modules, totalling eight lessons covering approximately two hours of material. Given the difficulties in sourcing identical materials, the examples used in the VR virtual photography studio and the Keyshot software differed slightly. However, the instructional content of the two groups was largely consistent, ensuring that the effects on their learning outcomes was minimal.

4 Results

4.1 Task Performance, Pretest, and Posttest Analyses

Although the statistical sample size was small, the pretest and posttest results were still sufficient for analysis. Table 2 presents the average time required for each task by participants after completing the eight-course modules. The experimental group (using VR as the instructional tool) took longer to complete each task compared with the control group (using Keyshot software on a computer). Observing the participants revealed that most of them were unfamiliar with VR, which led to a longer learning curve. On average, the participants required 126 s longer per task in VR than on a computer (approximately two-thirds more time); on the computer, simple actions could be easily completed with mouse clicks.

The scoring criteria were based on relevant reference materials and had six levels (0 to 5 points; Table 3). Participant performance was evaluated by three experienced 3D rendering software users and one design faculty member. Although individual differences existed among the participants, most of them had higher posttest scores than pretest scores (Table 4). The control group achieved a mean score increase of approximately 1.43 points, which was higher than the mean score increases of the experimental group (0.92 points). These findings suggest that a virtual photography studio could be effective for teaching lighting techniques. However, VR was less effective than computer-based methods as a teaching tool because it was less efficient and harder to operate. This was primarily attributed to most users being more accustomed to computer interfaces than VR interfaces. However, VR, as a novel learning tool, also had advantages. The effects of VR on learning motivation and student cognitive load and mental effort during the learning process were further explored through the feedback questionnaire and interviews. Moreover, the effects of these factors on affect their learning outcomes were investigated (Table 1).

4.2 Questionnaire and Interview Analysis Results

Integrating Studio Lighting Knowledge into Keyshot Improves Rendering Efficiency. Participants in both groups agreed that applying studio lighting knowledge to Keyshot rendering was useful. Understanding the logic of photographic lighting helped them systematically build a rendered image while avoiding repetitive and aimless adjustments of object angles and light positions. This approach also enhanced the overall quality of rendered results.

Participant C5: "*Keyshot lacks a clear sequence or logic, so I often wasted time repeating the same actions. Applying photographic lighting concepts helps me understand the process more logically.*"

Advantages of the VR Virtual Photography Studio. Using VR to learn photography lighting techniques offers several benefits, including considerably reducing the time and cost associated with traditional studio practice. Additionally, the engaging and interactive nature of VR enhances student learning motivation. Multiple participants noted that VR served as an effective bridge between the photography studio and 3D rendering software. It was also used to test lighting ideas.

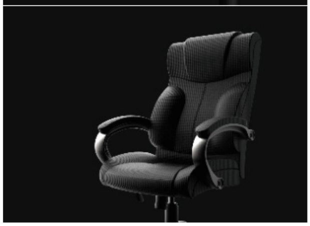
Participant E2: "*In VR, I gained a clearer understanding of how to set up lighting. It closely mimics the intuitive experience of physically moving lights in a real studio. Rendering on the computer does not provide this hands-on feeling.*"

Participant C2: "*From my personal experience, changing lighting elements in VR is smoother. I find using VR enjoyable. Working with Keyshot eventually became frustrating.*"

Participant E5: "*When I have an idea for a scene, I test it in VR first. It's much faster than testing the idea directly in Keyshot.*"

These advantages of VR are attributable to its provision of a virtual photography studio experience that enables students to engage with the lighting process in an immersive environment. This approach addresses the challenges of transitioning between 2D interfaces in Keyshot and the 3D environment of a real photography studio, helping students achieve more effective conversion between 2D and 3D objects.

Table 1. Main topics and demonstrations provided in the photography lighting course.

Main Topic		
Key light and fill light: Main light and contour lines, effects of different light positions		
Effect of light position on object: Gradient effects on the screen surface, backlight detail fill-in		
Enhancing object features through adjusting the size and angles of light sources: Highlighting product shape and contours, emphasizing leather textures and patterns, avoiding glare and reflections	 	 
Effect of light source angles, size, and colour temperature on shadows: Creating different scenarios (sunny, cloudy, and indoor conditions)		

(*continued*)

Table 1. (*continued*)

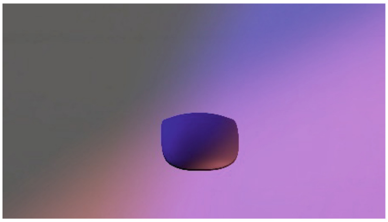
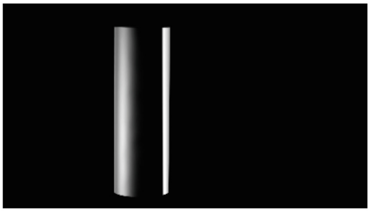
Using light source angles to create back-grounds: Combining object lighting with background lighting		
Rendering of metal materials		
Rendering of glass materials		

Table 2. Comparison of mean task completion times for the experimental and control groups (unit: seconds).

	Lesson 1	Lesson 2	Lesson 3	Lesson 4	Lesson 5	Lesson 6	Lesson 7	Lesson 8
Control group (Keyshot)	205 ± 37	253 ± 154	190 ± 68	188 ± 61	185 ± 74	245 ± 109	334 ± 206	270 ± 43
Experimental group (VR)	341 ± 105	399 ± 142	314 ± 119	276 ± 170	320 ± 183	437 ± 217	492 ± 337	359 ± 290

Participant E5: "*In Keyshot, light source adjustment occurs in a 2D space but is presented from a 3D perspective. After understanding relative positions through VR, my spatial awareness improved, and making adjustments was faster.*"

In addition, this study initially hypothesized that the sliders in the VR virtual photography studio used to adjust light brightness, coverage, and other effects (Fig. 8, left) would be unintuitive, and students would have difficulty relating them to real studio settings. However, the experimental results revealed that most participants were highly

Table 3. Learning outcome scoring criteria.

Scoring criteria	Standard
0	Poor
1	The object's outline is visible but incomplete; the background is not integrated
2	The object's outline is clear and the lighting highlights object details; the background is not integrated
3	Lighting on the object is complete, and the student has attempted to integrate the background
4	Lighting on the object is complete, and the background complements the object effectively
5	Lighting and composition are complete, and the image has aesthetic appeal

Table 4. Pretest and posttest scores of all participants.

Control group (Keyshot)			Experimental group (VR)		
Participant No	Pretest score	Posttest score	Participant No.	Pretest score	Posttest score
C1	2.5 ± 0.58	3.75 ± 0.5	E1	2.25 ± 0.5	3.25 ± 0.5
C2	3 ± 0	4 ± 0	E2	2.5 ± 1	3 ± 0
C3	1.5 ± 0.58	3.75 ± 0.96	E3	2.75 ± 0.5	3.75 ± 0.5
C4	1 ± 0	2.75 ± 0.96	E4	2.25 ± 0.5	3.5 ± 0.58
C5	2 ± 0	3.5 ± 0.58	E5	1.25 ± 0.5	2.75 ± 0.5
C6	2.25 ± 0.5	3 ± 0.82	E6	2 ± 0	3 ± 0.82
C7	2.5 ± 0.58	4 ± 0.82	E7	3 ± 0	3.25 ± 0.96
			E8	2 ± 1.41	2.25 ± 0.96
Mean	2.11 ± 0.67	3.54 ± 0.49	Mean	2.29 ± 0.57	3.21 ± 0.34

satisfied with the design of the sliders. They agreed that the sliders not only simplified complex lighting effects and enabled them to achieve their goals quickly but also facilitated a smoother transition to the similar slider functions in Keyshot.

Participant E1: "*The sliders are like digitized combinations of different studio lighting effects. It's easy to find the corresponding slider and value for a specific effect.*"

Participant E4: "*The parameters for adjusting light (colour, size, brightness range) in VR are similar to those in Keyshot, so it's easy to relate the two.*"

Moreover, the interface design of the VR virtual photography studio, which closely resembles that of Keyshot, helped participants efficiently identify corresponding functions when using Keyshot. This helped them apply lighting techniques and efficiently adjust lighting. For example, participants often added multiple lights in Keyshot to brighten dark areas, resulting in cluttered lighting (Fig. 7). By contrast, the three spotlight spheres provided in the VR studio correspond to the three light points in Keyshot (Fig. 8, right), effectively guiding participants to use standard key and fill lighting adjustments.

Participant C5: "*I kept adding lights—just a bit more here, a bit more there—to fix dark areas.*"

Participant E6: "*The light spheres in VR are similar to the light points in Keyshot. When I switch to Keyshot, I know exactly where to find a similar function instead of randomly experimenting.*"

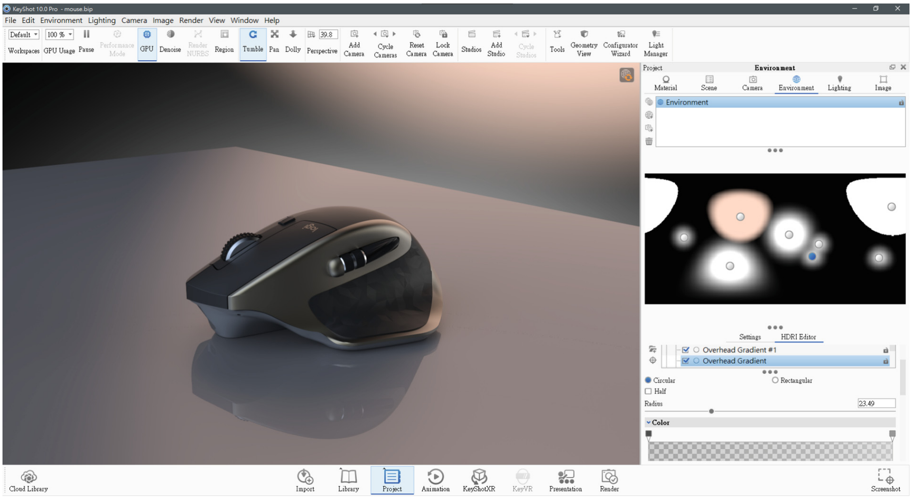


Fig. 7. Confusion caused by the overwhelming number of light sources provided in Keyshot.

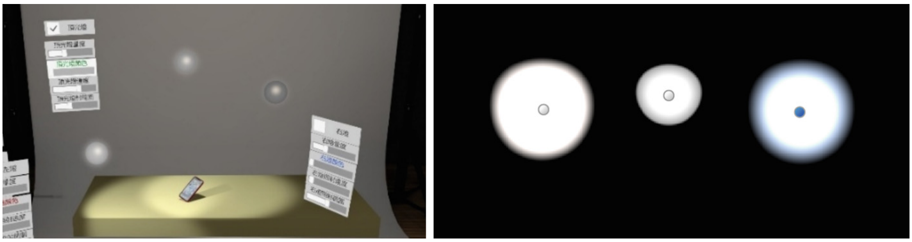


Fig. 8. Light adjustment interfaces of the VR virtual photography studio (left) and Keyshot (right).

Disadvantages of the VR Virtual Photography Studio. The experimental group took longer to complete tasks compared with the control group (Table 2). Participants attributed this to the higher complexity of using VR controllers for moving objects than using a computer mouse. They required additional time to adapt.

Participant E5: "*It took some time to get used to VR. I normally do not use controllers, so it took longer to adapt.*"

Another drawback mentioned by participants was the inability to preview the lighting effect on the photo. In real photography studios, multiple people typically operate the equipment; one person checks the viewfinder and directs the lighting adjustments. In Keyshot, users can directly observe and adjust the rendered image without an additional preview step. However, the VR virtual photography studio requires single-user operation. Without a viewfinder, users spent more time achieving the desired lighting setup.

Participant E2: "*A viewfinder is necessary so that I can see how the light affects the object while making adjustments.*"

Advantages of Keyshot Rendering Software. According to the participants, Keyshot as a professional 3D rendering software offers more refined visual effects than do the VR virtual photography studio**.** It also allows users to immediately see the outcomes of lighting adjustments.

Participant C7: "*Keyshot renders objects better than VR with more precise rendering effects. In Keyshot, I can immediately see how moving a light affects the object*."

Moreover, Keyshot's approach to adjustments is based on the relative positioning of objects and lights, rather than fixed numerical values. This flexibility supports creative and aesthetic experimentation, resulting in a rendering process that is more enjoyable than conventional photography. Users are more inclined to try different effects while anticipating the possibility of creating unexpected, visually pleasing outcomes.

Participant C4: "*In Keyshot, I can experiment freely and sometimes end up with something that looks surprisingly good. Real photography is focused solely on presenting the product in its best light, without room for creative exploration.*"

Disadvantages of Keyshot. Although Keyshot is a convenient and efficient tool for lighting in rendering, replacing a real photography studio with Keyshot can pose several challenges. The adjustments in Keyshot are disconnected from the theoretical foundations of photography; the software allows students to make adjustments based solely on the final image. This leads to difficulties for students who have lighting knowledge from real photography studios; they are likely to struggle when applying that knowledge systematically in Keyshot. For students without a foundation in photography, this is even more challenging.

Participant E3: "*I have no photography experience, so, although Keyshot is easy to adjust, I cannot figure out the exact position of the light source in 3D.*"

Participant C7: "*In Keyshot, I can achieve a satisfactory lighting result on the object, but if I were asked to replicate it in a real studio, I would not know how to position the lights.*"

Another problem lies in the transition from 2D to 3D objects and the process of adjusting lighting positions. Because Keyshot is displayed on a 2D computer screen, students with photography knowledge must rely on their imagination to understand the spatial relationship between objects and lights. This often leads to difficulties in accurately translating studio lighting techniques to Keyshot.

Participant C2: "*In real photography, I would position the key light and fill light diagonally, but, in Keyshot, I have to use the mouse to slowly adjust the light on the screen, and the light rarely appears where I imagine it.*"

Even students with strong spatial awareness often experienced frustration when they could not achieve their imagined lighting setup, requiring multiple attempts and consuming considerable time.

Participant C5: "*I know where the light should be, but I just cannot adjust it in Keyshot to match the image I have in my mind, which is frustrating.*"

For students without photography experience, their lack of understanding of lighting logic makes the lighting process difficult. Many resorted to basic material assignment and environmental lighting or did not strive for high-quality rendering results.

Participant C7: *"I usually adjust the overall environmental light rather than focusing on the details of light positioning."*

5 Discussion

The experimental data and interview analysis revealed several findings. First, when photography lighting knowledge was integrated into 3D rendering, even a basic understanding of photography lighting principles effectively enhanced students' rendering efficiency and the quality of product images produced. Second, the VR virtual photography studio could serve as a bridge between real-world photography and 3D rendering. Combining the advantages of spatial immersion and data-driven lighting adjustments, VR facilitated smooth application of lighting principles. The interactive nature of VR also increased students' motivation to learn. Third, student learning outcomes were improved by VR equipment; however, VR had limitations and this improvement was inhibited by their unfamiliarity with VR technology. Fourth, although Keyshot is a convenient tool for product design that enables users to quickly create high-quality renderings, the students' lack of spatial awareness when transitioning from 2D to 3D objects resulted in difficulty accurately positioning the lights. This led to frustration and suboptimal rendering outcomes, particularly when students could not apply their photography knowledge to the rendering process. Although various 3D rendering software tools such as Keyshot now incorporate VR elements, these functions are primarily designed for viewing the completed renderings in VR rather than making real-time adjustments. Therefore, 3D rendering software programs should be further developed to expand their VR-based functionality. For example, these programs should enable users to adjust the light source size, angle, and brightness directly in the VR environment instead of requiring them to complete the lighting process on a computer first and then switch to VR for viewing. Although building the VR photography studio for this study was challenging and the system was inferior to commercial software in terms of its interface and efficiency, it nonetheless demonstrated the potential of VR in improving the rendering workflow.

6 Conclusion

The objective of this study was to help students with no studio experience or limited photography experience to transfer their lighting knowledge to 3D rendering software. This research team constructed a VR virtual photography studio and compared the effectiveness of using VR and conventional 3D rendering software as teaching tools.

Participants completed a basic photography lighting course, and their performance was evaluated to compare the instructional outcomes of the types of two tools. The experimental results revealed that (1) integrating photography lighting knowledge into the 3D rendering process effectively improved students' rendering efficiency and the quality of product images. (2) The VR virtual photography studio acted as a bridge between real-world photography and 3D rendering, improving students' logical application of lighting techniques and increasing their learning motivation. (3) If current 3D rendering software programs are further developed to improve their VR-based interfaces and allow lighting adjustments directly in VR environments, they could greatly improve the precision of rendering tasks.

This study covered only basic photography lighting knowledge. Future research could incorporate more advanced photography techniques, such as focal length and depth of field adjustments, specialized lighting environments, light analysis, and grayscale pixel matrix maps, to further validate the effectiveness of VR-assisted studio photography courses. Additionally, while the current number of participants provides preliminary insights, increasing the sample size in future studies could enhance the reliability and validity of the findings.

Acknowledgments. This work was financially supported by National Science and Technology Council of Taiwan by grant number MOST 111–2410–H–011–029–MY2.

References

Abidin, M.I.Z., Razak, A.A., Joon, J.S., Anthasha, S.N.: Implementing understanding of photography principles to create effective photorealistic 3D rendering. In: Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003. (2003)

Cypress, B.S., Caboral-Stevens, M.: Sense of presence in immersive virtual reality environment: an evolutionary concept analysis. Dimens. Crit. Care Nurs. **41**(5), 235–245 (2022). https://doi.org/10.1097/dcc.0000000000000538

Dörner, R., Broll, W., Grimm, P., Jung, B.: Virtual and Augmented Reality (VR/AR) Foundations and Methods of Extended Realities (XR): Foundations and Methods of Extended Realities (XR) (2022). https://doi.org/10.1007/978-3-030-79062-2

Fu, Y., Zhang, N.: A study on the practicality of virtual reality technology in the teaching of photography in colleges and universities.(2021). https://doi.org/10.1109/IEIT53597.2021.00113

Fu, Y., Zhang, N.: A Study on the Practicality of Virtual Reality Technology in the Teaching of Photography in Colleges and Universities. 2021 International Conference on Internet, Education and Information Technology (IEIT) (2021, 16–18 April 2021)

Fuchs, A., Grimm, P.: Experience the theory: new perspectives through VR learning environments for photography education. In: 2023 International Electronics Symposium (IES) (2023)

Guo, K.: Application of virtual reality technology in the development of game industry. Highlights Sci. Eng. Technol. **15**, 102–106 (2022)

Hunter, F., Biver, S., Fuqua, P.: Light–science & Magic: An Introduction to Photographic Lighting. Focal Press (2007). https://books.google.com.tw/books?id=QzLuYhpQmocC

Ibrahim, A., Al-Rababah, A.I., Bani Baker, Q.: Integrating virtual reality technology into architecture education: the case of architectural history courses. Open House Int. **46**(4), 498–509 (2021)

Jong, S.J., Yuen, M.C., Khong, C.W.: A case study of integrating principles of photography and photorealistic for 3D rendering. IEEE (2007)
Juan, M.C., Estevan, M., Mendez-Lopez, M., Fidalgo, C., Lluch, J., Vivo, R.: A virtual reality photography application to assess spatial memory. Behaviour Inf. Technol. **42**(6), 686–699 (2023). https://doi.org/10.1080/0144929x.2022.2039770
Kamińska, D., et al.: Virtual reality and its applications in education: survey. Inf. (Switzerland) **10**, 318 (2019). https://doi.org/10.3390/info10100318
Kobayashi, H., Nagao, K.: VR training system to help improve photography skills. Appl. Sci. **13**(13), 7817 (2023). https://doi.org/10.3390/app13137817
Li, S.: Research on application of virtual reality technology in teaching and training. In: 2009 Second International Conference on Computer and Electrical Engineering (2009)
Loureiro, S.M.C., Bilro, R.G., de Aires Angelino, F.J.: Virtual reality and gamification in marketing higher education: a review and research agenda. Spanish J. Mark.-ESIC **25**(2), 179–216 (2020)
Macpherson, C., Keppell, M.: Virtual reality: what is the state of play in education? Australas. J. Educ. Technol. **14**(1) (1998)
Makransky, G., Petersen, G.B.: The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. Educ. Psychol. Rev. **33**(3), 937–958 (2021). https://doi.org/10.1007/s10648-020-09586-2
Pellas, N., Dengel, A., Christopoulos, A.: A scoping review of immersive virtual reality in STEM education. IEEE Trans. Learn. Technol. **13**(4), 748–761 (2020)
Seidler, D.R.: Revit Architecture 2020 for Designers. Bloomsbury Academic (2019). https://books.google.com.tw/books?id=rZMyEAAAQBAJ
Wang, P., Wu, P., Wang, J., Chi, H.-L., Wang, X.: A critical review of the use of virtual reality in construction engineering education and training. Int. J. Environ. Res. Public Health **15**(6), 1204 (2018)
Tamauchi, K., Hasegawa, O., Shimada, G., Kurokawa, T.: Masterclass in Advanced Commercial Photography: Unveiling the Secret Techniques of Professional Commercial Photographers. Sharp Point Press (2012).. https://books.google.com.tw/books?id=XzgjMwEACAAJ

The Application of Sharestart Teaching Method for Combining VR / AI in 3D Modeling Learning

Yu-Hsu Lee(✉) and Zi-Cong Hsu

addNational Yunlin University of Science and Technology, Yunlin 64002, Taiwan, Republic of China
{jameslee,m11331012}@yuntech.edu.tw

Abstract. The rapid advancement of technology has transformed design education, particularly in 3D modeling, rendering, and hand-drawing courses. This study explores the integration of the Learning by Teaching (Sharestart) method with AI image generation and VR tools in 3D modeling courses for industrial design students. Over a four-week curriculum, 20 s-year students are introduced to VR hardware, curve and surface manipulation, and Bing Copilot's AI image generation to construct vehicle components. Students replicate the process using Alias NURBS modeling, enhancing their understanding of form generation and construction techniques.

To evaluate effectiveness, the study adopts a mixed-methods approach, combining qualitative data from observations and interviews with quantitative assessments using the System Usability Scale (SUS). All participants achieving SUS scores above 65 confirm the system's usability. Retrospective interviews highlight the strengths and limitations of combining AI, VR, and traditional tools, offering insights for curriculum improvement.

This research demonstrates how modern technologies can enhance traditional 3D design education, providing students with interactive and practical learning experiences. The findings contribute to the ongoing discourse on integrating cutting-edge tools in design education, offering a framework for future pedagogical innovations.

Keywords: AI Image Generation · Virtual Reality · 3D Modeling · Sharestart

1 Introduction

1.1 Background and Motivation

In recent years, due to the declining birth rate and curriculum adjustments, computer-aided design (CAD) courses in industrial design programs have been changed from mandatory to elective. The original intention of this adjustment was to encourage students to first develop core design competencies and then invest additional time and effort in professional skills if they have a strong interest. However, in actual teaching practice, many students struggle to master relevant software due to the workload from other courses and a lack of sufficient self-directed practice time, which ultimately affects

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 163–180, 2025.
https://doi.org/10.1007/978-3-031-93715-6_11

their learning progress. Although instructors provide instructional videos for after-class practice, students generally lack motivation and often only acquire basic operational skills, with little interest in further exploration. As a result, after completing their final project, students tend to discontinue the application of their acquired knowledge in design practice.

To address these challenges, integrating new technologies and innovative teaching methods serves as an experimental approach to enhance both teaching and learning. Traditional 3D modeling training requires extensive time for learning and practice, yet the rapid advancement of technology has led to a fundamental transformation in design education. Modern technologies such as AI image generation and VR modeling have reshaped traditional teaching methods and tools. For instance, the combination of 3D printing, AI image generation, and VR modeling not only enhances efficiency but also significantly reduces reliance on manual craftsmanship and physical models.

This study aims to integrate traditional surface modeling instruction in design education with emerging technologies (such as VR and AI) to enable students to quickly grasp three-dimensional form concepts, compensating for the lack of time and hands-on experience. Given that the new generation of learners has been exposed to digital tools from an early age, educational approaches must shift from passive learning to interactive learning. This research employs the Learning by Teaching (Sharestart) method, where students engage with pre-recorded instructional videos for self-practice before class, supplemented by classroom discussions to stimulate interest and enhance self-directed learning capabilities.

1.2 Research Problem

This study primarily adopts the innovative Learning by Teaching (Sharestart) method, allowing next-generation students to engage in self-directed learning within traditional 3D modeling courses. By utilizing AI and VR technologies, students can develop a perceptual understanding of three-dimensional forms. The research problem consists of two main aspects:

Challenges in Design Education. How to enhance students' understanding of transportation design, especially for those who are completely unfamiliar with automotive design. The current 3D modeling courses have limited efficiency, as students' learning outcomes are constrained by time limitations and a lack of access to professional tools. While improving tool acceptance and course efficiency, it is also crucial to balance the development of students' creativity.

The Need for AI and VR Integration. Traditional Alias surface modeling can accurately process curved surfaces, but students often find it difficult to master quickly. The integration of VR modeling and AI image generation can compensate for the challenges in surface learning, helping students rapidly build form concepts and a sense of proportion.

1.3 Research Objectives

This study integrates the Learning by Teaching (Sharestart) method into traditional 3D modeling tool (Alias software) education, incorporating AI image generation and VR modeling as supplementary tools for developing form-thinking skills. This study focuses on three key objectives: the impact of the Learning by Teaching method, the stimulation of design ideas through emerging technologies, and the cognitive development of transportation design. The specific objectives are as follows:

Enhancing Learning Experience and Skill Mastery. By integrating the Learning by Teaching method, AI image generation, and VR modeling, this study provides students with a progressive learning framework, helping them acquire fundamental skills and tool proficiency essential for transportation design.

Fostering Creativity and Practical Application Skills. Image generation is utilized for form exploration and refinement training, while VR rapid modeling enhances students' design creativity and ideation process.

Providing a Teaching Model that Combines Theory and Practice. By incorporating AI and VR into the curriculum, students engage in an immersive learning experience, allowing them to grasp the core concepts of automotive design while developing adaptability to future design tools and creative problem-solving abilities.

2 Literature Review

2.1 Technology and Design Education

The rapid advancement of technology is transforming design education, particularly in 3D modeling, rendering, and hand-drawing courses. According to Huson (2006), 3D printing technology not only reduces the time and cost associated with the model-making process but also improves precision, allowing students to quickly materialize their design concepts. Rios et al. (2023) pointed out that AI generative technologies, such as the SHAP-E model, make the design process more creative, particularly during the early conceptualization stage. Additionally, Berg & Vance (2017) mentioned that the introduction of VR tools not only enhances immersive learning experiences but also helps students intuitively understand the construction of complex curves and surfaces.

2.2 Learning by Teaching Method

The Learning by Teaching (Sharestart) method has been widely explored in education. According to Thomas (2022), encouraging students to teach others can significantly enhance learning motivation and facilitate knowledge internalization. Cinar et al. (2024) further pointed out that integrating AI and VR into teaching methods provides learners with more interactive opportunities, fostering creativity and improving learning efficiency. However, the integration of these technologies in design education remains in its early exploratory stage, particularly in industrial design courses.

Lee and Pang (2023) categorized VR modeling into four stages: interface introduction, curves, surfaces, and SubD integration, using instructional videos to help students learn in an immersive environment. This study adopts a similar approach in VR education, guiding students to install all instructional videos on their VR devices at the beginning of the course, allowing them to engage in self-paced learning after class.

2.3 AI and VR in Education

Research on the application of Artificial Intelligence (AI) and Virtual Reality (VR) in education has rapidly increased in recent years. Jun and Nichol (2023) found that AI technology not only accelerates the generation of creative forms but also provides diverse design options; for instance, the SHAP-E model demonstrates higher efficiency in surface design generation compared to traditional modeling tools. Lee and Chiu (2023) and Lee and Lin (2023) conducted experimental studies using different AI tools to explore their applicability in industrial design processes, specifically in shape divergence (concept exploration) and shape convergence (design refinement). However, Hong et al. (2023) pointed out that AI-generated models may lack fine details, requiring further refinement by designers. On the other hand, Agkathidis and Gutierrez (2016) emphasized that the integration of VR in design education can provide an immersive learning environment, significantly enhancing students' practical skills and engagement.

2.4 Combining Traditional and Modern Tools

Traditional design tools, such as Alias and hand sketching, continue to play a crucial role in design education; however, they have a steep learning curve and complex operations. Drogemuller et al. (2024) highlighted that integrating parametric design tools (e.g., Grasshopper) with VR can help students rapidly construct geometric shapes while maintaining design consistency. Krish (2011) stated that combining parametric modeling techniques with digital design tools can enhance the efficiency of creative design, particularly for novice students. These approaches make design courses more adaptable to modern learning needs.

2.5 Research Gap

Although significant progress has been made in the application of AI and VR in education, research on the integration of these technologies with the Learning by Teaching (Sharestart) method remains limited (Thomas, 2022). Additionally, existing studies primarily focus on the technical performance of these tools rather than their actual effectiveness in industrial design education (Huson 2006). This study aims to bridge these gaps by employing a mixed-method approach, combining quantitative and qualitative analysis to assess the practical impact of AI and VR tools on enhancing students' creativity and learning efficiency.

3 Research Methodology

3.1 Research Design

This study adopts an action research approach, combining qualitative and quantitative methods to explore the application and impact of Artificial Intelligence (AI), Virtual Reality (VR) tools, and the Learning by Teaching (Sharestart) method in industrial design education. The primary objective is to assess how these modern technologies assist students in mastering 3D modeling skills, particularly in the design and manipulation of curves and surfaces.

This study consists of the following key elements:

Mixed-Methods Approach. Qualitative Methods: Non-participatory observation and retrospective interviews are conducted to gather student feedback on the teaching method and technological tools.

Quantitative Methods: The System Usability Scale (SUS) is used to measure the usability of VR tools, Alias software, Bing Copilot, and Vizcom.

Curriculum Structure. The 18-week semester course is divided into four modules (as shown in Fig. 1). Alias and VR (Gravity Sketch) fundamental training for 4 weeks each, and AI image generation for 2 weeks. After that, Advanced Alias modeling for automotive design for 8 weeks. Each session lasts 3 h per week, gradually introducing technological tools and design concepts to enhance students' skills. The curriculum covers VR hardware and software operations, curve and surface modeling techniques, and ultimately integrates Bing Copilot, Vizcom and Alias to complete the transportation component design.

Alias basic (Sharestart)	**VR modeling (Sharestart)**	**AI Image Generation (Demo)**	**Alias (Class Demonstration)**
4 weeks	**4 weeks**	**2 weeks**	**8 weeks**
Familiarizing with Basic Operations	Familiarizing with Control and Surface Techniques	Familiarizing with Control and Surface Techniques Shape Divergence and Convergence (Iteration)	Understanding Detailed Component Segmentation and Connectivity Surface Requirements for Engineering Applications

Fig. 1. The time allocation and teaching objectives of each course module in this study.

Technological Tools Used. Alias Software is utilized for precise design reconstruction, using the 2024 version with the following hardware specifications: the computer hardware is ASUS M90 workstation (Intel Core i7–12700, 8GB DDR5, Integrated Intel HD Graphics, 1TB + 256GB M.2 PCIe SSD). The operating System is Windows 11 Pro.

The VR Modeling provides an immersive modeling environment, utilizing Meta Quest 2 headsets (one per student) along with Gravity Sketch software. AI Tools include Bing Copilot and Vizcom, used to generate creative form concepts.

Since Alias follows the Learning by Teaching (Sharestart) method, all course materials and instructional videos are uploaded to a cloud drive in advance, allowing students to access them and practice at home. For VR learning, during the first week of software and hardware setup, four instructional videos (Lee & Peng 2023) covering different progress stages are preloaded onto the students' VR devices, enabling them to take the VR equipment home for self-practice. For AI learning, students are required to research AI image generation online before the module begins. The first lesson of the AI module includes a demonstration on how to generate and modify images using two AI tools, utilizing both text and image inputs.

3.2 Participants and Tasks

Participants. At the beginning of this study, 24 students from the Industrial Design Department, including second- and third-year students, enrolled in the elective course. By the end of the course, 12 students (8 s-year and 4 third-year students) completed the course and submitted their assignments. The participants' ages ranged from 19 to 21 years old, with a gender distribution of 58% male and 42% female. All participants possessed basic 3D modeling skills, but none had prior experience in transportation design.

The course enrollment criteria required students to have completed basic design courses and be familiar with fundamental 3D modeling software operations. Additionally, students needed to demonstrate an interest in surface modeling and form design. During the first class session, students were informed of the teaching plan and course structure. Enrollment was contingent on students' consent to participate in all activities, including classroom learning, model design, and questionnaire surveys.

Experimental Tasks. Participants were required to complete assigned tasks during classroom activities and self-practice sessions over four weeks. The learning activities and tasks included: 1.Using VR tools to learn curve and surface modeling techniques – Creating vehicle body components such as headlight curves and surfaces. 2.Generating biomimetic automotive forms using Bing Copilot and Vizcom – Designing a shark-inspired sports car and integrating the generated images into a VR environment to complete rapid modeling using Gravity Sketch. 3. Completing 3D modeling of automotive design using Alias software – Enhancing understanding of the modeling process and refining the final 3D vehicle form.

3.3 Training Procedure and Materials

The Alias foundational learning process (as shown in Fig. 2) spans four weeks, with the weekly learning progress as follows:

Week 1: Course introduction and instructional video overview.

Week 2: Object selection and transformation.

Week 3: Object scaling and deformation.
Week 4: Proficiency assessment – integrated application.

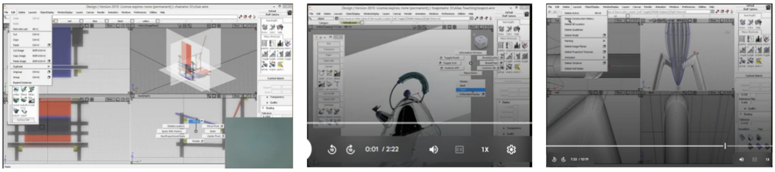


Fig. 2. The instructional videos for Alias fundamentals.

The VR learning process is divided into the following four stages:

Week 1: Introduction to VR hardware and software operations, helping students familiarize themselves with the modeling environment.

Week 2: Teaching curve modeling techniques, including curve generation, deformation, and creative applications.

Week 3: Learning surface modeling techniques, focusing on form construction and detail refinement (Fig. 3 shows a classroom example).

Week 4: Integrating all acquired skills—using Bing Copilot to generate design concepts and modeling in a VR environment (Fig. 4 shows a classroom example).

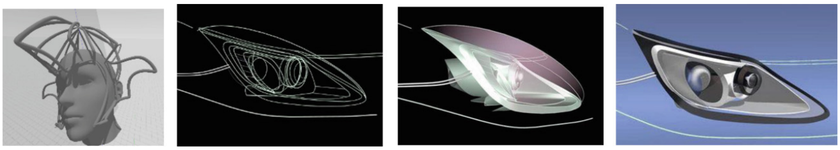


Fig. 3. The instructional example of VR modeling.

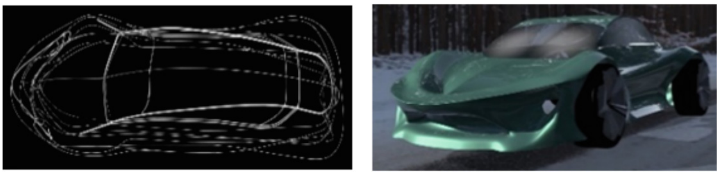


Fig. 4. The classroom example – drawing curves and surfaces in VR based on AI images.

The AI image generation learning process focuses on understanding the differences between various adjectives and nouns.

Week 1: Generating design concepts using text descriptions (Copilot).

Week 2: Refining forms through hand-drawn sketches (Vizcom).

During the class, students are asked to share their experiences regarding how different text inputs influence the generated images. Figure 5 shows the examples in classroom

The Alias automotive modeling classroom example is shown in Fig. 6, with the weekly progress as follows:

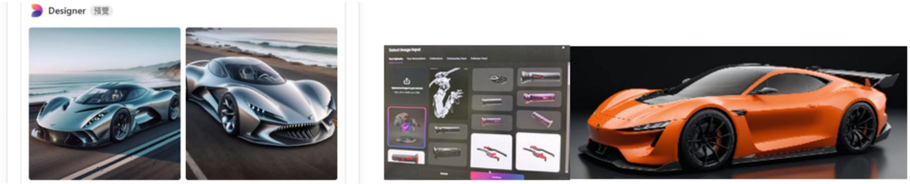


Fig. 5. AI image generation of a streamlined shark-inspired sports car using Copilot (left) and Vizcom (right).

1. Sketching the main body curves of the car.
2. Constructing the cockpit and basic surfaces.
3. Building the car body surfaces.
4. Segmenting components – modeling headlights and doors.
5. Designing side mirrors and wheel rims.
6. Refining design details, including blending surfaces.
7. Finalizing headlight segmentation.
8. Surface stitching and integration.

During this phase, course handouts are provided to students, guiding them through step-by-step construction. Students are encouraged to discuss learning challenges and difficulties with instructors and peers in class.

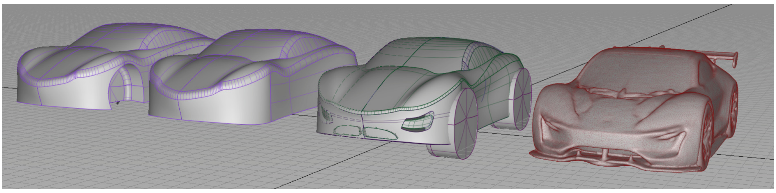


Fig. 6. The classroom example – Alias automotive modeling.

After completing the Learning by Teaching (Sharestart) course, students are required to reconstruct their designs using Alias software with NURBS modeling techniques, ultimately generating a transportation vehicle model. Upon course completion, participants must complete the SUS questionnaire and retrospective interviews, providing both quantitative and qualitative data.

3.4 System Usability Evaluation

This study employs the System Usability Scale (SUS) to evaluate the usability of the tools, including VR tools, Alias software, and Bing Copilot. The SUS questionnaire consists of 10 items, assessing both positive and negative aspects of each tool. The questionnaire adopts a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). The questionnaire is administered at the end of the course, with all participants providing separate responses for each of the three tools. Scoring method:

Odd-numbered questions: Response score – 1

Even-numbered questions: 5 - Response score

SUS score calculation: Sum the scores of all 10 questions and multiply by 2.5 to obtain the final score (ranging from 0 to 100).

Score aggregation: Compute the SUS scores for each participant and calculate the mean and standard deviation for the three tools.

Result comparison: Compare scores with the SUS benchmark score of 68:

SUS $>$ 68: The tool demonstrates good usability.

SUS $<$ 68: The tool requires improvement.

It is important to note that SUS only provides an overall perception of system usability and does not account for the impact of different user backgrounds on the ratings. Future studies should incorporate a broader range of usage scenarios for a more comprehensive analysis.

4 Student Learning Outcomes

The Learning by Teaching (Sharestart) method emphasizes students' self-learning abilities. As this course was an elective, students were not required to meet specific learning outcomes. Instead, the first ten weeks of foundational training were designed to help students quickly gain a sense of achievement using technological tools while also recognizing the challenges they might face in advanced surface modeling. Overall, during the initial phase of the foundational course, students who lacked interest in automotive design or found Alias too difficult had already withdrawn from the course. By the final stage, the remaining 12 students successfully completed the AI-generated shark-inspired automotive concept and refined their designs before transitioning to VR-based modeling. Finally, they constructed the model using Alias curve and surface tools. Table 1 presents examples of final projects from two second-year and two third-year students.

4.1 Quantitative Evaluation Results

This study utilized the System Usability Scale (SUS) to evaluate the usability of Alias, VR tools, and AI tools. The results are as follows:

Alias had an average score of 42.92 (Grade: F).

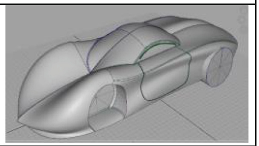
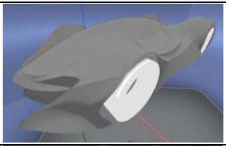
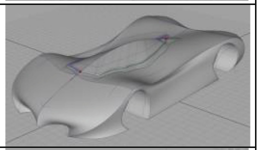
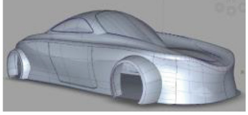
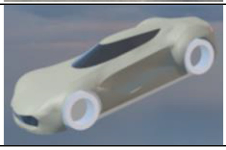
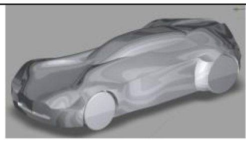
VR tools had an average score of 63.33 (Grade: D).

AI tools had an average score of 73.75 (Grade: B).

For Alias Tool: The score was significantly lower than the SUS benchmark of 68, reflecting the usability challenges associated with traditional NURBS modeling tools. Students may struggle with complex functionality and high learning difficulty, especially when their fundamental modeling skills are not yet fully developed. The students' rating and distribution chart were shown in Fig. 7.

For VR Tool: The score was close to 68, indicating that the learning experience generally meets usability standards but still requires improvements to enhance ease of use. The immersive environment of VR may help students better understand design concepts, but further optimization is needed to improve operational convenience.

Table 1. The final projects from four students (2 s year and 2 third year).

	AI Image generation	VR modeling	Alias modeling
S 1			
S 2			
S 3			
S 4			

For AI Tool: With a score exceeding the SUS benchmark (73.75) and a B rating, the results indicate good overall usability. AI enables rapid generation of creative forms and reduces modeling time, but its capability for fine-tuning details may still require improvement.

Rating	Alias	VR	AI
A	0	1	2
B	0	6	6
C	0	0	0
D	3	2	3
F	9	3	1
Total students	12	12	12

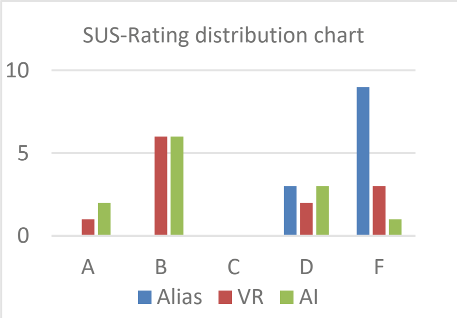


Fig. 7. The students' SUS rating and distribution chart.

The result shows that comparing to traditional Alias modeling tool, the VR tool provides a basic level of acceptable learning support, however, future improvements should focus on enhancing operational convenience and feature completeness. On the other hand, the AI tool demonstrated good usability and design potential making it one of the most suitable tools for supporting industrial design education.

4.2 Qualitative Analysis

The qualitative component of this study aims to gain deeper insights into students' experiences using the three tools (Alias, VR, and AI) through retrospective interviews. The goal is to explore the advantages, limitations, and students' learning experiences related to these tools in an educational context. This qualitative data will complement the quantitative results, providing a more comprehensive analysis and guiding future curriculum design.

The interview questions were designed around the following themes:

- Learning experience with each tool:

  What was your overall learning experience this semester (across all subjects)?
  Which course or activity required the most time?
  In which subject or skill did you learn the most?
  What was your learning experience with Alias, VR, and AI this semester?

- Teaching suggestions:

Suggestions for teaching and learning – How do you think Alias, VR, and AI should be taught?

What were the most challenging aspects of learning Alias, VR, and AI?
Should more or less content be taught?
Should students be required to bring their own laptops for class?

The Interview Results and Key Findings

1. Overall Learning Experience This Semester: Which Course or Activity Required the Most Time? What Subject or Skill Did You Learn the Most?

Regarding the first question, students generally reported that product design was the course in which they spent the most time this semester, with Alias modeling consuming the largest portion of their study time. This is primarily due to the steep learning curve of Alias software and its different operational logic compared to other modeling tools, such as Blender and Rhino. Some students also mentioned that they used VR modeling tools as a supplementary method for shape evaluation, while AI tools played a crucial role in creative ideation and product concept visualization.

2. Learning Experience with Alias, VR, and AI This Semester

The feedback on Alias software was polarized. Some students appreciated its powerful modeling capabilities, while many beginners found the basic operations unintuitive, leading to a high learning curve. Additionally, some students mentioned that Alias operates differently from other 3D modeling software such as Rhino and Fusion 360, which caused confusion in the learning process.

- Positive Feedback for Alias:

"Alias offers many unique modeling features and commands that other 3D software lacks. It is very useful for advanced applications."

"Alias is the only modeling software I know how to use, so I cannot compare it with others, but overall, it provides a comprehensive set of professional tools."

- Negative Feedback for Alias:

"The initial learning process was very frustrating because the basic operations are not intuitive. Many commands require memorizing execution sequences, which can be confusing."

"It is easy to mix up left, right, and middle mouse button operations. More time is needed to get used to the controls."

"The download and installation process were somewhat difficult. Additionally, compared to other modeling software, Alias lacks a command bar, making it unclear which objects should be selected before executing specific commands."

Most students found that VR tools performed well in rapid concept modeling and proportion adjustments, providing a more intuitive 3D modeling experience. However, prolonged use of VR caused discomfort, especially for those who wear glasses, making the experience less user-friendly.

- Positive feedback for VR:

"VR is very useful, especially in the early stages of form development. It is much more intuitive and faster compared to traditional modeling software."

"It is an excellent tool for quickly shaping proposals, easy to use, and allows for a clear, direct visualization of model proportions."

"It was quite fun and significantly helped with creative ideation."

- **Negative feedback for VR:**

"Using VR for extended periods causes eye strain, making it unsuitable for long modeling sessions."

"For users with nearsightedness or those who wear glasses, the experience is less friendly, and prolonged use can cause discomfort."

"VR lacks numerical input for moving objects, making it difficult to quantify precise dimensions, which reduces the modeling accuracy and reference value."

Students widely recognized AI tools as highly beneficial for product concept ideation and aesthetic design proposals. The biggest advantage of AI was its ability to generate creative ideas rapidly, with minimal learning effort required. However, some students reported limitations in the free versions of AI tools, stating that results were sometimes unpredictable, requiring multiple iterations to fine-tune outcomes.

- **Positive feedback:**

"AI tools are very useful for presenting product concepts and can generate interesting forms for aesthetic exploration."

"The efficiency of generating quick visual proposals is extremely high, and the learning cost is low. It is easy to explore and establish an effective workflow."

"It significantly helps with product ideation, offering many novel design inspirations."

- **Negative feedback:**

"The free version of AI tools is sometimes unstable, and the results do not always align with expectations."

"It requires multiple refinements and adjustments to obtain results that better match specific design needs."

3. Challenges in learning Alias/VR/AI: Should more or less be taught? should students be required to bring their own laptops?

According to interview feedback, students encountered multiple challenges when learning Alias modeling software, including difficulties with interface navigation, understanding commands and modeling workflows, and applying learned skills in practical tasks. Additionally, some students mentioned that the software's English-only interface and the difficulty in correcting mistakes contributed to a steep learning curve. Student Feedback including:

"The interface is difficult to navigate, and it takes time to get used to the controls."

"Alias is entirely in English, making it harder to remember commands. More supplementary materials would help."

"Thinking through the modeling sequence is challenging and requires additional time to grasp."

"When modeling a car, ensuring no gaps in the surfaces is difficult. Sometimes, I can't even tell where the errors are."

"At first, I didn't understand the meaning of some Alias annotations like P, C, and T, but later, I figured them out."

Some students also suggested that Alias instruction should focus more on practical applications, as simply listening to lectures or watching tutorial videos was insufficient for mastering the software:

"The hardest part is applying the tools and commands in real projects. Just hearing about them or watching tutorials doesn't help much if you don't use them hands-on."

Should More or Less Time Be Allocated for Alias Training?

Regarding the allocation of class time for Alias training, most students felt that more time should be dedicated to teaching, especially for beginner guidance and advanced applications. Some suggested that a structured learning schedule could help students systematically adapt to Alias's modeling processes. Students supporting more teaching time:

"The learning curve for Alias is steep. We need more lessons—absolutely necessary!"

"A structured weekly learning schedule would help reduce the risk of falling behind."

"Some Alias commands are not intuitive and difficult to learn at first. More learning time would make adaptation easier."

However, some students who found the current teaching time sufficient:

"The current class schedule is just right. For building a full car model, the lessons already cover enough practical features."

Regarding to students' selves prepare laptops their own. Students expressed mixed opinions on whether bringing personal laptops (NBs) should be mandatory for learning Alias. Some students supported the requirement, arguing that it would allow for more flexible practice outside of class, while others were concerned about resource fairness, noting that the school's provided computers were already sufficient. Students Supporting the Requirement to Bring Personal Laptops:

"Alias is difficult to master. Most students already bring their laptops to class, so making it mandatory would be reasonable."

"Having my own laptop allows me to practice more flexibly, which improves my learning efficiency."

However, some students against the requirement to bring personal laptops:

"It shouldn't be mandatory since some students may not have suitable laptops, which would create fairness issues."

"The school's computer lab (Room 327) is well-equipped. There's no need for personal laptops."

Challenges in Learning Alias, VR, and AI. Students reported that Alias has a steep learning curve, requiring more instructional support and extended practice time. Common challenges included difficulty in mastering commands, remembering functions after periods of non-use, and differences in interface versions across software updates. Students' feedback including:

"Alias is difficult to grasp and may require teachers to pay closer attention to students' progress."

"Alias takes time to learn. If I don't use it for a while, I forget some commands or actions."

"I feel like there isn't enough class time to teach Alias. If students could be more engaged or have extended practice sessions, it would be more helpful."

"Updating tutorial videos regularly would be helpful, as older versions have different interface layouts and terminology."

"Alias needs more practice time. Either extend class hours or offer additional practice sessions."

"Video tutorials for Alias are very helpful. I suggest adding more case studies to allow us to practice with real examples."

Regarding to the challenges in VR learning. Students generally had a more positive experience with VR compared to Alias, as it was perceived as more intuitive with a lower learning barrier. However, they still encountered challenges, which can be categorized into three key areas:

1. Adapting to VR Equipment - Some students experienced mild VR motion sickness and required time to adjust.

"I felt slightly dizzy at first, but after some time, I got used to it."

"VR is a new technology for me, so I had to completely readjust my learning process."

2. Teaching Approach for VR - Students saw VR primarily as a tool for quick visualization tool rather than for deep learning about detailed modeling. They preferred more practice time over excessive theoretical instruction.

"VR is fine. If it's just for quick design visualization, there's no need for extensive insttion."

"The current balance of VR teaching is good - fewer commands, easier to grasp than Alias."

3. Accessibility and Equipment Use - Most students noted that VR headsets are expensive, making it unreasonable to require students to bring their own.

"VR headsets are expensive. It's best to use the ones provided by the instructor."

"If I had to buy my own VR headset, I would drop this course. The cost barrier is too high."

Regarding to the challenges in AI learning. Students generally found AI tools intuitive and easy to learn, with a lower entry barrier compared to Alias. However, key challenges revolved around controlling AI-generated results, making adjustments, and mastering tool-specific functionalities (e.g., Vizcom).

1. Controllability of AI Tools - Although AI tools can generate creative concepts quickly, students struggled with controlling output precision and predictability, such as:

"AI isn't difficult, but the results can be unpredictable."

"More guidance on how to fine-tune AI outputs would help students learn better prompt techniques."

2. Specific Operational Challenges in AI Tools - Some students found Vizcom more complex to use, particularly when refining generated images.

"With AI, the main challenge is knowing how to refine outputs in Vizcom, but it's not too difficult."

"Rendering requires repeated adjustments. More instructions on detailed settings would be helpful."

3. Learning Approach: Emphasis on Exploration - Most students preferred a self-directed, exploratory learning style for AI, with teachers providing occasional guidance rather than rigid instruction.

"AI learning should focus on self-exploration, with teachers offering tips rather than fixed methods. This makes it easier to apply in real-world design work."

5 Teaching Feedback

1. Teaching and learning suggestions – How should Alias/VR/AI be taught?

Most interviewed students expressed satisfaction with the current teaching approach for Alias, VR, and AI. They found that a combination of instructional videos and in-class guidance effectively supported their learning. However, due to the steep learning curve of Alias, some students requested additional learning resources and extended practice time to ensure familiarity with the tool. Students' feedback including:

"The current teaching approach is great."

"VR and AI teaching methods are well-structured, allowing us to learn and apply them effectively."

"Instructional videos for Alias are very helpful since they can be replayed for detailed review. I have no particular concerns about VR and AI."

"I think the instructor provides sufficient instructional videos, making it easy to follow along."

Although most students were satisfied with the teaching approach, Alias remained the most challenging tool to learn. Some students found its commands complex and required a longer adaptation period. Additionally, since Alias is difficult to retain after periods of non-use, students suggested longer practice sessions or additional resources to reinforce their learning.

2. Should more AI content be taught?

The interviews revealed that students generally supported the idea of increasing AI instruction, but they preferred an exploratory learning approach rather than overly detailed instruction.

- Students supporting more AI instruction:

"AI tools should be taught more—this is an essential skill for students!"

"The AI learning content is well-structured and can be applied to other courses as well."

"It would be great to have more guidance on how to fine-tune AI results by adjusting input techniques."

- Students who found the current AI teaching sufficient:

"AI learning is not particularly difficult, and the current amount of instruction is just right."

"The current teaching method is good—students can explore AI tools further through assignments."

Regarding to students selves prepared their own laptops for AI learning. Most students did not object to bringing their own laptops, as AI tools primarily rely on software-based processing rather than high-cost hardware like VR headsets.

"AI learning can be done on personal laptops. Either school-provided computers or personal devices are fine."

"It doesn't matter which computer we use, as long as it can run AI tools."

"Requiring personal laptops would make AI tools more accessible for students."\

However, a few students suggested that the school should provide equipment to ensure equal access for all students.

3. Should more VR content be taught?

Students had mixed opinions on the current balance of VR instruction. While most students found the existing curriculum adequate, others expressed a desire for more in-depth training to improve learning outcomes.

- Students supporting more VR instruction:

"It would be helpful to have additional VR training, especially if the school can provide equipment."

"It's sometimes difficult to keep up with the instructor. More time for hands-on learning would be beneficial."

- Students who found the current VR teaching sufficient:

"The current VR teaching ratio feels about right."

"VR is primarily for visualizing shapes, so excessive instruction might not be necessary."

Regarding whether students should be required to bring their own VR headsets, opinions were nearly unanimous. Most students strongly opposed this requirement, citing the high cost of VR hardware as a major barrier.

"Self-providing VR devices would be too difficult—thankfully, the instructor provides them."

"VR equipment should be supplied by the school; otherwise, some students may be unable to participate."

"VR headsets are expensive, and requiring students to bring their own would make the course inaccessible to many."

6 Conclusion and Recommendations

This study serves as an exploratory teaching practice focusing on the integration of AI, VR, and traditional 3D modeling tools in industrial design education. With the instructor's years of experience in teaching Alias software, along with pre-existing instructional videos for Alias and VR, students were informed about the course structure before enrollment. The modular teaching approach allowed students to engage with multiple technology-assisted design tools.

The results indicate that all students who remained in the course successfully completed the assignments for each module, demonstrating that the Learning by Teaching (Sharestart) method is well-suited for students with a strong motivation and interest in learning. Those who were unable to adapt to this teaching method or found it challenging to keep up with the coursework had withdrawn from the course before midterm. While the aesthetic quality of students' vehicle designs still has room for improvement, the primary objective of this study was to evaluate students' acceptance of new technological tools and their adaptability to a self-directed learning method. From this perspective, the study successfully achieved its initial goals.

VR Modeling. In addition to traditional NURBS-based 3D modeling tools like Alias, future courses could incorporate AI-generated visualization techniques (e.g., AI image generation), Sub D modeling, or VR-based 3D modeling. Providing students with a broader selection of modeling techniques would allow them to explore different workflows and toolsets that align with contemporary design industry practices. However, the major challenges faced by both students and instructors in adopting VR technology is hardware investment. Due to the hype surrounding VR has diminished, and while VR is highly beneficial for large-scale product design (e.g., furniture and transportation design), the cost of acquiring VR hardware remains a burden for students. Although schools can invest in VR equipment, the management and maintenance of these devices pose additional challenges for instructors and institutions.

AI integration in immersive design environments could be future research direction. Future studies could explore the integration of AI tools within immersive environments such as VR or MR (Mixed Reality). By adopting both qualitative and quantitative evaluation methods, researchers can assess the effectiveness of AI-generated designs within immersive or hybrid spatial computing environments. Investigating AI-assisted generative design within VR/MR could provide valuable insights into how emerging technologies influence industrial design workflows and creativity.

References

Agkathidis, A., Gutierrez, R.U.: Incorporating digital tools with ceramic crafting: design and fabrication of light diffusing screen shells. J. Int. Assoc. Shell Spatial Struct. (2016)

Berg, L.P., Vance, J.M.: Industry use of virtual reality in product design and manufacturing: a survey. Virtual Reality **21**, 1–17 (2017)

Cinar, O.E., Rafferty, K., Cutting, D., Wang, H.: AI-powered VR for enhanced learning compared to traditional methods. Electronics **13**(23), 4787 (2024)

Drogemuller, A., Matthews, B.J., Cunningham, A., Yu, R., Gu, N., Thomas, B.H.: Hey building! novel interfaces for parametric design manipulations in virtual reality. In: Proceedings of the ACM on Human-Computer Interaction, vol. 8(ISS), pp. 330–355 (2024)

Hong, M.K., Hakimi, S., Chen, Y.Y., Toyoda, H., Wu, C., Klenk, M.: Generative AI for product design: Getting the right design and the design right (2023). arXiv preprint arXiv:2306.01217

Huson, D.: Digital fabrication techniques in art/craft and designer/maker ceramics. In: NIP & Digital Fabrication Conference, vol. 22, pp. 172–175 (2006)

Jun, H., Nichol, A.: Shap-e: Generating conditional 3d implicit functions. arXiv preprint (2023)

Krish, S.: A practical generative design method. Comput. Aided Des. **43**(1), 88–100 (2011)

Lee, Y.H., Chiu, C.Y.: The impact of AI text-to-image generator on product styling design. In: Mori, H., Asahi, Y. (eds.) Human Interface and the Management of Information. HCII 2023. Lecture Notes in Computer Science, vol. 14015, pp. 502–515. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-35132-7_38

Lee, Y.H., Lin, T.H.: The feasibility study of AI image generator as shape convergent thinking tool. In: Degen, H., Ntoa, S. (eds.) Artificial Intelligence in HCI. HCII 2023. Lecture Notes in Computer Science(), vol. 14050, pp. 575–589. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-35891-3_36

Lee, Y.H., Peng, S.W.: Building VR learning material as scaffolding for design students to propose home appliances shape ideas. In: Zaphiris, P., Ioannou, A. (eds.) Learning and Collaboration Technologies. HCII 2023. Lecture Notes in Computer Science, vol. 14041, pp. 144-160. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-34550-0_10

Rios, T., Menzel, S., Sendhoff, B.: Large language and text-to-3D models for engineering design optimization. In: IEEE Symposium Series on Computational Intelligence (SSCI) (2023)

Thomas, G.: How to do your research project: A guide for students (2022)

Augmented and Mixed Reality Procedural Task Training Effectiveness and User Experience

Emily Rickel(✉) and Barbara S. Chaparro

Department of Human Factors and Behavioral Neurobiology, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA

emilyrickelphd@gmail.com, chaparb1@erau.edu

Abstract. The use of augmented reality (AR) and mixed reality (MR) technologies for training is growing due to benefits like increased immersion, safer training, and reduced costs. However, the effectiveness and user experience of AR/MR training, particularly for head-mounted displays (HMDs), remains unclear. The purpose of this study is to investigate user perceptions and retention of AR/MR training for a procedural task delivered through an HMD. This two-part study utilized a within-subjects experimental design with 30 participants to determine how instruction method (paper vs. AR vs. MR) and time of procedure recall (immediate vs. post-test vs. retention) influenced completion time, perceived task difficulty, perceived confidence in successfully completing the task, workload, user experience, and trainee reactions. Results showed notable differences between instruction methods for user experience and preference, with significantly higher user experience ratings for MR and lower preference rankings for AR. Findings also show decreased performance, increased perceived task difficulty, and decreased confidence as time since training increased, with no significant differences in these measures between instruction methods. Completion times and workload were also similar between instruction methods. This research highlights objective and subjective differences between paper-, AR-, and MR-based training experiences, offering insights into which type of training may be suited for a particular use case. Recommendations for appropriately matching training modalities and scenarios, as well as for how to successfully design AR/MR training experiences, are discussed.

Keywords: Augmented/Mixed Reality · Training · User Experience

1 Introduction

Use of extended reality (XR) technologies as training solutions is growing (Stachiw 2023). XR is an umbrella term that includes augmented reality (AR), mixed reality (MR), and virtual reality (VR). VR experiences completely immerse users in the digital environment (Marr 2021). AR and MR experiences allow users to maintain visual awareness of the real world. Digital content in AR is simply overlaid onto a user's view of their surroundings and is not responsive to physical elements (Brigham 2017). MR is characterized by greater integration of digital content within real world, allowing

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 181–199, 2025.
https://doi.org/10.1007/978-3-031-93715-6_12

digital content to acknowledge and interact with the real world (Stanney et al. 2021). For example, MR head-mounted displays (HMDs) like the Microsoft HoloLens 2 work by scanning, mapping, and superimposing the user's surroundings with virtual objects. These virtual objects can be anchored to physical landmarks, and obscured by physical objects (Microsoft 2022).

Advantages to XR-based training include reduced costs, safer training, and increased immersion. However, challenges with XR-based training, such as cognitive, perceptual, and technical limitations, can negatively impact user perceptions and training effectiveness. Current literature related to XR training effectiveness is still limited given the technology's rapid evolution, particularly for AR/MR experiences (Kaplan et al. 2021) and HMDs (Han et al. 2022). Additionally, there is a need to assess the impact of AR/MR HMD training on knowledge retention and performance over time (Daling & Schlittmeier 2024; Werrlich et al. 2017), as well as to determine which types of tasks (e.g., procedural, cognitive) are more amenable for XR training (Kaplan et al. 2021).

The purpose of this study is to investigate user perceptions and retention of AR/MR training delivered through an HMD for a procedural task. Findings intend to provide insight into the implications of adopting AR/MR HMDs for procedural task training.

2 Method

2.1 Experimental Design

This two-part study utilized a within-subjects 3 × 3 experimental design. The independent variables were time of procedure recall (immediate vs. post-test vs. retention-test) and instruction method (paper vs. AR vs. MR). Dependent variables included performance, completion time, difficulty, confidence, workload, user experience, trainee reactions, and cybersickness.

2.2 Participants

Participants were recruited from a university located in the southeastern United States and its surrounding community. To be eligible to take part in this study, participants were required to be eighteen years old or order with normal or corrected-to-normal vision and full use of both hands and arms. Prior to recruiting participants, the study protocol was reviewed and approved by an Institutional Review Board (IRB) to ensure the rights and welfare of participants were protected before, during, and after data collection. Participants who consented to participate in the first study session were compensated $10 USD. Those who consented to the second study session were compensated $20 USD.

2.3 Materials

Task. Participants were trained to complete a series of origami models using instructions delivered via paper, AR, or MR. Defined as the art of paper folding (Georgia Technical Institute of Technology, n.d.), origami has been demonstrated as a suitable task for studying the acquisition of procedural skills in prior studies (Novick & Morse 2000; Tenbrink & Taylor 2015; Wong et al. 2009; Zhao et al. 2020). Six unique models were used to prevent learning effects that may impact the study's results if the same model was used for all instruction methods. Three models were used as practice models to familiarize participants with the instructions and instruction method (i.e., paper, AR, MR), while three other models were used for training and retention assessment. The chosen models were similar in number of steps and types of folds and rated as comparable in difficulty by pilot participants. Participants completed each model using a square piece of paper measuring 8.5 by 8.5 inch.

For each instruction method, instructions were displayed to participants one step diagram at a time. Paper instructions were printed on separate cards measuring 8.5 × 5.5 inch. held together in a binder. AR instructions consisted of the same step diagrams presented virtually through the Microsoft HoloLens 2. MR instructions displayed the same virtual step diagrams through the Microsoft HoloLens 2 and were supplemented by three-dimensional virtual cues anchored to the participant's workspace. These virtual cues included arrows, dashed lines, and text that mirrored the notations presented on the paper and digital diagrams.

Participants completed origami models while seated in front of a table-top study station. The study station consisted of a grid measuring 12 × 12 inches and a quick response (QR) code. Participants were prompted to fold their origami model on the grid surface and to align their paper to the grid as shown in the step diagrams. The QR code was used in the AR and MR conditions to display the virtual content in relation to the study station (see Fig. 1).

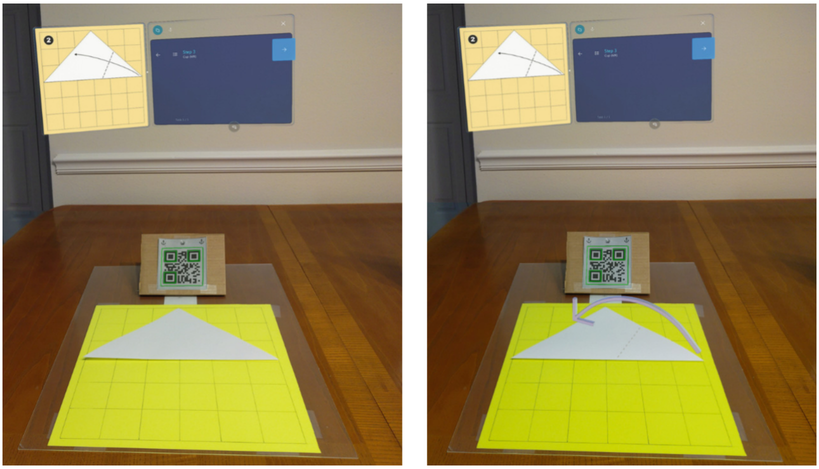


Fig. 1. MR instructions (right) displayed the same virtual step diagrams as AR instructions (left) and were supplemented by three-dimensional cues anchored to the table-top study station.

Device and Application. The Microsoft HoloLens 2 was used to display instructions presented in the AR and MR conditions. The device utilizes spatial mapping technology to construct three-dimensional models of the user's physical surroundings. Upon looking through the visor, users see virtual content displayed on top of their real-word environment. Virtual content can also be anchored to physical objects and surfaces.

The AR and MR training conditions were built using the Microsoft Dynamics 365 Guides application. This application enables the creation of instructional guides composed of text, images, videos, and three-dimensional virtual objects that can be anchored to specific locations within the user's physical surroundings in relation to a QR code displayed in a location central to the user's workspace. Participants were permitted to interact with the application using their preference of the following input methods: hand tracking, eye tracking, and voice commands.

2.4 Measures

Paper Folding Test. Following completion of a demographics questionnaire, participants completed the Paper Folding Test (PFT), a measure of the ability to manipulate or transform spatial content into other arrangements. The PFT comprises of twenty questions that participants must complete within six minutes. Each question presents a series of diagrams representing a square piece of paper being folded one to three times with a hole punched through the folded paper. Participants must choose between five diagrams what the paper and hole pattern would look like if the paper is unfolded. The PFT was scored by tallying the number of correct answers out of all twenty questions (Ekstrom et al. 1976). Higher PFT scores indicate a better spatial manipulation ability.

Performance. Scoring rubrics were created to evaluate the accuracy of the completed origami models. Each rubric was piloted multiple times with a range of low- to high-quality models to ensure the rubric was comprehensive and generalizable to a variety of model attempts. Three raters were trained to use the rubric by teaching them how to complete each model, presenting them with examples of pass/fail models for each rubric item, and rating at least five low- to high-quality models for each of the three origami patterns used for training and retention assessment. Feedback was provided on training ratings, as needed. After learning the rubrics, raters scored all participants' second training, immediate, post-test, and retention attempts. Performance scores range from 0 to 7, with higher scores indicating a more accurate completed model.

Completion Time. The amount of time required to complete each training (with instructions) and recall (without instructions) model was measured in seconds. The timer started when the participant verbally indicated they understood the task and were ready to begin. The timer was stopped when the participant verbally indicated they completed the task.

Difficulty and Confidence. Perceived difficulty was measured using a 7-point rating scale (1 = Very Difficult; 7 = Very Easy). Task difficulty was collected following each origami model attempt. Instruction method difficulty was collected following each completed instruction method condition. Perceived confidence in successfully completing the task was measured following each origami model attempt using a 7-point rating scale (1 = Not at all Confident; 7 = Very Confident). Higher ratings indicate the task was perceived to be easier with a greater sense of confidence in success.

Workload. Workload was measured using a modified version of the National Aeronautics and Space Administration Task Load Index (NASA-TLX; Hart & Staveland

1988) called the Raw TLX (RTLX; Hart 2006). The RTLX consists of six items that correspond to six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. This metric is rated on a 21-point scale and forgoes the paired comparison process utilized in the NASA-TLX. Higher RTLX ratings indicate participants perceived the task as more demanding.

User Experience. User experience was measured using the short version of the User Experience Questionnaire (UEQ-S; Schrepp et al. 2017). The UEQ-S consists of eight items providing insight into two dimensions: pragmatic quality (i.e., aspects related to the user's task or goals, such as efficiency and clarity) and hedonic quality (i.e., aspects not related to the user's task or goals, such as pleasure). This metric is rated on a 7-point semantic differential scale. Higher UEQ-S ratings indicate better user experience. Ratings can be categorized into one of five benchmarks: Excellent, Good, Above Average, Below Average, and Bad (Hinderks et al. 2018).

Trainee Reactions. Trainee reactions were collected using six questions adapted from Long et al. (2008) and three open-ended questions. The first six items captured technology satisfaction, enjoyment, and relevance of course content using a 5-point scale (1 = Strongly Disagree; 5 = Strongly Agree). Higher ratings indicate a more positive perception of the training experience. The three open-ended questions collected qualitative data regarding participants' likes and dislikes about each training method, as well as recommendations for improving each training method.

Cybersickness. Prevalence of adverse symptoms such as dizziness, nausea, and visual stress following exposure to the AR/MR HMD (i.e., cybersickness) was measured at the end of the first study session using the Simulator Sickness Questionnaire (SSQ; Kennedy et al. 1993). The SSQ consists of sixteen symptoms rated on a 4-point scale (0 = None; 3 = Severe) to indicate symptom severity across three subscales: nausea, oculomotor, and disorientation. Ratings can be categorized into one of six benchmarks: 0 – No Symptoms; <5 - Negligible Symptoms; 5–10 - Minimal Symptoms; 10–15 - Significant Symptoms; 15–20 - Concerning Symptoms; and >20 - Bad (Stanney et al., 1997).

Post- and Retention-Test Open-Ended Questions. Participants were asked a series of open-ended questions to collect their perceptions of the instruction methods and strategies they used to perform the procedure at the end of the first (post-test) and second (retention-test) study sessions. Post- and retention-test questions included preference rankings and explanations of using paper, AR, and MR instructions from most to least preferred, as well as stating whether they would choose AR or MR instructions and why. Additional questions posed after the retention-test included an explanation of strategies used to recall the procedure for each origami model, as well as whether the participant believed the method of instruction affected their ability to recall each model.

2.5 Procedure

After receiving an overview of the study and providing their consent to participate, participants filled out a demographic questionnaire, completed the PFT, and learned how to interpret origami diagrams. Before putting on the HoloLens 2, participants were given an overview on how to adjust the headset fit and brightness of virtual content. Once the device was fitted comfortably, participants completed the device's eye calibration procedure and learned how to interact with virtual content using gestures and voice commands.

The three instruction method conditions were counterbalanced across all participants. For each instruction method, participants completed a practice model. Participants were prompted to fold using "hard" creases and to make sure their model aligned with the grid as shown in the diagram before moving onto the next step. After the practice model, participants were introduced to the model used for training and retention assessment. This model was completed twice with the instructions and an example model available for reference before completing the model a third time without access to the instructions or example model. Participants were given the final diagram of each completed model and allowed to reference it during all recall attempts. The researcher measured completion time for all training and recall models. Perceived task difficulty and confidence were collected after participants completed each training and recall model. Following each instruction condition, participants completed the perceived instruction method difficulty, workload, user experience, and trainee reaction questions. After completing all three instruction conditions, participants folded each recall model again to assess their post-task performance. Time to completion, perceived task difficulty, and confidence were collected again following completion of each post-task model. At the end of the first study session, participants responded to the post-test open-ended questions and SSQ, and received their first compensation payment.

Participants who completed the first study session and were interested in participating in the second study session were scheduled to return one week later. During the second study session, participants folded each recall model. Time to completion, perceived task difficulty, and confidence were collected following completion of each retention-test model. Participants were also asked to respond to the retention-test open-ended questions. Finally, participants were debriefed and received their second and final compensation payment. Each participant took approximately 2.5 h to complete the study (two hours for the first study session, 30 min for the second study session).

3 Results

Analyses of variance (ANOVAs) were conducted to determine the effect of instruction method and time of procedure recall on the dependent variables. Unless otherwise stated, outliers identified as data points greater than ± 3 standard deviations from the mean are included in the following results because including these outliers did not substantially impact the interpretation of the ANOVA outputs. Satisfaction of the assumption of homogeneity of variance was determined using Mauchly's test of sphericity. Results in which Mauchly's test of sphericity was significant, indicating violation of the assumption, are reported using the Greenhouse-Geisser correction.

3.1 Demographics

Thirty participants (13 male, 17 female) completed this study. Participant ages ranged from 18 to 37 years ($Mdn = 21.5$, $IQR = 6$). Eight participants wore prescription glasses under the AR/MR HMD during training sessions. Three participants reported being left-handed. Twenty-three participants indicated prior use of XR HMDs, with eight reporting they owned an XR HMD. Among those having prior experience with XR HMDs, 11 participants reported using AR/MR HMDs for at least one hour and 14 reported using VR HMDs for at least one hour. Five participants reported prior experience with the Microsoft HoloLens 2.

The average number of correct responses on the PFT measuring spatial manipulation ability was 14.13 ($SE = 0.61$). These results are comparable to prior studies assessing college students ($M = 13.8$, $SD = 4.5$; Ekstrom et al., 1976) and the general population ($M = 12.7$, $SD = 3.5$; Burte et al. 2018). A Pearson correlation was conducted to investigate the relationship between PFT scores and performance. There were statistically significant, moderate positive correlations between PFT scores and paper performance ($r(28) = .60, p < .005$), between PFT scores and AR performance ($r(28) = .45, p = .012$), and between PFT scores and MR performance ($r(28) = .43$, $p = .018$). These results indicate that as spatial manipulation ability increased, performance for all instruction methods also increased.

3.2 Performance

Three raters used scoring rubrics to evaluate the accuracy of participants' origami model attempts. Percent agreement ranged from 75% to 85%. Cronbach's alpha coefficients exceeded the acceptable threshold of .70, indicating acceptable inter-rater reliability.

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on performance. Overall performance scores can range from 0 to 7, with higher scores indicating more accurate models. The main effect of recall time showed a significant difference in performance across recall time periods, $F(1.47, 42.71) = 25.05$, $p < .005$, partial $\eta^2 = .46$. Retention models ($M = 3.07$, $SE = 0.37$) were significantly less accurate than the immediate ($M = 4.97$, $SE = 0.15$) and post models ($M = 4.25$, $SE = 0.26$). Post models were also significantly less accurate than the immediate models (see Fig. 2). There was no statistically significant main effect of instruction method, $F(2, 58) = 0.63$, $p = .538$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 1.24$, $p = .297$.

3.3 Completion Time

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on completion time, measured in seconds. An assessment of studentized residuals greater than ±3 standard deviations identified three outliers. These outliers were replaced with the average time for all other completion times of that condition. The main effect of time of procedure recall showed a significant difference in completion time across recall time periods, $F(1.52, 43.93) = 4.39$, $p = .027$, partial $\eta^2 = .13$. Post-hoc analyses did not identify significant differences between

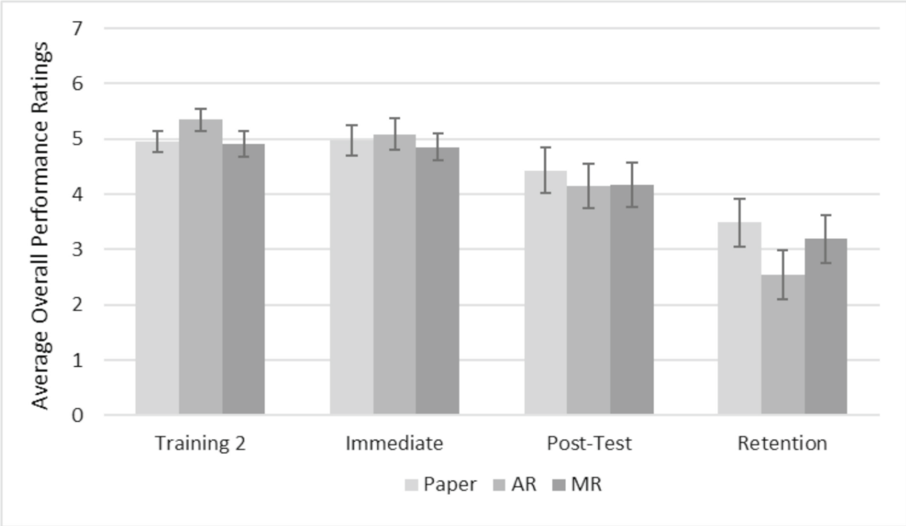


Fig. 2. Comparison of performance between instruction methods over time. Error bars represent ± 1 standard error.

immediate (M = 147.03, SE = 8.67), post-test (M = 149.24, SE = 7.95), or retention (M = 174.49, SE = 12.12) completion times. There was no statistically significant main effect of instruction method, $F(2, 58) = 0.07$, $p = .934$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 0.94$, $p = .442$.

An exploratory two-way repeated measures ANOVA was conducted to investigate differences in completion time between the two training sessions. The main effect of recall time showed a significant difference in completion time between training sessions, $F(1, 29) = 36.41$, $p < .005$, partial $\eta^2 = .56$. Training 1 (M = 220.62, SE = 11.64) took significantly longer to complete than training 2 (M = 180.51, SE = 10.36). There was no statistically significant main effect of instruction method, $F(2, 58) = 1.53$, $p = .226$, or interaction between instruction method and time of procedure recall, $F(2, 58) = 0.58$, $p = .561$.

3.4 Difficulty

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on perceived instruction method difficulty. Perceived instruction method difficulty was self-reported by participants using a 7-point scale (1 = Very Difficult to 7 = Very Easy) after completing each instruction method during Session 1. There was no statistically significant difference of perceived difficulty ratings between instruction method, $F(2, 58) = 0.76$, $p = .473$.

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on perceived task difficulty. Perceived task difficulty was self-reported by participants using a 7-point scale (1 = Very Difficult to 7 = Very Easy) after each model attempt. The main effect of recall time showed a significant difference in perceived task difficulty across recall time periods, $F(1.59, 45.97) = 31.55$, $p < .005$, partial $\eta^2 = .52$. Retention models ($M = 3.52$, $SE = 0.30$) were perceived to be significantly more difficult than immediate ($M = 5.41$, $SE = 0.16$) and post-test models ($M = 4.66$, $SE = 0.23$). Post-test models were perceived to be significantly more difficult than immediate models (see Fig. 3). There was no statistically significant main effect of instruction method, $F(2, 58) = 0.07$, $p = .932$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 0.76$, $p = .555$.

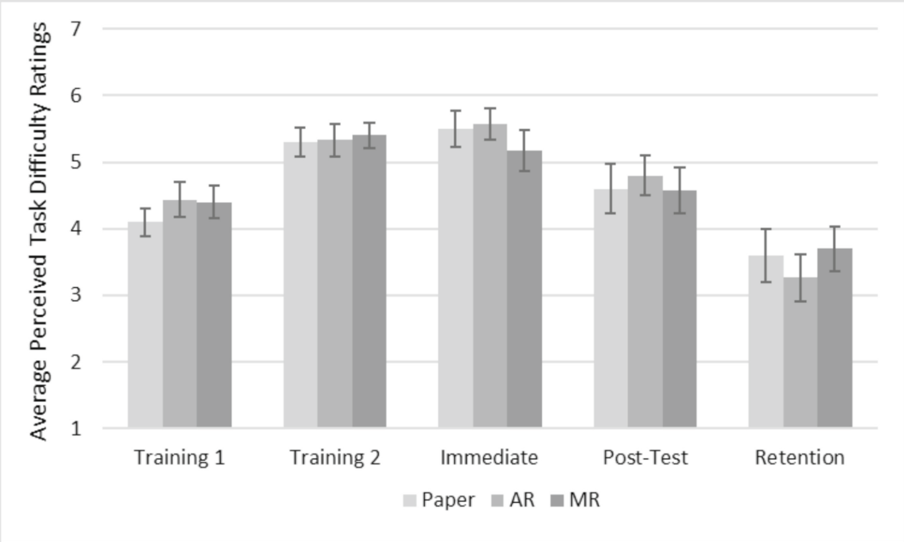


Fig. 3. Comparison of perceived task difficulty between instruction methods over time. Error bars represent ±1 standard error. 1 = Very Difficult; 7 = Very Easy.

3.5 Confidence

A two-way repeated measures ANOVA was conducted to determine the effect of instruction method and time of procedure recall on perceived confidence. Perceived confidence was self-reported by participants using a 7-point scale (1 = Not at all Confident to 7 = Very Confident) after each model attempt. The main effect of recall time showed a significant difference in perceived confidence across recall time periods, $F(1.61, 46.71) = 33.40$, $p < .005$, partial $\eta^2 = .54$. Participants were significantly less confident about their retention model success ($M = 3.60$, $SE = 0.35$) compared to their immediate ($M = 5.71$, $SE = 0.20$) and post-test models ($M = 5.00$, $SE = 0.28$). Participants were also significantly less confident about their post-test models compared to their immediate models. There was no statistically significant main effect of instruction method, $F(1.64, 47.42) = 0.19$, $p = .782$, or interaction between instruction method and time of procedure recall, $F(4, 116) = 1.10$, $p = .358$.

3.6 Workload

A series of one-way repeated measures ANOVAs was conducted to determine whether raw ratings of each NASA-TLX dimension differed between instruction methods. There were no statistically significant differences in workload between instruction methods (see Fig. 4): Mental, $F(1.44, 41.68) = 1.64$, $p = .210$; Physical, $F(2, 58) = 0.07$, $p = .932$; Temporal, $F(2, 58) = 3.15$, $p = .050$, partial $\eta^2 = .10$; Performance, $F(2, 58) = 0.27$, $p = .765$; Effort, $F(2, 58) = 3.00$, $p = .057$, partial $\eta^2 = .09$; and Frustration, $F(1.63, 47.18) = 3.03$, $p = .068$, partial $\eta^2 = .10$.

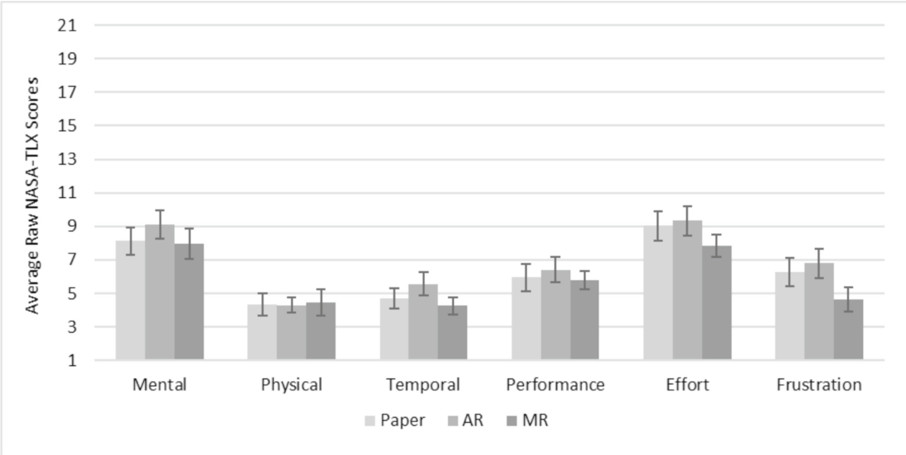


Fig. 4. Comparison of workload ratings between instruction methods. Error bars represent $\pm$ 1 standard error. Higher ratings indicate that participants perceived the task as more demanding or that they performed poorly.

3.7 User Experience

The 8-item UEQ-S was completed by participants following their exposure to each of the three instruction methods. The UEQ-S was analyzed by averaging ratings for the first four items to produce a pragmatic quality score, averaging the last four items to produce a hedonic quality score, and averaging all eight items to produce an overall user experience score. Benchmarks for each score (Excellent, Good, Above Average, Below Average, and Bad) were calculated using the analysis tool provided by the UEQ developers (Hinderks et al. 2018).

A series of one-way repeated measures ANOVAs was conducted to determine whether pragmatic quality, hedonic quality, and overall scores differed across instruction methods (see Fig. 5). There was no statistically significant difference in average pragmatic quality scores between instruction methods, $F(2,58) = 0.54$, $p = .587$. Benchmarks for pragmatic quality were Good for Paper, Above Average for AR, and Good for MR. There was a statistically significant difference in average hedonic quality scores between instruction methods, $F(1.61, 46.64) = 70.60$, $p < .005$, partial $\eta^2 = .71$. MR scores ($M = 6.23$, $SE = 0.12$) were significantly higher than AR scores ($M = 5.70$, $SE = 0.20$), which were significantly higher than Paper scores ($M = 3.57$, $SE = 0.24$). Benchmarks for hedonic quality were Bad for Paper, Excellent for AR, and Excellent for MR. There was a statistically significant difference in average overall user experience scores between instruction methods, $F(2, 58) = 20.54$, $p < .005$, partial $\eta^2 = .42$. Paper scores ($M = 4.62$, $SE = 0.17$) were significantly lower than AR ($M = 5.55$, $SE = 0.17$) and MR scores ($M = 5.91$, $SE = 0.13$). There was no statistically significant difference between AR and MR scores. Benchmarks for overall scores were Below Average for Paper, Good for AR, and Excellent for MR.

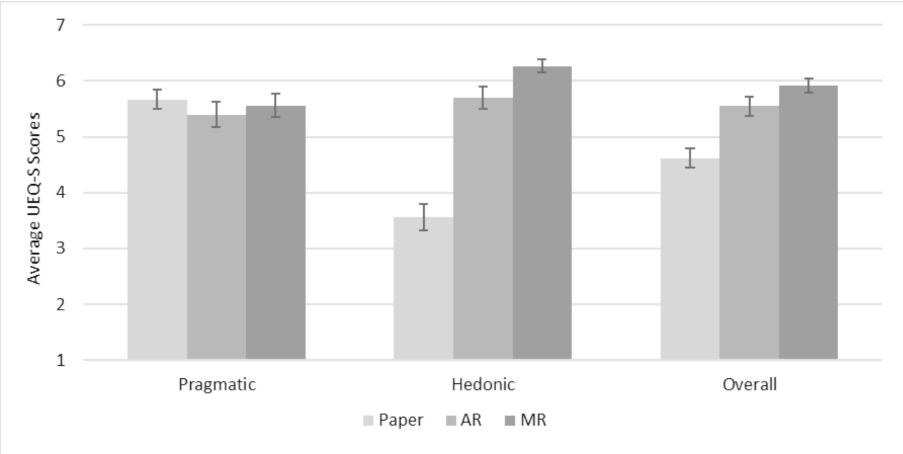


Fig. 5. Comparison of user experience ratings between instruction methods. Error bars represent ± 1 standard error.

3.8 Trainee Reactions

Trainee reactions were collected using a 6-item questionnaire adapted from Long et al. (2008) after participants completed training with each instruction method. Participants rated each item on a scale of 1 = Strongly Disagree to 5 = Strongly Agree. Item numbers referenced in the following paragraphs correspond to item numbers listed in Table 1.

A series of one-way repeated measures ANOVAs was conducted to determine whether the six individual trainee reaction items differed between instruction methods. There were significant differences between instruction methods for Item 3, $F(2, 58) = 6.39$, $p < .005$, partial $\eta^2 = .18$, and Item 4, $F(1.53, 44.21) = 5.49$, $p = .013$, partial $\eta^2 = .16$. Paper ($M = 4.97$, $SE = 0.03$) was rated significantly higher than AR ($M = 4.53$, $SE = 0.12$) and MR ($M = 4.57$, $SE = 0.10$) for Item 3, indicating that participants were better able to navigate through training content using the paper instructions. Paper ($M = 4.83$, $SE = 0.07$) was also rated significantly higher than AR ($M = 4.37$, $SE = 0.18$) and MR ($M = 4.43$, $SE = 0.09$) for Item 4, indicating that it was easier for participants

Table 1. Trainee Reaction Questionnaire Items.

Item Number	Question
1	The training content was clear
2	I could easily understand the training content
3	I was able to navigate through the training content
4	I found the training content easy to use
5	I was satisfied with the presentation of the training content
6	I had a positive learning experience

to use the paper instructions. There was no significant difference in ratings for Item 1, $F(2, 58) = .053$, $p = .590$; Item 2, $F(1.62, 46.99) = 0.79$, $p = .437$; Item 5, $F(1.53, 44.27) = 2.56$, $p = .086$; or Item 6, $F(2, 58) = 0.03$, $p = .969$.

Participants were also asked to answer three open-ended response questions regarding their likes, dislikes, and recommendations for improvement for each instruction method. To summarize participants' likes and dislikes, participants reported the paper instructions to be more familiar and reported paper easier to navigate and use compared to the AR/MR instructions. However, participants did not enjoy having to take their hands off their origami paper to use the paper instructions. They also reported liking the voice commands and eye gaze features offered by the AR/MR training that enabled hands-free interaction with the instructions. Several participants shared that the MR virtual cues (e.g., dotted lines to indicate fold placement, arrows to indicate fold direction) were the best aspect of the MR training because they promoted accuracy and reduced the need to move their head to view the training materials, but some also perceived the MR virtual cues as distracting and obstructive. Participants also noted more navigation issues with the AR/MR training compared to the paper instructions, primarily referencing issues with lag. Other AR/MR dislikes included limited field of view (FOV), increased head movement to reference AR instructions, and the visor tint that negatively impacted participants' ability to see their real-world surroundings.

When asked how to improve the paper training, thirteen participants suggested additional written instructions (e.g., "Adding words to the instructions on how and where exactly I'm supposed to make certain folds and creases would be beneficial") or visual instructions (e.g., "More pictures indicating different angles," and, "Put a small image of what it should look like when you are done with the step up in the corner"). Three participants suggesting putting all the paper instructions on one page to reduce the need to flip between steps (e.g., "It would be easier to have all the steps on one sheet, that way you can see how each step leads to the next"). Regarding MR training improvements, twelve participants suggested redesigning the virtual content to make it less obstructive, such as making the virtual cues thinner or more transparent, implementing the ability to customize the color of the virtual cues, reducing the number of virtual cues by only showing the dotted fold lines and removing the arrows, and toggling the presence and absence of the virtual cues depending on whether the user's hands are in the workspace (e.g., "Have the arrows and words that pop up on the grid disappear when your hands are there, but reappear when you move your hands away, so you can reference the instructions at any time just by staring at the grid"). Five participants suggested adding folding animations or auditory instructions to the AR/MR training conditions. Four participants recommended better alignment of the MR virtual cues to the real world.

3.9 Cybersickness

Cybersickness was measured using the SSQ, a 16-item questionnaire distributed to participants at the end of their first study session. The following benchmarks can be used to facilitate interpretation of SSQ ratings: 0 – No Symptoms; <5 Negligible Symptoms; 5–10 Minimal Symptoms; 10–15 Significant Symptoms; 15–20 Concerning Symptoms; and >20 Bad (Stanney et al., 1997). Oculomotor discomfort ($M = 17.18$, $SD = 15.28$) was the highest score among the subscales and indicated Concerning Symptoms. Disorientation ($M = 14.85$, $SD = 14.13$) and total SSQ scores ($M = 10.60$, $SD = 8.68$) were Significant, while nausea ($M = 6.04$, $SD = 7.72$) was Minimal.

3.10 Post- and Retention-Test Open-Ended Questions

At the end of the first and second sessions, participants were asked to rank their preference for instructions provided using paper, AR, or MR, from most to least preferred. Additionally, after the retention-test, participants were asked to describe the strategies they used to recall each model and how each instruction method affected their ability to recall the procedure.

Preference. Participant rankings for paper, AR, and MR were labeled from most (1) to least (3) preferred instruction method. A Friedman test was conducted to determine if there were differences in how instruction methods were ranked. There was a statistically significant difference between average ranks of the instruction methods collected at the end of the first study session, $\chi^2(2) = 6.07$, $p = .048$. MR was ranked first by 15 participants ($M = 1.80$), paper was ranked first by 9 participants ($M = 1.83$), and AR was ranked first by 6 participants ($M = 2.37$). Post-hoc analysis with Wilcoxon signed-rank tests conducted with a Bonferroni correction found a significant increase in average rank for paper compared to AR ($p = .021$). There was no significant difference between AR and MR ($p = .063$) or paper and MR ($p = .922$) average rankings. There was no significant difference between average ranks collected at the end of the second study session, $\chi^2(2) = 5.40$, $p = .067$. MR was ranked first by 16 participants ($M = 1.70$), paper was ranked first by 8 participants ($M = 2.00$), and AR was ranked first by 6 participants ($M = 2.30$). These results suggest that AR is the least preferred method of instruction.

Comments from Participants Who Preferred Paper. Participants who ranked paper instructions as their most preferred instruction method explained they were more familiar with paper instructions and felt the paper was less obtrusive and effortful. For example, one participant commented on their familiarity with paper instructions by stating, "Using paper instructions is just a habit, it is what I work with all the time. It is like having a textbook and homework laid out on a desk. I am used to that setup," while another said, "The MR was helpful, interesting, and different, but if I was learning something for the first time, I would want to stick with something I was used to using, like paper." Perceptions of AR/MR obtrusiveness primarily stemmed from the MR virtual cues, "MR was obstructive. The 3D content obstructed my view of the paper, the grid, and my hands. I could not see what I was folding." The headset hardware also contributed to its obstructiveness, "Even when I adjusted the headset, it was a little blurry to look through it. Paper is very sharp and clean." Paper was also considered to be less effortful to navigate through the training materials, "I can flip through it quickly or close it altogether if I don't want it. But with AR and MR, I have to stare at the button for a few seconds to flip the page. I can't flip multiple pages or flip very quickly in AR and MR because I have to wait for it to respond to my input." Several participants who ranked paper as their most preferred method noted that AR presented the training materials in a very similar manner, but with the added difficulty of having to learn and use a new device and application (e.g., "AR seemed unnecessary for this task and the headset started to weigh on my neck a little bit. I didn't feel like it added anything I couldn't have just gotten from the paper," and, "Paper and AR are more or less the same thing, but AR involves wearing a headset. Why wear a headset when you don't have to?").

Comments from Participants Who Preferred AR. Participants who ranked AR instructions as their most preferred instruction method stated they found AR more exciting, enjoyed hands-free navigation of the training materials, and perceived it to be less obstructive. For those who preferred AR, AR was considered more exciting than paper, "The paper instructions were easy to follow. I just preferred the AR more because it was interactive, entertaining, and interesting." Other participants shared, "I like how innovative the AR is, and I like the fact that I don't have to remove my hands from the origami to look at the instructions," and, "The AR allowed me to fold better. When I had to flip the paper pages, the model would move since I had to take my hands off of it. It was inconvenient to have to take my hands off the model." The perceived obtrusiveness of the MR virtual cues was exacerbated by the HoloLens 2 limited FOV, "AR was the easiest to work with. MR was difficult to get used to. It was weird having to look down at it because it [the virtual cues] wouldn't show up if you just glanced down, so you had to make sure [to move your head] to really look at it. And when I was folding, it would get in the way so I couldn't see if my folds were exactly straight or not."

Comments from Participants Who Preferred MR. Participants who preferred MR instructions most indicated they found MR more integrative and comprehensive, providing benefits that outweigh the disadvantages of using a headset. Several participants noted that because the MR instructions were better integrated with the participants' workspace, they experienced less cognitive workload and reduced head movement. For instance, one participant stated, "With MR, I don't have to interact with so many other things around me. I just look at the [origami] paper and fold it as I am looking at it, instead of having to look up at the AR and take time to look away from my paper to find the screen, then reorient myself to what I was doing with the paper," and another participant commented, "Because the MR placed the instructions right on my paper, I didn't have to hold the information in my head as long or keep looking back at the instructions, as I did with the paper and AR instructions." Additionally, those who preferred MR training believed it provided more information than the other instruction methods, "I felt like the MR provided the most in-depth instruction and it was easiest to follow along because it was so immersive." Other participant comments that support this notion include, "I liked that MR provided multiple sources of information. I was able to be more precise with my folds because I had so many cues to rely on," and, "MR was a lot more helpful. It took the guesswork out of it. The dotted lines and arrows really showed exactly where to fold." Participants who ranked MR first also noted that AR did not provide additional benefits that outweigh the disadvantages of using the headset. Representative participant comments for this point include, "I didn't feel like the AR contributed anything more than paper, but AR made it more cumbersome to complete the task," and, "For the AR, I feel like it wasn't providing any value. It was like using technology for technology's sake, because it was almost exactly what the paper offered."

Recall Strategies. At the end of the second study session, participants were asked to reflect upon and share strategies used to recall the trained procedures. Twenty-four participants stated they generated and utilized mental images as they recalled the models. Eight participants indicated they visualized the MR virtual cues. Of these participants, six commented this visualization strategy facilitated their ability to recall the models they learned using MR (e.g., "I was picturing the dotted lines and arrows, which was helpful."), one said they did not think it helped their recall ability, and one stated it hindered their ability to accurately recall the model (e.g., "The MR was so distracting that I could not think of the steps, only the arrows."). Other recall strategies mentioned by participants include relying on muscle memory, leveraging the diagram of the finished model provided to participants during all recall models, and using points on the grid to help them determine where to make folds. In general, recall strategies did not differ between instruction methods, with the exception of visualizing the virtual cues that were only available during MR training.

Perceptions on Whether Instruction Method Matters. At the end of the second study session, participants were also asked whether they believed the different methods of instruction impacted their ability to recall each procedure. Twenty-nine participants agreed that the instruction method influenced their recall performance. Six participants stated the AR/MR training experiences were more fun and exciting. Of these six participants, three believed this positively contributed to their ability to recall the origami models (e.g., "I think I remembered the AR and MR models better because it was more exciting to use something other than the book."), while two participants thought the fun experience negatively contributed to their recall performance (e.g., "I was more focused on trying to figure out how to use the headset. It was cool, but with the paper instructions, there was nothing else to focus on but the instructions and I think I learned better as a result."). Six participants also mentioned the MR training was more helpful than the paper and AR training, enabling them to better remember the MR models. For example, one participant supported this notion by stating, "By placing the instructions directly on the [origami] paper, the MR provided more landmarks than the AR and paper instructions. I would argue this helped me remember the MR model better," while another participant explained, "The MR instructions provided more guidance, which helped me learn the model better and led to better [memory] encoding." However, three participants stated they became reliant on the MR virtual cues, which negatively impacted their recall performance. For example, one participant shared, "There were some steps that I completely missed because I was relying on the MR but did not have the lines on the [origami] paper anymore. I was definitely relying on the MR more than the other two [instruction] methods," while another participant said, "I wonder if I relied more on the MR rather than actually learning from it. I felt a lot more lost once the MR virtual elements were gone compared to just losing the instructions with the other two methods."

Overall, almost all the participants believed their recall performance in the second study session was impacted by how they learned each origami model. More participants commented they liked the AR/MR training, but they had differing opinions as to whether it helped or hindered their ability to recall each model.

4 Discussion

This work provides insight into objective and subjective differences between paper-, AR-, and MR-based procedural training experiences. The current study found that paper, AR, and MR training modalities were comparable in regard to performance, completion time, perceived task difficulty, perceived instruction method difficulty, perceived confidence in successfully completing the task, workload, pragmatic quality, and retention recall strategies. Differences were found regarding user experience and trainee reactions, such that MR received higher user experience ratings and AR was least preferred by participants. Increases in spatial manipulation ability were correlated with increases in performance for all instruction methods.

4.1 Practical Implications

These findings can be generalized to help determine which type of training is best suited for a particular use case. Paper instructions are recommended when time available for training is limited, as trainees are more likely to be familiar with this instruction modality and therefore would not have to spend additional time learning a new way of learning. Additionally, the results of this study suggest that it is likely not worthwhile to invest in transitioning paper instructions to an AR experience if the AR experience is simply a virtual recreation of the paper experience, unless it is beneficial for trainees to have hands-free access to the training materials so they can keep their hands on task. Because MR was subjectively reported to reduce head movements and cognitive workload, it may be suitable for longer training sessions because it may minimize risk of fatigue or injury that could result from having to perform repetitive bodily movements to reference multiple, separate sources of information. However, it would be important to monitor trainees' cybersickness symptoms as they complete AR/MR training to ensure trainees are not experiencing adverse effects. MR virtual content would also have to be designed to reduce obtrusiveness and piloted to confirm it is properly aligned with the physical world. The design of virtual content presented in AR/MR training experiences should be an iterative process that involves input from an interdisciplinary team that includes instructional designers, human factors practitioners, XR developers, and representative end-users.

Organizations planning to implement AR/MR training should allocate time for trainees to acclimate to these technologies. Familiarization periods should include customizing device fit, completing eye calibration, interacting with virtual content, and learning input methods (e.g., gestures, voice commands, controllers). Trainees should also practice utilizing input methods with tasks similar to what they will be performing during the training session. Examples of such tasks include activating, deactivating, placing, moving, or resizing virtual content. Trainees should also learn how to properly hold and clean the device so it can be sanitized between users. Skipping this familiarization period and simply handing off AR/MR devices to trainees who are inexperienced with using such devices will likely have a negative impact on training effectiveness, user experience, and trainee acceptance of the technology.

4.2 Limitations

Limitations of this study include those associated with the sample, methodology, and generalizability to other XR devices and training tasks.

Sample Limitations. Because the sample consisted primarily of college students, demographics and results of the current study may not completely generalize to the general population. Results may have been impacted by individual differences not captured in this study, such as learning style, acceptance of XR technologies, motivation to learn and perform, self-efficacy, and spatial abilities not measured by the PFT.

Methodological Limitations. This study was conducted in a controlled, laboratory environment that may have limited ecological validity. Participant dropout was minimized by reducing the number of study sessions, but this resulted in participants completing their training during a single two-hour session. While participants were offered multiple opportunities to take breaks during the training session, participants could have experienced fatigue during the initial session, negatively impacting their ability to attend to the training materials or recall the procedure from memory. A single, extended training session is also not representative of all training delivery schedules. Moreover, there are limitations regarding how the dependent variables were measured in this study. For one, performance was assessed using a non-validated rubric and individual differences between the three raters who evaluated participants' completed origami models likely impacted their ratings due to some rubric items yielding more room for subjectivity. Additionally, several dependent variables were collected using self-reported measures. The accuracy of self-reported measures can be hindered by participants forgetting pertinent details or responding in a manner they believe will be viewed favorably by the researchers.

Generalizability Limitations to Other Devices and Tasks. This study only utilized one AR/MR headset. Because different headsets have different attributes (e.g., input methods, hardware features) that can impact user experience and training effectiveness, it may not be appropriate to generalize the findings of the current study to all XR devices used for training. Additionally, the results of this study may be limited to procedural tasks and may not be applicable to cognitive or affective tasks. Also, the task performed in this study was a tabletop task completed by an individual that utilized smaller equipment, an experience which may differ from tasks that require standing or walking while interacting with larger equipment. This task also utilized a limited range of MR virtual cues (e.g., dotted lines, arrows) that may not be applicable to other tasks.

4.3 Future Research

Recommendations for future research include replicating the study with a larger, more diverse sample to determine what, if any, impact individual differences may have on AR/MR training effectiveness and user experience. For example, the current study found significant relationships between spatial manipulation ability and task performance, which should be further explored in future research. The use of XR training in a real-world setting and with other training delivery schedules, such as multiple, shorter sessions, could also be investigated. It is recommended for future studies that aim to assess performance to consider how precise accuracy must be measured to determine task success. The need for more precise measures of accuracy may require the use of more objective tools. For example, a more precise measure of accuracy for the procedural task completed in this study would be to evaluate completed origami models using rulers and protractors. The current study could also be replicated using other types of tasks (e.g., cognitive, affective), MR virtual cues, and XR devices used for training, including other AR/MR headsets, AR/MR handheld devices, and VR devices. Other procedural tasks could be also examined, especially those that involve standing or walking to interact with larger or smaller objects, or those that require collaboration with others.

Acknowledgments. The authors would like to thank dissertation committee members Beth F. W. Atkinson, Elizabeth L. Blickensderfer, and Joseph R. Keebler for their contributions to study design and data analysis, as well as Darrin Knaggs, Reid Santiago, and Zachary Phillips for their contributions to data collection.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Brigham, T.J.: Reality check: Basics of augmented, virtual, and mixed reality. Med. Ref. Serv. Q. **36**(2), 171–178 (2017)

Burte, H., Gardony, A.L., Hutton, A., Taylor, H.A.: Knowing when to fold 'em: problem attributes and strategy differences in the paper folding test. Pers. Individ. Differ. **146**, 171–181 (2018)

Daling, L.M., Schlittmeier, S.J.: Effects of augmented reality-, virtual reality-, and mixed reality– based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. Hum. Factors **66**(2), 589–626 (2024)

Ekstrom, R.B., French, J.W., Harman, H.H., Dermen, D.: Manual for kit of factor referenced cognitive tests. Educ. Testing Serv. (1976)

Georgia Technical Institute of Technology. (n.d.). History of origami. https://paperdev.gatech.edu/kinetic-joy/history-origami

Han, X., Chen, Y., Feng, Q., Luo, H.: Augmented reality in professional training: a review of the literature from 2001 to 2020. Appl. Sci. **12**(3) (2022)

Hart, S.G.: NASAtask load index (NASATLX); 20 years later. Proc. Hum. Factors Ergon. Soc. Ann. Meeting **50**(13), 904–908 (2006)

Hart, S.G., Staveland, L.E.: Development of NASATLX (task load index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)

Hinderks, A., Schrepp, M., Thomaschewski, J.: User experience questionnaire (2018). https://www.ueq-online.org/

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

Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**(3), 203–220 (1993)

Long, L.K., DuBois, C.Z., Faley, R.H.: Online training: the value of capturing trainee reactions. J. Workplace Learn. **20**(1), 21–37 (2008)

Marr, B.: Extended Reality in Practice. Wiley (2021)

Microsoft. HoloLens 2 (2022). https://www.microsoft.com/enus/hololens/hardware

Novick, L.R., Morse, D.L.: Folding a fish, making a mushroom: the role of diagrams in executing assembly procedures. Mem. Cognit. **28**(7), 1242–1256 (2000)

Schrepp, M., Hinderks, A., Thomaschewski, J.: Design and evaluation of a short version of the user experience questionnaire (UEQS). Int. J. Interactive Multimedia Artif. Intell. **4**(6) (2017). https://doi.org/10.9781/ijimai.2017.09.001

Stachiw, S.: The future of XR training: What's to come in 2024. Training Industry (2023). https://trainingindustry.com/articles/learning-technologies/the-future-of-xr-training-whats-to-come-in-2024/

Stanney, K.M., Kennedy, R.S., Drexler, J.M.: Cybersickness is not simulator sickness. Proc. Hum. Factors Ergon. Soc. Ann. Meeting **41**(1), 1138–1142 (1997)

Stanney, K.M., Nye, H., Haddad, S., Hale, K.S., Padron, C.K., Cohn, J.V.: Extended reality (XR) environments. In: Salvendy, G., Karwowski, W., (eds.), Handbook of human factors and ergonomics (5th ed., pp. 782–815). Wiley (2021)

Tenbrink, T., Taylor, H.A.: Conceptual transformation and cognitive processes in origami paper folding. J. Problem Solving **8**(1) (2015)

Werrlich, S., Eichstetter, E., Nitsche, K., Notni, G.: An overview of evaluations using AR for assembly training tasks. Int. J. Comput. Inf. Eng. **11**(10), 1129–1135 (2017)

Wong, A., et al.: Instructional animations can be superior to statics when learning human motor skills. Comput. Hum. Behav. **25**(2), 339–347 (2009)

Zhao, F., et al.: Origami folding: Taxing resources necessary for the acquisition of sequential skills. Plos One **15**(10), e0240226 (2020)

Seamless Augmented Reality Support for a Computer-Assisted Surgery System for Minimally Invasive Orthopedic Surgeries

Julian Schlenker[1(✉)], Hendrik Vater[2], Enes Yigitbas[2], and Alexander Dann[1]

[1] metamorphosis GmbH, Paderborn, Germany
julian.schlenker@metamorphosis.ai
[2] Paderborn University, Paderborn, Germany

Abstract. There is a general trend in orthopedic surgery towards minimally invasive surgery. Because only small incisions are performed to make the inside of the patient accessible, a surgeon is unable to visually assess the surgical site. To overcome this limitation, X-ray imaging is used to provide the surgeon with insights about the actions taken inside the patient. However, given the projection of 3D structures onto the 2D image plane, this is still a challenging task for a surgeon. To further assist a surgeon, computer-assisted surgery (CAS) systems are used to help a surgeon to precisely navigate the tools and implants used during a procedure. While these systems provide valuable insights, current state-of-the-art CAS systems come with major drawbacks, including high acquisition and operational cost, additional hardware in an already crowded operating room (OR), and added complexity and change to the traditional surgery workflows. To address these issues, we propose an augmented reality (AR) extension of a novel CAS system, focusing on the cephalomedullary nailing procedure of the femur bone. As the extended system purely operates based on the acquired intraoperative X-rays, the AR extension provides real-time guidance between X-ray acquisitions and reduces the number of X-ray images taken during the procedure. This is achieved by augmenting instruction onto the surgeons' tools which are tracked by the AR glasses. To evaluate the feasibility and usability within a surgical setting with medical professionals, the solution has been tested in multiple experiments with cadaveric specimens. It was well received by the surgeons and almost all of the AR-guided drillings were successful.

Keywords: Augmented Reality · Computer-Assisted Surgery · Orthopedic Surgery · Usability

1 Introduction

There has been an ongoing trend towards minimally invasive surgery within the medical field of orthopedics [1]. During minimally invasive surgeries, the surgery team can not directly see important anatomical structures, tools, or

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 200–219, 2025.
https://doi.org/10.1007/978-3-031-93715-6_13

implants as the surgical site is not laid bare but instead, only a few small incisions are made to perform the surgical procedure [1]. While this brings a multitude of improvement for the patient, including but not limited to reduced infection risk and faster recovery [2,12], the surgeon, on the other hand, faces additional challenges during the procedure as a result of the restricted view [2]. X-ray imaging has proven to be an invaluable tool in the medical field as it is helpful in, e.g., noninvasive and painless diagnostics, treatment planning, and intraoperative guidance [2,16]. However, multiple problems come with the use of X-ray-based imaging techniques.

First off, X-ray images are generated using ionizing radiation which, depending on the dose and exposure time, increases the risk of cancer and can also have an effect on the tissue resulting in, e.g., cataracts, skin reddening or hair loss [16]. To ease these risks, exposure to ionizing radiation should be kept as short and as low a dose as possible. Secondly, in the process of fluoroscopic X-ray imaging, the real-world 3D structures, including anatomy and tools, are projected onto a 2D X-ray image. From this 2D projection, 3D information has to be inferred by the surgeon. This increases the difficulty of a procedure and requires a significant amount of skill and experience from the surgeon to be executed in a fast and secure manner [6]. This is amplified even further by the inconvenient display of X-ray images on a separate screen that hinders the hand-eye coordination of the surgeon [6] and spatially disconnects the acquired X-ray image from the surgical site. On top of that, the image acquisition is noncontinuous, meaning that each X-ray image is only valid at the time it was acquired and any change at the surgical site are only reflected in a newly acquired X-ray image [1]. As each X-ray is merely a snapshot, any changes in-between have to be extrapolated by the surgeon.

To overcome these problems, various types of CAS systems have been developed to support surgical teams in difficult procedures and reduce the likelihood of errors. However, incorporation of most of these systems into the surgery workflow can be quite cumbersome as they usually come with specific tracking hardware and can require additional setup and surgery steps [1,10]. Moreover, proper setup of said hardware and correct execution of the additional steps requires the surgical team to be familiar with the system, introducing an additional learning curve to the procedure at hand [1,10]. Another shortcoming of the usual CAS systems lies in the presentation of the information on a separate screen [1]. Similar to the problem with the visualization of X-ray images, the surgeon's attention constantly has to shift between the surgical site and the information displayed on some external screen [8]. As a result, the surgeon's hand-eye coordination suffers, potentially introducing errors into the procedure [6].

A CAS system by metamorphosis GmbH which, as of writing this, is still under active development and has not yet received clearance as a medical device, resolves the majority of the described issues for minimally invasive orthopedic surgery. By removing the need for reference bodies and trackers from the OR and sticking closely to traditional surgical workflows without enforcing additional steps, the system seamlessly integrates into the OR. It purely works on the intraoperative X-ray images by matching visible tools, implants and the patient anatomy in 2D and 3D. Based on this information, it provides the surgeon with highly useful information based on the current and previous X-ray images. This context based nature also means, that the surgeon only has to acquire a new X-ray image to receive a new system response, almost no interaction with the system required. The only two remaining issues are (1) that the system response is displayed on an external monitor, introducing a decoupling from the surgical site, and (2) that the system has no real-time feedback, only providing new information with a newly acquired X-ray.

To solve these remaining issues, an AR extension of the described CAS system utilizing a head-mounted device (HMD) is a natural choice. This is because it (1) enables the system to display valuable information directly in the surgical site by augmenting, e.g., instruments in use and it (2) enables the system to provide real-time feedback to the surgeon by live tracking objects in the OR based on sensor information provided by the HMD. Thus, the central research question (RQ) of this paper is to investigate to which extent a head-mounted AR device can be seamlessly integrated into the existing CAS system to support minimally invasive orthopedic surgeries. To answer this RQ, the described CAS system by metamorphosis GmbH is extended to interface with an HMD, namely the Magic Leap 2 (ML2). The focus of the AR support lies on supporting the surgeon to align a drilling machine with the correct drilling trajectory during the distal locking (DL) of an intramedullary nail (IMN). It is important to note that the system by metamorphosis GmbH is ideal for this use case as it performs a new registration with each X-ray, meaning that both the tracking of the instrument and the localization of the HMD only need to be accurate during a very limited time frame between two X-rays. This contrasts most conventional systems which have an initialization step at the beginning to register different coordinate frames or introduce additional hardware to achieve the same [17]. This either makes the localization of the HMD too inaccurate over the time span of an entire surgery or adds the burden of additional hardware in an already crowded OR.

We evaluated the feasibility and usability of our proposed AR solution within a surgical setting where multiple experiments with cadaveric specimens have been conducted. In addition, the accuracy of the used object tracking library is evaluated quantitatively on a dataset specifically recorded for this evaluation. Our evaluation results show that the AR solution was well received by the surgeons and almost all of the AR-guided drillings were successful. The quantitative evaluation of the object tracking reveals sufficient accuracy for the given use case and therefore supports these qualitative finding.

The remainder of the paper is structured as follows. In Sect. 2, we describe the medical background and related work. Section 3 presents the system design and the implementation of our AR solution. In Sect. 4, we present the evaluation and its main results. Finally, Sect. 5 gives a summary and outlook for future work.

2 Background and Related Work

This section introduces the required medical background and recent related work to be able to put this paper into context.

2.1 Medical Background

In this subsection, the medical application of treating femoral fractures with the IMN procedure is briefly described. This is one of the procedures supported by the CAS system by metamorphosis GmbH and the one chosen for the feasibility of the AR extension of the existing CAS system.

The overall goal of this procedure is to insert a metal rod, the IMN, lengthwise into the fractured femur bone and lock it in place by additionally inserting both a locking element and locking screw, as seen in Fig. 1. Once successfully implanted, the IMN fixates the fractured bone of the patient and helps both with healing and stabilizing the bone long term [2].

The IMN procedure involves multiple different steps from finding the correct entry point, inserting the nail, finding the ideal nail position, proximal locking and DL [4,14]. All of these steps are supported by metamorphosis GmbH's CAS system, but, for the scope of the AR extension, the focus lies on the DL step. As the DL step is one of the final steps of the IMN procedure, we assume all previous steps to have been successfully performed and consider them out-of-scope for this paper.

The DL step, as the name already suggest, takes place at the distal part of the femur, i.e., the part of the femur closest to the knee joint. At the time of this procedure step, the nail is already fully inserted into the femur and is locked in the proximal region, i.e., the region closest to the hip joint. The goal of the DL step is to place between one and three screws within the distal holes of the IMN to lock the nail in place distally. Before being able to insert the screw, for each desired hole, the surgeon uses a drilling machine to pre-drill the bone on the screw trajectories. The traditional approach using only X-ray imaging to perform a free-hand drilling, the workflow is to align the X-ray machine such that the desired hole appears both as round as possible and centered in the X-ray image. Once correctly aligned, the surgeon can use the X-ray machine as a reference to align the drilling machine with. This method is called down-beam positioning technique and requires many X-rays to precisely align the X-ray machine as described and involves multiple visual approximations by the surgeon, making it time consuming, prone to errors and require a significant amount of X-rays [5]. It is important to note, that for short IMNs an external physical targeting device exists that attaches directly to the nail and provides the surgeon with physical guidance, eliminating the necessity of the described approach. However, when using long IMNs, the nail bends during the insertion, making the use of a physical targeting device infeasible.

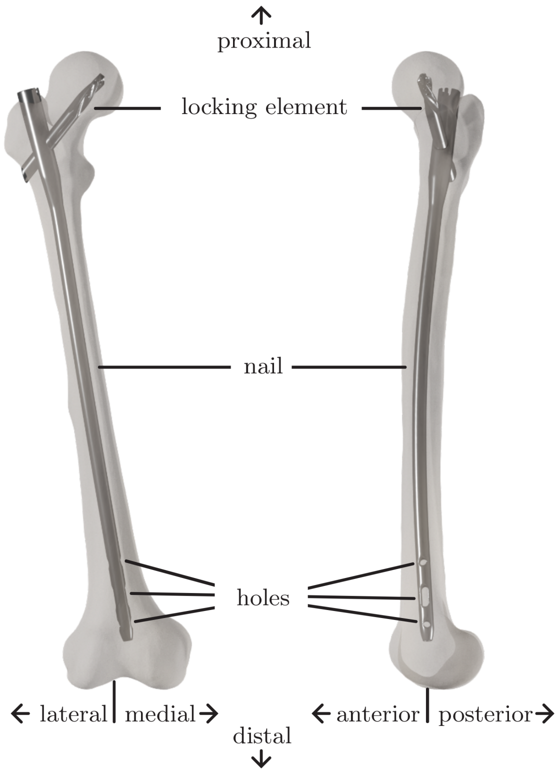


Fig. 1. 3D rendering of the IMN in the femur. The locking element, nail and the DL holes are annotated in the image. The DL holes have to be drilled before locking screw are inserted. All relevant medical direction are provided with the associated direction arrows.

2.2 Related Work

There is already a substantial amount of literature on the topic of AR-based CAS systems for the OR. In the following, we draw on prior research that focuses on the usage of state-of-the-art AR HMDs that do not fully rely on additional external sensors and/or consider evaluations based on cadaveric tests.

Two studies from Ma et al. [11] and Tu et al. [15] considered the usage of AR in the DL of long IMNs, which naturally relates them to this paper. Notably, both considered tibial IMNs instead of femoral ones. As the nature of the underlying challenges are closely related, it is still safe to compare these approaches.

Ma et al. uses electromagnetic (EM) tracking to determine the positioning of the IMN inside the leg and by extension the positions of the locking holes. Another component of their setup is the integral videography (IV) technology, which is a static AR device that can be used to overlay the real world with virtual content (for a more precise inspection of IV, view [11]). The positioning of the IV system is determined using an optical tracker that is fixed to the system and an optical tracking camera that is calibrated such that the internal coordinate frame of the tracker can be brought into relation with the EM tracking. The IV system can then superimpose the 3D information about the locking holes onto an image of the surgery site to assist with the DL procedure. Additionally, the electric drill that is used during the procedure is also tracked using an optical marker.

Tu et al. also use an EM tracking system to determine the position of the IMN and by extension the locking holes. Contrary to Ma et al., the electric drill is also equipped with an EM sensor instead of an IV system. Tu et al. uses the HoloLens 2 HMD (Microsoft, Redmond, United States) for visualization. A registration cube with yet another EM tracker is introduced so that the HMD can use the data collected by the EM tracking system. Due to its distinct shape, the cube is easily trackable for the HMD and due to the EM tracker it is easily detectable for the EM tracking system. Once the calibration is done, the HMD can visualize the drilling trajectories that are necessary to perform DL on an IMN.

The experimental results presented in either publication are fairly promising. A clear advantage of the presented EM-based method over the typical workflow used in orthopedic surgery is its potential to eliminate the necessity for intra-operative fluoroscopy. However, EM-based tracking is vulnerable to the presence of ferromagnetic metals [15] and the high amount of additional hardware as well as the extra setup that comes with it are considered serious drawback. It is not possible to naturally incorporate the presented ideas into the surgical workflow that is typically used for the DL of IMNs. EM-based CAS support is not the norm in the OR. The system presented by Tu et al. was also tested on a cadaveric specimen.

While the application of AR in maxillofacial surgery is not located in an orthopedic context, the consideration of research from this field is interesting nevertheless as it can help form an understanding of what is generally feasible with current AR hardware. One recent study [3] used the HoloLens 2 HMD (Microsoft, Redmond, United States) to investigate whether the 3D-printed cutting guides that are usually deployed in maxillofacial surgery could be replaced using AR. They used a QR-code-based registration cube to register the HMD to a phantom of the lower mandible. The HMD was then used to project cutting lines that were drawn onto the phantom by the participants of the study in multiple trials. The resulting accuracy was satisfactory with deviations of 1.03 mm and 1.27 mm for two of the applications and unsuitable for a third application with an accuracy of over 2.5mm. Overall, this shows that when a stable registration is accomplished, the accuracy of the HMD is sufficient for tasks with small error margins.

Other applications of AR in the context of surgery attempt to visualize the perspectives from which X-ray images were acquired to help with the interpretation of the 2D X-ray images in the 3D space, e.g. in [1,6]. In the first publication Andress et al. [1] used the first generation Microsoft HoloLens HMD (Microsoft, Redmond, United States) to estimate the position of a radiopaque marker while simultaneously acquiring an X-ray image showing the same marker.

In [6], the authors presented a related approach where the HMD is again registered to the intraoperative X-ray machine, the so called C-arm. Here, two first-generation Microsoft HoloLens HMDs were used. One was mounted to the C-arm and registered with the image intensifier of the C-arm such that their relative positioning and orientation are known. The second HMD is worn by the surgeon and both HMDs are using the simultaneous localization and mapping (SLAM) algorithm to create a spatial mapping of the OR. The spatial mapping of the surroundings is used to constantly update the transformation between the HMD that is worn by the surgeon and the HMD that is mounted to the C-arm. As a result, the coordinate system of the HMD that is worn by the surgeon can be brought into relation with the coordinate system of the C-arm. Now, whenever an X-ray image is acquired from the C-arm, an interactive flying frustum (IFF) can be generated. An IFF is a frustum that is located such that one base face is located at the X-ray source, while the other is located at the image intensifier. The X-ray image is then visualized in the frustum. It is oriented such that it always retains the orientation in which it was generated. Using the registration between the surgeon's HMD and the C-arm, the IFF can then be visualized by said HMD, and the input options of the HMD can be used to move it along the frustum. While the idea helps interpret the 2D X-ray images in 3D space, the cumbersome calibration and setup, as well as the need for a tracking device on the C-arm make it rather difficult to use the system in a real-world surgery workflow.

Another study [7] used the first generation HoloLens to quantitatively assess whether hologram stabilization may be improved when the HMD is used in collaboration with the image processing SDK Vuforia (PTC, Boston, United States). The setup was tested in a simulated neuronavigational application. A neuronavigational phantom was continuously tracked using Vuforia's feature detection algorithms, where the features visible in the sensory input of the HMD are compared with a 3D computer model of the object in the memory of the device to gain an estimate of the object's 3D position in relation to the HMD. This estimate was then used to stabilize the position of the hologram that is shown in the HMD. The authors found that by using Vuforia, the perceived drift of the hologram and the error in surface point localization could be reduced significantly in comparison to the usage of the HMD without Vuforia's tracking capabilities. While the HMD used here is outdated, the tests show that camera-based object tracking could be valuable in expanding the possibilities and improving the accuracy of HMD-based spatial recognition and hologram stability. Notably, the whole setup shown in this publication is based on inside-out tracking, i.e. does not rely on external tracking hardware.

In summary, while past applications of AR in CAS systems generally helped with visualization of the surgical site and produced promising results, they suffer from extra sensors and hardware which makes the integration into the real-world OR difficult. Additionally, there have only been few publications that use newer

HMDs such as the HoloLens 2 and the ML2 and there does not seem to be any publication that investigated the object tracking accuracy that is possible with such an HMD. Newer hardware could prove to be the key to developing AR-based CAS systems which do not rely on extra sensors or markers but instead can be integrated seamlessly into the OR by utilizing inside-out tracking. The system proposed in the context of this paper aims to build upon the shortcomings of these earlier systems and to fill their above-mentioned deficits.

3 System Design and Implementation

In this section, we describe the high level architecture and implementation details of our AR extension of the novel CAS system by metamorphosis GmbH.

3.1 System Design

The system can be divided into three subsystems and one actor, as shown in Fig. 2. The three subsystems are the *C-Arm*, the *CAS System* and the *HMD*. Both the *C-Arm* and the *CAS System* subsystems are considered to be black box systems. The *C-Arm* subsystem provides a video interface *IXRay* for obtaining the acquired X-ray images. The *CAS System* subsystem connects to the *IXRay* interface of the *C-Arm* and provides the *ICAS* interface. This subsystem is responsible for processing the intraoperative X-ray images, registering anatomy, implants and tools based on the X-ray and generating instruction based on the registrations. Thus, it has to implement a complex image processing pipeline and business logic to achieve its goals. The *HMD* subsystem connects to the *ICAS* interface to retrieve current instructions computed from newly acquired X-rays. It is responsible for visualizing the instructions and the tracking of the surgeon's tool. The subsystem is further divided into two components: The *Unity Application* and the *Tracking* component. The *Unity Application* component is responsible for the business logic of the AR application and the visualizations presented to the user. It is also, from a users perspective, the entry point into the entire system. The *Tracking* component is responsible for the inside-out tracking of real-world objects based on the available sensor data. It provides an interface to retrieve poses of the tracked object to which the *Unity Application* attaches to. A sophisticated *Tracking* component in combination with the *HMD* component enables the overall system to eliminate trackers and reference bodies. Lastly, the only actor in the system is the surgeon, interacting with both the *C-Arm* to acquire new X-rays on demand and with the HMD to receive AR instruction based on the information provided by the *CAS* and the *Tracking* component. It is important to note that in a real-word OR, the surgeon would usually interact with the *C-Arm* through an additional actor: the X-ray technician.

Following the directions of the interfaces, it again becomes apparent, that the *Unity Application* is the entry point into the designed systems, as it only attaches to interfaces without providing any. Further it follows, that the system inputs come from (1) the *C-Arm* component in the form of new X-rays and (2)

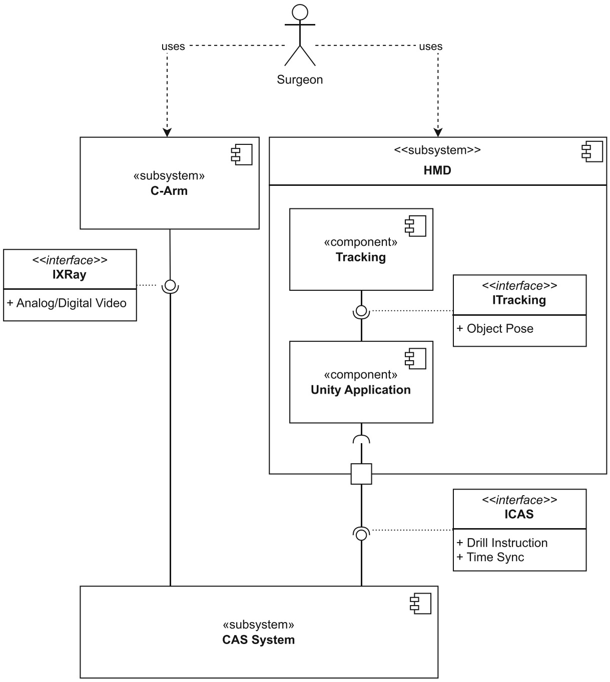


Fig. 2. Simplified high level architecture of the AR extension of the CAS system by metamorphosis GmbH.

the *Tracking* component in the form of tracked object poses retrieved from the sensory data available on the HMD subsystem.

3.2 Implementation

After introducing the overall system design, this subsection expands the system design with specific implementation details.

The *CAS System* component is a placeholder for the CAS System by metamorphosis GmbH we are aiming to extend with AR capabilities in the scope of this paper. This system has already fully implemented the interface with the *C-Arm* component, therefore these two subsystems are already given. This also means, that the supported C-arms for the overall system reflect the supported C-arms by the metamorphosis GmbH CAS system, i.e., as of writing this, most common C-arms, including distorted image intensifiers.

As the *HMD* subsystem, we used the Magic Leap 2 (ML2) as a modern and powerful AR device. For the *Tracking* component, the commercially available tracking solution Vuforia Engine (PTC, Boston, United States) is used. As of writing this paper, this is the industry leading tracking solution and it provides support for the chosen ML2 platform.

Lastly, the *Unity Application* component, as the name suggests, is the main application running on the HMD and is implemented using the Unity Game Engine. Its main purpose is to join the information provided by the CAS system and the object tracking to provide the *Surgeon* actor with precise and intuitive support during DL. To achieve this, this component has to overcome the major challenge of aligning the coordinate frames of the CAS system and HMD, accurately enough to reliably perform DL. Aligning the two coordinate frames has both a temporal and spatial component to it.

The temporal component is caused by the processing delay of the CAS system. At the time at which the current instruction is provided to the HMD, the X-ray used to compute the instruction is already as old as the processing time of the CAS system. Therefore, the instruction the HMD receives corresponds to the tool position at the time of the X-ray acquisition, not the time of receiving the instruction. To eliminate this issue, two measures are taken. First, the CAS system and the HMD need a synchronized clock to accurately communicate at which point in time the processed X-ray was acquired. For this purpose, the *ICAS* supports a time synchronization handshake. Secondly, the *Unity Application* has to store enough past tracked poses provided by the *Tracking* to compensate for the maximal processing delay of the CAS system. Then, upon receiving a new instruction, the *Unity Application* searches for the past tracked pose which has the smallest time difference to the time of X-ray acquisition. Additionally it is insured, that the time delta between X-ray acquisition and the closest tracked pose is less than $50ms$. If that is not the case, the surgeon receives an audiovisual instruction through the HMD, requesting a new X-ray while ensuring that the tool is being tracked simultaneously.

The spatial component of the coordinate frame alignment is caused by that fact that the HMD and the CAS system have two independent coordinate frames. For the scope of this paper, we do not provide AR support for drill tip placement. This means, that the position of the drill tip on the bone surface over the locking hole of the IMN is performed as if there is no AR support, with the benefit being that the tip movement instructions are displayed directly in the surgeon's field of view through the HMD. As a result, we only need to reliably display rotation instruction around the drill tip to align the drill with its target trajectory, as shown in Fig. 3. These instructions are always split into angle adjustments along two axes. We utilize knowledge about the patient position and the HMD world frame to build a common understanding of axes along which the angle instruction are defined. The *Unity Application* also implements additional 3D visualizations of the matched patient anatomy, tool and implants and additional head-up elements to improve overall usability and user feedback.

While the presented implementation is based on the ML2 HMD, the system architecture employs a hardware abstraction layer to ensure device independence. This is achieved through standardized data protocols and the use of standards like OpenXR to achieve hardware independence.

This implemented system is capable of solving multiple issues found and discussed in Subsect. 2.2. While existing proposed AR solutions mostly suffer from

Fig. 3. Surgeon's view during drill alignment. Left image illustrates a tilt instruction by 1.8° up and 11.2° left. Right image illustrates the green confirmation while being properly aligned. (Color figure online)

multiple external sensors and additional expensive and clumsy hardware, this system only relies on the already present intraoperative X-ray machine, a computer for the metamorphosis GmbH CAS system and a single additional HMD worn by the surgeon. This introduces only few and relatively small hardware components into an already crowded OR, while stripping the procedure of any cumbersome initial registration processes often required by CAS systems. The truly unique benefit lies in the capability of metamorphosis GmbH's CAS system to perform new registration with every new intraoperative X-ray acquired. This way, the HMD SLAM and its object tracking only have to be highly accurate for the brief periods of time between two X-rays, not relying on an initial registration performed at the very beginning of a procedure or additional hardware to achieve accurate tracking over long time periods. The following section will explore how the implemented system performed both based on quantitative results, cadaveric experiments and surgeon interview.

4 Evaluation

This section describes the qualitative and quantitative evaluation conducted to be able to properly assess the feasibility of the proposed AR extension of the CAS system by metamorphosis GmbH.

4.1 Qualitative Evaluation

The qualitative evaluation is based on both, observations during and interviews after surgery sessions with cadaveric specimens. This aims to paint a picture of how the system performs in a setting close to the real-world OR and how well the acceptance rate is with the surgeons using the system.

Cadaver Experiments. For the cadaver experiments, multiple different surgeons performed the DL procedure using the AR-extended CAS system by metamorphosis GmbH. During these procedures, each drilling of a locking hole is recorded as hit or miss. A drilling is considered a hit if the surgeon drills through both cortices and through the distal locking hole of the IMN successfully. This is verified through a true lateral X-ray, as required by the down-beam positioning technique, after each drilling with the drill bit left inside the specimen. When the drill bit is rendered inside the DL hole on the X-ray, the drilling went through the locking hole. Any drilling which does not go through the desired locking hole is considered a miss. Using these criterions, 42 drillings have been performed across 8 cadaver labs over the span of almost two years using the AR-extended CAS system. Of these 42 drillings, 37 are considered a hit while 5 are considered a miss, leaving us with an 88% success rate. It is important to note that the 5 missed drillings only occurred during the first three months and therefore in a very early stage of the prototype. Additionally, these missed drilling can be traced back to slipping of the drill bit during the drilling, the absence of a drill sleeve, causing soft tissue to wrap around the drill bit and abnormally hard bones. These results clearly demonstrate the feasibility of the proposed AR extension and showcase the high accuracy of the entire CAS pipeline.

All procedures were conducted following strict ethical guidelines and applicable regulatory requirements, with informed consent from participating surgeons. The use of cadaveric specimens upholds the fundamental medical ethics principle of "first, do no harm" by enabling critical surgical innovations to be thoroughly validated before clinical implementation, thereby maximizing patient safety while respecting the wishes of donors who contributed to medical advancement.

Expert Interviews. Informal expert interviews with some of the surgeons were conducted to gain basic insight into their willingness to use such a system in the OR and gather a first round of critical feedback. A list of questions was asked and the surgeons' answers were noted in the form of bullet points. Due to the rather loose format of the interviews, there are no objective results in the form of questionnaires. The answers should nevertheless be sufficient to give an initial indication of the subjective usefulness of the prototype and by extension, one could infer in which direction the concept should be developed in the future. Three of the surgeons who participated in the cadaveric experiments were interviewed. One takeaway from the interviews is that continuous feedback in the drilling phase is helpful. Usually, the surgeon has to drill along a certain trajectory without any feedback, and knowing when one diverges from this trajectory is valuable. The object tracking subjectively felt accurate enough to the surgeons. They all felt confident in following the instructions. The exception was that sometimes the tracking was a bit "jumpy" or "jittery" which reduced the confidence in the accuracy of the instructions. The perspective of the drilling machine that seems best for the object tracking algorithm was seen as slightly uncomfortable, but this could be partly alleviated by changing the height of the table, etc. The reduction in the field of view that naturally occurs from the wearing of the HMD was seen as unproblematic for the given application, since the surgeon mostly operates alone in the cephalomedullary nailing surgery, but could be problematic when there are more people involved and the whole situation is more dynamic. The slight delay in tracking behind the real object was not seen as a problem for any of the surgeons. The surgeons noted that there is a learning curve to the usage of the system, as one has to understand the instructions and the way the drilling machine is supposed to be held initially. This, however, is an issue which can be addressed with improved training and incorporating the feedback from further usability studies. A negative point towards the usage of the ML2 HMD specifically was that the surgeon could not wear glasses underneath, which is possible with the HoloLens 2 HMD, for example. The ML2 does provide the possibility to use prescription insert for overcoming this issue, it is just not as seamless as using ones glasses directly. Overall, the feedback indicated that the positive point of continuous feedback is important and would be a great benefit during surgery in the future. They also indicated that they would be willing to use the system in surgeries in the future.

4.2 Quantitative Evaluation

The goal of the quantitative evaluation is to get an understanding of the actual achieved accuracy by the object tracking performed on the HMD. While there exist many datasets to measure the performance of general tracking solutions, we aim to evaluate the performance on real-world data. This means that we do not only want to record the data directly on the target HMD ML2 with its real sensors, but also record the real target object, the drilling machine used in the OR. Additional to the sensor data, the ground-truth (GT) pose of the tracked object relative to the HMD has to be determined and recorded accurately. To achieve this, a test bench involving a robotic arm was designed to record an evaluation dataset. In the following, the designed test bench and the data recording process are briefly explored and the evaluation results based on the recorded datasets are presented.

Test Bench and Data Recording. For the test bench, both an attachment for the ML2 and the drilling machine have been designed and 3D-printed to securely fixate the HMD and the tracked object during sequence recording, respectively. The setup of the test bench can be seen in Fig. 4. The sequences which are meant to be recorded have to record frames which resemble single frames from a continuous video. To achieve this, there are generally two approaches: (1) you let the robotic arm constantly move around at a certain speed and capture the frames in real time or (2) you sample many poses along a previously computed smooth path and move the robot to each of these poses sequentially and capture a frame at each pose, not in real time. While the first approach is closer to how the data is captured in the actual use case, including potential motion blur, it presents multiple technical issues. These include precise synchronization of the captured sensor data and the pose reported by the robotic arm and the writing speed of the sensor data onto the ML2 storage. Therefore, the second approach is chosen, computing all poses of each sequence in advance, moving the robot to each pose sequentially and ensuring the ML2 has captured all of its sensor data before continuing to the next pose. This way, the movement of the robot arm between two recorded poses is not of interest, also making the planning of the robot movement easier.

The algorithm developed for sampling the poses for each sequence has several parameters which allow for variations of the movement speed, i.e., the relative distance between two sampled poses, the area in which the poses are sampled and the bounds, speed and origin of the possible rotational changes of the sampled poses. This allows to capture different sequences for different use cases.

Evaluation Results. For the evaluation, 12 different sequences have been recorded with 1700 frames each. Each sequence starts at a position roughly 50cm away from the ML2 camera and the parameters are chosen so that the camera frame of the ML2 can only be left partially by the tracked object, allowing for continuous tracking throughout the entire sequence. Six of the 12 sequences focus on sampling rotations around three different references, with small and large rotational steps each. The different reference points can be seen in Fig. 5. Specifically the rotation around the *Drill Tip* represents the exact use case found during the drill angle instruction as seen in Fig. 3. All of these sequences are prefixed with "exp_around_". The next four sequences sample both large and small translational movement, each in combination with small and large rotations. These sequences are prefixed with "exp_large_mov_" and "exp_small_mov_", respectively. The last two sequences only move the target object translationally and are prefixed with "exp_mov_".

These 12 sequences are then evaluated with three different scores: the s_{ADD}, s_D and s_R. The s_{ADD} score measures the overlap of the tracked and GT model by computing the average distance between the vertices of the tracked and GT model which is then used in an area under curve score [13]. This results in a score $s_{ADD} \in [0, 1]$, where 1 is a perfect match and 0 represents an average vertex distance of 10cm or higher [13]. The s_D score simply describes the Euclidean distance of the tracked *TCP* reference point (see Fig. 5) and the GT *TCP* reference point in mm. This gives an idea of how accurate the tracking is with regards to translational error. Lastly, the s_R score represents the angle between tracked and GT drilling trajectory vectors in deg. This gives an idea of how accurate the tracking is with regards to rotational error, as, e.g., required during the described drill angle instruction for the user.

In Table 1, the scores for all recorded sequences are shown. For these evaluations, the tracking is initialized using the first ground truth pose of each sequence and therefore only focuses on the tracking accuracy, not the tracking initializa-

Fig. 4. Photo of the test bench for recording object tracking sequences with GT poses. On the left side, the ML2 is statically mounted on a tripod. On the right side, the robot arm HCR-3A from Hanwha Robotics can be seen, with the a medical grade drilling machine attached to it.

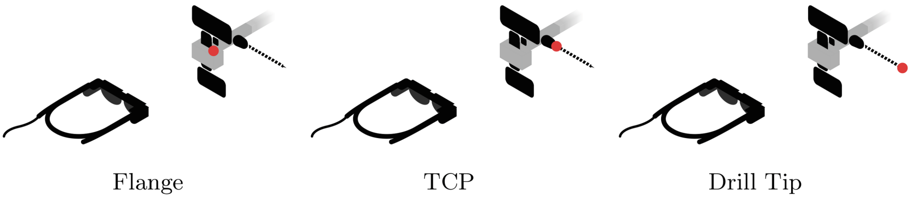


Fig. 5. Illustration of the three reference points for the rotations: flange, TCP and drill tip. Each one is marked by a red dot in their respective image.

tion. Based on the resulting values, we can expect an average tracking error with respect to the drilling angle of approximately 1 *deg* across different scenarios. The expected translational error of the tracked TCP across all scenarios is approximately 2.5 *mm*. It is important to note that this score is more susceptible to inaccuracies introduced by the entire pipeline of the test bench and is therefor likely not highly accurate. In comparison to [13], also the s_{ADD} score is very good, i.e., a good tracking of the entire match is achieved.

Table 1. Table showing the s_{ADD}, the s_D and the s_R score for the 12 recorded evaluation sequences based on tracking with the Vuforia Engine. The best scores per column are printed bold while the second best scores are underlined.

Sequence	s_{ADD}	s_D (mm)	s_R (deg)
exp_around_tcp_small_rot	0.972	1.524	0.672
exp_around_tcp_big_rot	0.967	1.679	0.825
exp_around_drill_tip_small_rot	<u>0.974</u>	2.323	0.665
exp_around_drill_tip_large_rot	0.972	2.191	<u>0.545</u>
exp_around_flange_small_rot	0.962	1.558	1.079
exp_around_flange_large_rot	0.957	2.499	1.080
exp_large_mov_and_large_rot	0.950	3.827	0.980
exp_large_mov_and_small_rot	0.944	3.624	1.226
exp_small_mov_and_large_rot	0.932	5.018	1.592
exp_small_mov_and_small_rot	**0.983**	<u>1.281</u>	**0.281**
exp_mov_large	0.946	3.001	1.447
exp_mov_small	0.970	**1.006**	1.212
avg.	0.960	2.460	0.967

Based on these values, it can be concluded that the pure tracking performance of Vuforia is sufficient to be used in a CAS system for real time guidance. Specifically the results for the accuracy of the drilling angle instruction used for proper drilling trajectory alignment are well within a range suitable for real world applications.

4.3 Threats to Validity

While both the quantitative and qualitative evaluations yield very promising results, it is important to point out potential weaknesses and improvements of the evaluation to tell the full story. Therefore, this subsection explores the potential threats to validity and potential improvements which can be done in the future.

Regarding the qualitative evaluation, it should be noted that a commercially available drilling machine by Makita is used throughout. This drilling machine has a feature rich texture and an unreflective surface. In contrast, a typical surgical drilling machine exhibits fewer features and uniform metallic or black surfaces, rendering the object tracking task more difficult. Hence, additional hands-on cadaveric experiments should be conducted with a proper surgical drilling machine in use.

The quantitative evaluation, on the other hand, was performed using a surgical drilling machine with a metallic surface. The good tracking performance using this drilling machine suggest that the qualitative evaluation with this surgical drill will still yield equivalent results. However, the test bench introduced for recording the evaluation datasets represents a highly controlled environment. This contrasts the very dynamic and uncontrolled environment typically present in a real-world OR. Additionally, the initialization and report-of-failure steps of the tracking pipeline are excluded from the evaluation by providing the initial pose from the recorded GT dataset and designed the dataset in a way such that the tracked object only ever partially leaves the camera frame. A reliable initialization and report-of-failure steps are crucial steps in every object tracking pipeline to achieve a good usability. This is because a perfect tracking algorithm that only is initialized in a few perfect conditions or is constantly invalidated by a poor report-of-failure metric is not usable in a real-world scenario. Therefore, additional evaluation datasets should be recorded which allow evaluation of the entire object tracking pipeline, including initialization and report-of-failure. Lastly, the inaccurate positioning of the AR device relative to the robot base while recording the evaluation datasets introduces an error for the evaluation. While this error should be marginal for the angle evaluation, the translational accuracy of the GT drilling machine suffers from this accuracy. It is difficult to put precise numbers to the inaccuracies introduced by these factors but it is safe to say that they are greater than a millimeter.

4.4 Discussion

From the presented cadaveric experiments, it could be found that the system is usable in the context of a surgery with an 88% success rate. The few drillings that did miss, however, occurred in very early stages of the prototype development and can also be traced back to causes beyond the proposed AR-extended CAS system.

Informal expert interviews showed that the continuous real-time feedback introduced by the AR HMD was conceived as helpful and that they felt confident following the provided instructions. They also revealed potential for improvement regarding usability, specifically the slightly unnatural handling of the drilling machine to ensure optimal tracking performance.

Additionally, quantitative evaluations of the used tracking based on an evaluation dataset specifically created for this application have been performed. The evaluation of the datasets shows an expected drilling angle deviation of approximately 1 *deg* and an expected TCP position deviation of approximately 2.5 *mm*. Comparing to the results from Heining et al. [9] where a similar procedure is performed with AR support using marker-based tracking, our results numerically underlined that the tracking is accurate enough for the task at hand and therefore supports the statements of the surgeons and experts. Heining et al. [9] scores an average translational error of 3.99 *mm* and an average rotational deviation of 4.3 *deg*, however it is worth noting that these values refer to the final position of the implant, not purely the tracking accuracy of the tool, and therefore these results have a longer chain for accumulating evaluation errors.

The central RQ of this work was to investigate to what extent an HMD can be integrated into a CAS system to support minimal invasive surgeries. For one, this system should only use the HMD as extra hardware for the surgeon to use and secondly, it should integrate itself smoothly into the workflow of the surgery. The presented prototype is an AR extension of the existing novel CAS system by metamorphosis GmbH, where the surgeon only interacts with the HMD as extra hardware and no further sensors, markers or reference bodies are used. There is only a computer on which the fluoroscopy-based CAS system is running but apart from starting the application and connecting the computer to the X-ray machine, there is almost no further interaction with that computer. The consensus among the surgeons was that the continuous feedback subjectively makes the DL step easier and more reliable - this is also shown numerically. While a learning curve was observed, improvement could already be seen after a few uses. At this point, it seems that the prototype and the concept behind are a good way to integrate an HMD into the surgery to gain the benefits of visualization and tracking without the necessity for further sensors or markers.

5 Summary and Outlook

In this paper, we have presented an AR extension of the novel CAS system by metamorphosis GmbH. This extension aims to further improve the accuracy and usability of existing CAS system and its support for different minimally invasive orthopedic surgeries. Minimally invasive surgery is a complex task since the surgical site is not laid open but only small incisions are performed to complete a surgery with minimal damage to the patient tissue and therefore reduces blood loss and the patients recovery time. To perform a surgery under such circumstances, X-ray imaging is used to provide the surgeon with information about the position of the inserted tools and implants inside the patient. While this is very helpful and crucial for the surgeon, the projection of 3D information onto a 2D image still makes it difficult to precisely determine 3D information.

To tackle this problem, we have introduced an AR extension to support the drill alignment process in the DL step of the cephalo-medullary nailing procedure of the femur. The AR instructions are based on the output of the extended fluoroscopy-based CAS system that matches the anatomy, tools and implants visible in the current X-ray in 3D. Based on this 3D information, the proper drilling trajectory is calculated, transmitted and used by the AR extension to guide the user in real time based on the sensor information provided by the AR device.

Qualitative evaluations based on cadaver experiments and expert interviews with medical doctors show that such an extension, combining AR with a novel CAS system has the potential to reduce the probability of incorrect drillings in the DL procedure. The presented results and positive feedback from the surgeons clearly hint at the potential of the solution. Quantitative evaluations based on a dataset specifically recorded for this use case back up the promising results of the qualitative evaluation. The evaluated datasets show an average error for the drilling angle of just under $1deg$. Overall it can be concluded that the proposed AR extension is highly promising and well suited for integration with an existing CAS system, answering the central RQ.

In future work, multiple paths could be taken to improve upon the given prototypical implementation. First off, the qualitative evaluation could be extended to cover additional scenarios, including scenarios where the tracked object leaves the camera frame to also evaluate the initialization and report-of-failure capabilities. The presented prototype could naturally be extended to the locking of other intramedullary nails such as the tibial intramedullary nail. Another interesting application is supporting the surgeon in the drilling of pedicle screws. The surgeon is presented with a similar trajectory alignment problem and the concept could support the surgeon in aligning the drill and holding said alignment.

References

1. Andress, S., et al.: On-the-fly augmented reality for orthopedic surgery using a multimodal fiducial. J. Med. Imaging **5**(2), 021209 (2018)
2. Bekos, A., Sioutis, S., Kostroglou, A., Saranteas, T., Mavrogenis, A.F.: The history of intramedullary nailing. Int. Orthop. **45**(5), 1355–1361 (2021). https://doi.org/10.1007/s00264-021-04973-y
3. Brunzini, A., Mandolini, M., Caragiuli, M., Germani, M., Mazzoli, A., Pagnoni, M.: Hololens 2 for maxillofacial surgery: a preliminary study. In: International Conference on Design, Simulation, Manufacturing: The Innovation Exchange, pp. 133–140. Springer (2021)
4. DePuy Synthes: Surgical technique guide tfna. https://www.jnjmedtech.com/en-US/product/tfn-advanced (2024), Accessed 15 Jul 2024
5. Diotte, B., et al.: Multi-modal intra-operative navigation during distal locking of intramedullary nails. IEEE Trans. Med. Imaging **34**(2), 487–495 (2014)
6. Fotouhi, J., et al.: Interactive flying frustums (IFFs): spatially aware surgical data visualization. Int. J. Comput. Assist. Radiol. Surg. **14**(6), 913–922 (2019)
7. Frantz, T., Jansen, B., Duerinck, J., Vandemeulebroucke, J.: Augmenting microsoft's hololens with vuforia tracking for neuronavigation. Healthcare Technol. Lett. **5**(5), 221–225 (2018)
8. Gavaghan, K., et al.: Evaluation of a portable image overlay projector for the visualisation of surgical navigation data: phantom studies. Int. J. Comput. Assist. Radiol. Surg. **7**(4), 547–556 (2012)
9. Heining, S.M., Raykov, V., Wolff, O., Alkadhi, H., Pape, H.C., Wanner, G.A.: Augmented reality-based surgical navigation of pelvic screw placement: an ex-vivo experimental feasibility study. Patient Saf. Surg. **18**(1), 3 (2024)
10. Hestehave, R.A., Gundtoft, P.H., Nielsen, C.L., Brink, O., Rölfing, J.D.: Poor usability of computer-assisted navigation for hip fracture surgery. Arch. Orthop. Trauma Surg. **144**(1), 251–257 (2024)
11. Ma, L., et al.: Three-dimensional augmented reality surgical navigation with hybrid optical and electromagnetic tracking for distal intramedullary nail interlocking. Int. J. Med. Robot. Comput. Assist. Surgery **14**(4), e1909 (2018)
12. Mohiuddin, K., Swanson, S.J.: Maximizing the benefit of minimally invasive surgery. J. Surg. Oncol. **108**(5), 315–319 (2013)
13. Stoiber, M., Elsayed, M., Reichert, A.E., Steidle, F., Lee, D., Triebel, R.: Fusing visual appearance and geometry for multi-modality 6dof object tracking. arXiv preprint arXiv:2302.11458 (2023)
14. Stryker: operative technique gamma3. https://www.stryker.com/us/en/trauma-and-extremities/products/gamma3.html (2024), Accessed 15 Jun 2024

15. Tu, P., Gao, Y., Lungu, A.J., Li, D., Wang, H., Chen, X.: Augmented reality based navigation for distal interlocking of intramedullary nails utilizing microsoft hololens 2. Comput. Biol. Med. **133**, 104402 (2021)
16. U.S. Food and Drug Administration: Medical x-ray imaging. https://www.fda.gov/radiation-emitting-products/medical-imaging/medical-x-ray-imaging (2024), Accessed 15 Jun 2024
17. Wallace, N., et al.: Computer-assisted navigation in complex cervical spine surgery: tips and tricks. J. Spine Surgery **6**(1), 136 (2020)

Immersive Active Shooter Response Training and Decision-Making Environment for a University Campus Building

Sharad Sharma(✉) and Pranav Abishai Moses

Department of Data Science, University of North Texas, Denton, TX, USA
sharad.sharma@unt.edu, pranavabishaimoses@my.unt.edu

Abstract. There is a critical need to improve emergency response, decision-making, and safety for the most critical threats to public spaces today. Active shooter events are one of the most critical threats that require training for high-pressure decisions that need to be made during such situations. This paper presents the prototype of an immersive active shooter response training and decision-making environment for a university campus building. The immersive active shooter response training environment is developed in Unity 3D and is based on run, hide, and fight modes for emergency response. We have presented a multi-user virtual reality (VR) platform where experiments for active shooter response can be conducted using computer-controlled (AI) agents and user-controlled agents. This platform can be used as a teaching and educational tool for navigation and performing VR evacuation drills for active shooter events. A user study was conducted to evaluate the immersive VR training environment with 195 participants. The experiment sought to examine how the proposed tool influenced participants' understanding of the safest actions to take during an active shooter situation. The evaluation includes Group Environment Questionnaire (GEQ), Presence Questionnaire (PQ), System Usability Scale (SUS), and Technology Acceptance Model (TAM) Questionnaire. The findings suggest that the participants' knowledge, intrinsic motivation, and self-efficacy showed a significant increase immediately after the training. The results show that the majority of users agreed that the sense of presence was increased when using the immersive emergency response training environment for an active shooter evacuation environment. Through the use of an immersive VR platform, trainees develop a heightened sense of spatial awareness and an understanding of how to navigate the building in high-stress situations, thus increasing the chances of survival and successful evacuation.

Keywords: Virtual reality · immersive VR · building evacuation · training simulation

1 Introduction

The use of immersive active shooter training drills allows for training for situations that could not be tested in real life due to legal issues and possible health risks to participants. The multi-user virtual reality (VR) environment goes beyond traditional or tabletop exercises by immersing participants in scenarios where they interact with realistic, dynamic,

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 220–232, 2025.
https://doi.org/10.1007/978-3-031-93715-6_14

and often unpredictable situations. The use of an immersive active shooter response training environment improves the realism of the training by allowing participants to engage with each other in real-time scenarios, making decisions, responding to changes, and adapting to dynamic situations that mirror real-world challenges. VR enhances the spatial awareness of the users by allowing them to understand how best to navigate or make decisions based on the layout of the building. Sharma et al. [1–4] have developed an active shooter response training environment for a building evacuation in a collaborative virtual environment. They have presented an immersive security personnel training module and a civilian training module for an active shooter event in an indoor building. They have developed an experimental platform for conducting immersive training for performing virtual evacuation drills.

A multi-user VR environment places participants directly in the same environment using VR headset where each participant enters a fully immersive digital world. Participants can navigate around the space using VR headsets controllers, interact with objects in the environment, and respond to auditory cues, such as gunshots or cries for help. This increases the emotional and psychological engagement of the training which helps create a "sense of presence" in the environment. Traditional exercises for active shooter or fire evacuation drills often rely on pre-scripted events that unfold predictably. They allow for basic learning and repetition, but they don't fully replicate the complexity and unpredictability of real-world events. On the other hand, a multi-user VR environment allows for more advanced learning by replicating scenarios and user interactions that are dynamic and often unpredictable. The multi-user VR environment allows for the incorporation of dynamic and random elements such as an unexpected fire, a change in lighting, or the appearance of additional threats, keeping participants to critically think under pressure for various what-if scenarios. In a multi-user active shooter environment, participants can learn through experience, which aligns with a constructivist approach to learning. It allows for making choices and experiencing outcomes based on different what-if conditions, and replay scenarios, allowing participants to refine their decision-making and correct previous errors. In active shooter response training situations, VR can simulate the unpredictability of a threat. The active shooter's location can be changed forcing trainees to make on-the-spot dynamic decisions about how to react. Moreover, training in VR can replicate real-life emergencies such as sirens, loud gunshots, or the chaotic sounds of people in a disaster unfolding.

This paper presents a multi-user VR platform for conducting immersive training for an active shooter event for a university campus building. We have developed an immersive virtual reality training environment for active shooter events using the Unity game engine by integrating it with Meta Quest 3 hardware as shown in Fig. 1. The immersive nature of VR training and incorporation of fire and smoke creates a higher level of stress, which is necessary for learning how to manage anxiety and operate under pressure. The proposed multi-user VR platform can also be helpful for performance reviews such as decision speed, movement efficiency, and decision-making strategy. The data gathered from user participation can provide data-driven insights into how participants performed during a stressful and anxiety-inducing emergency situation.

Fig. 1. Immersive VR environment for active shooter response in a building on the university campus.

The rest of the paper is structured as follows. Section 2 briefly describes the related work for immersive training for an active shooter response, and disaster response training. Section 3 describes the implementation of an immersive VR active shooter response environment. Section 4, describes the evaluation of the immersive active shooter training environment. Section 5 discusses the drawn conclusions. Finally, Sect. 6 states acknowledgments.

2 Related Work

A Collaborative virtual reality environments (CVE) represent a powerful tool for active shooter response training. Their immersive and interactive nature enhances the realism of training scenarios and helps improve teamwork and communication among participants. As research continues to demonstrate the effectiveness of VR in training applications, the integration of CVEs into emergency preparedness programs is likely to become increasingly prevalent.

Active shooter response training in a VR environment can be impactful because of its ability to simulate real-life emergencies in a controlled setting. Studies have shown that VR can significantly enhance situational awareness and decision-making skills among participants [5, 6]. For instance, immersive VR environments can replicate the stress and urgency of an active shooter situation which enables the users to experience and react to simulated threats without the risks associated with them [7, 8]. The collaborative aspect of VR enables teams to practice communication and coordination, which are vital components of effective emergency response [9]. Immersive VR has been used as an education, training, and emergency response tool for an aircraft evacuation [10], a library evacuation [11], a subway evacuation [12], a megacity evacuation [13], a night club evacuation [14], a university campus evacuation [15].

VR environments can improve the psychological preparedness of individuals facing emergency scenarios. By engaging in realistic simulations, participants can develop a better understanding of their responses to stress and anxiety, which are common during such incidents [8]. The use of VR in training has been linked to increased motivation and engagement, as users often find the immersive experience more compelling than traditional training methods [16]. This heightened engagement can lead to better retention of training protocols, such as the run, hide, and fight protocols, which are critical for survival during active shooter events [17].

3 Implementation of Immersive Active Shooter Environment

The implementation of this project is divided into five main phases: Modeling, Unity Integration, GUI and User Interaction, Photon Integration, and VR Integration. Each phase builds upon the previous, culminating in a fully immersive, multi-user virtual environment.

3.1 Phase 1: Modeling

The first phase involved creating a detailed, to-scale replica of the campus building using 3D software's such as 3ds Max and Google Sketch UP. The 3D digital model of the physical building includes all its architectural details, dimensions, textures, and features and is represented in a three-dimensional virtual environment. As the 2D architectural model was available, it was easy to create 3D model of the building. The creation of a 3D model of the campus building involved:

- Extruding the 2D floor plan with precise dimensions to a 3D floor plan.
- Adding walls, windows, doors, and other structural elements.
- Incorporating architectural details like stairs, columns, and roofs.
- Including features like lighting fixtures, furniture, textures, and other signage that are part of the space.
- Textures were applied to give it a realistic appearance.
- Adding lights and environment setup.

Real-time images were captured for textures to be applied to carpet, walls, signage, etc. Textures were applied to give it a realistic appearance by adding finishes like paint, flooring, wall coverings, and other surface details. 3D modeling software allowed for the application of materials and textures that mimic real-world surfaces, adding depth and realism to the model. The incorporating lighting was a crucial step in recreating the building's atmosphere. By setting up virtual light sources for natural light and interior lights we were able to replicate how the building would appear at different times of day or under different conditions. This was important as the model was used for simulations or visualizations of emergency scenarios for active shooter response training.

Figure 2 shows the initial 2D models of the campus building. We exported the 2D model into Sketch-Up to create a 3D model of the building. As shown in Figs. 3, the building is large, and creating the 3D models along with adding textures, furniture, and other details required significant effort.

3.2 Phase 2: Unity Integration

In the second phase, the Sketch Up model was imported in Unity 3D, which is a 3D gaming engine. Additional elements for interactivity such as opening the door and windows, proximity triggers, etc. were integrated in the VR environment. The building was

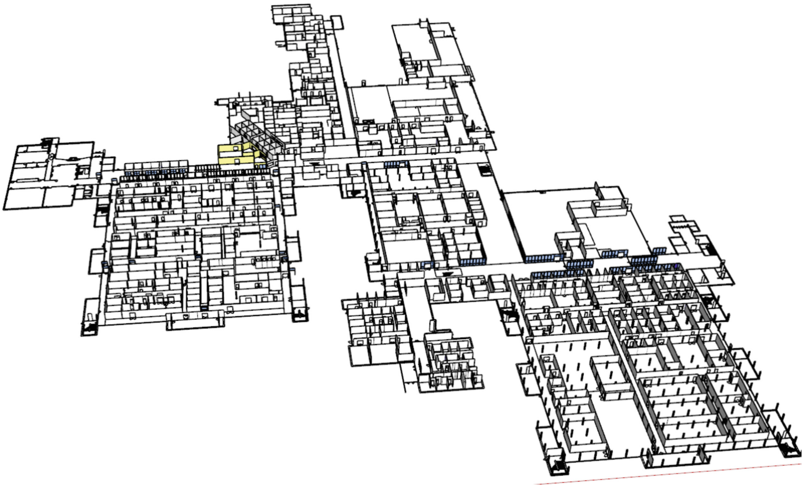


Fig. 2. 2D model of the UNTY building exported to sketch up (1st floor Plan).

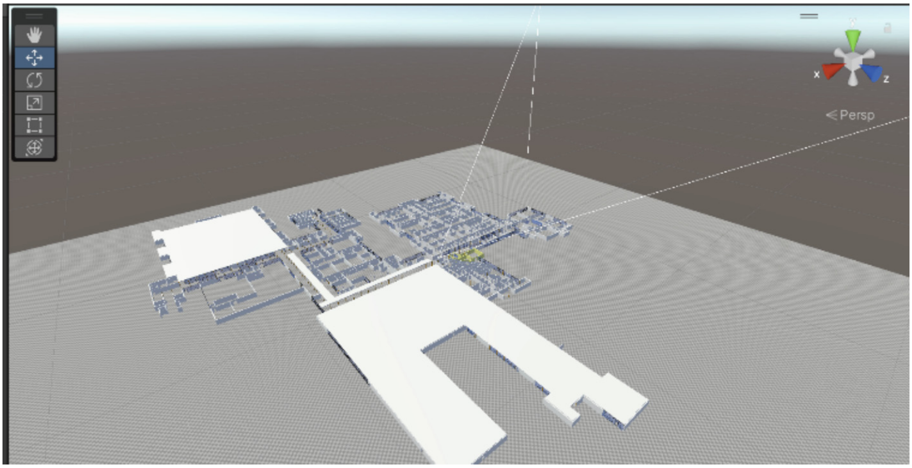


Fig. 3. Initial modeling: 2D model of the campus building exported to sketch up and extruded to create a 3D model (2nd floor plan).

modeled to scale. During this stage, we incorporated C# behavior scripts, looping, and key triggered animations. The use of smoke and fire as well as some consistent colored light flickering functionality was implemented through C# programming. The modeled virtual agents in the environment have a waypoint algorithm attached as a component to enable them to navigate toward their goal in the environment. The combined interactions of the agents and the environmental hazard created a more realistic experience for both immersive and non-immersive participants.

3.3 Phase 3: GUI and User Interaction Phase

Once the 3D model was fully integrated in Unity 3D, the focus shifted to designing user interfaces (UI) and enabling interactions within the virtual environment. Key UI elements, such as “Start,” “Game,” and “End” panels, were developed. A third-person player controller was introduced, allowing users to explore the environment and navigate seamlessly through mouse and keyboard in the non-immersive environment. Additional interactive elements, including functional doors and non-player characters (NPCs) or AI agents, were incorporated to create a dynamic, immersive, and lifelike virtual world.

Fig. 4. AI behavior for NPCs in the active shooter environment.

This phase also involved implementing algorithms for NPCs or AI agent behavior in an active shooter events environment. We present two ways of modeling user behavior. First, by defining rules for AI agents or NPCs (Non-Player Characters). Second, by providing controls to the users-controlled agents or PCs (Player characters) to navigate in the VR environment as autonomous agents with a keyboard/ joystick or with an immersive VR headset. The user-controlled agents can enter the CVE and can respond to emergencies like active shooter events, bomb blasts, fire, and smoke. We have presented a multi-user virtual reality (VR) platform where experiments for active shooter response can be conducted using computer-controlled (AI) agents and user-controlled agents. As shown in Figs. 4 and 5 we have already implemented behavior for NPCs in the active shooter environment at the campus building. We have modeled the following behaviors for computer-controlled agents (AI agents) so that they can interact with user-controlled agents in a CVE.

- Hostile
- Non-hostile

- Selfish
- Leader-following

3.4 Phase 4: Photon Implementation Phase

This phase utilized Photon Unity Networking (PUN), a robust networking framework compatible with Unity, to enable multi-user functionality. PUN allows up to 20 users to interact simultaneously within the environment without incurring operational costs. Custom scripts were developed for room management, player synchronization, NPC behavior synchronization, and UI functionality, ensuring smooth and consistent interactions for all active users. This phase was critical for achieving the collaborative, multiplayer aspect of the virtual environment. C# scripts were incorporated for the implementation of a PUN system that allowed multiple users to collaborate and communicate with one another. The users were able to create a room on the server using a unique application ID. Other users as clients were also able to join the room to participate in the active shooter response environment for campus building. The photon network in Unity 3D allowed all users to view and interact with other user-controlled agents in real-time.

3.5 Phase 5: VR Integration Phase

The final phase involved integrating virtual reality (VR) support using the MetaXR SDK. This enabled VR devices, such as the Meta Quest 3, to interact with and navigate the virtual environment. Dedicated VR-compatible UI elements were designed and implemented to enable operation with these devices.

Fig. 5. Developed active shooter environment based on run. Hide, and fight.

The VR integration expanded the project's potential applications by providing an immersive and intuitive experience, allowing users to interact with the environment naturally through VR input systems. The collaborative immersive environment was implemented in Unity 3D and is based on run, hide, and fight approach for emergency response.

Figure 5 shows our developed CVE environment for active shooter events using Meta Quest 3 touch controllers for the course of action, visualization, and situational awareness for active shooter events.

4 Evaluation of Active Shooter Response Training Environment

A user study was conducted to evaluate the immersive VR training environment with a total of 195 participants at the university campus building (refer Figs. 3, 4, 5 and 6). Phase 1 of the user study included 80 participants whereas Phase 2 of the user study included 115 participants. Each session included 4 participants in the user study. Phase 1 user study included all participants in a multi-user VR environment using a monitor, mouse, and keyboard (non-immersive). On the other hand, phase 2 user study included 2 users on Meta Quest 3 (immersive environment) and 2 users on computer and keyboard (non-immersive) in the same multi-user environment. The experiment sought to examine how the proposed active shooter multi-user VR environment influenced participants' understanding of the safest actions to take during an active shooter situation. The post-evaluation includes questions on the Group Environment Questionnaire (GEQ) [18], Presence Questionnaire (PQ) [19], System Usability Scale (SUS) [20], and Technology Acceptance Model (TAM) Questionnaire [21]. Figure 6 shows the user study conducted for the evaluation of active shooter response training environment.

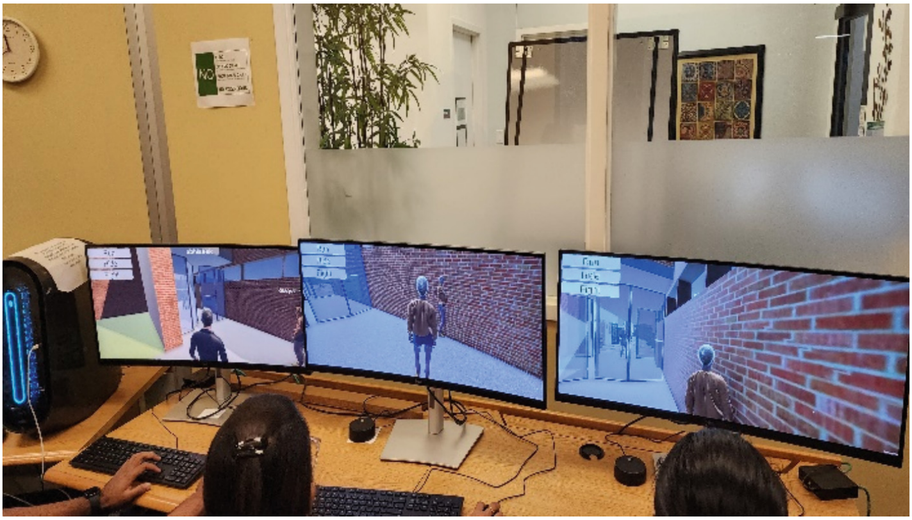


Fig. 6. User study in a multi-user VR environment or collaborative VR environment.

Figure 7 illustrates engagement in the environment by major and gender, with fields like Business Analytics, Data Science, Information Science, and Health Informatics showing the highest levels of engagement. Gender had little impact, though females in Health Informatics were slightly more engaged.

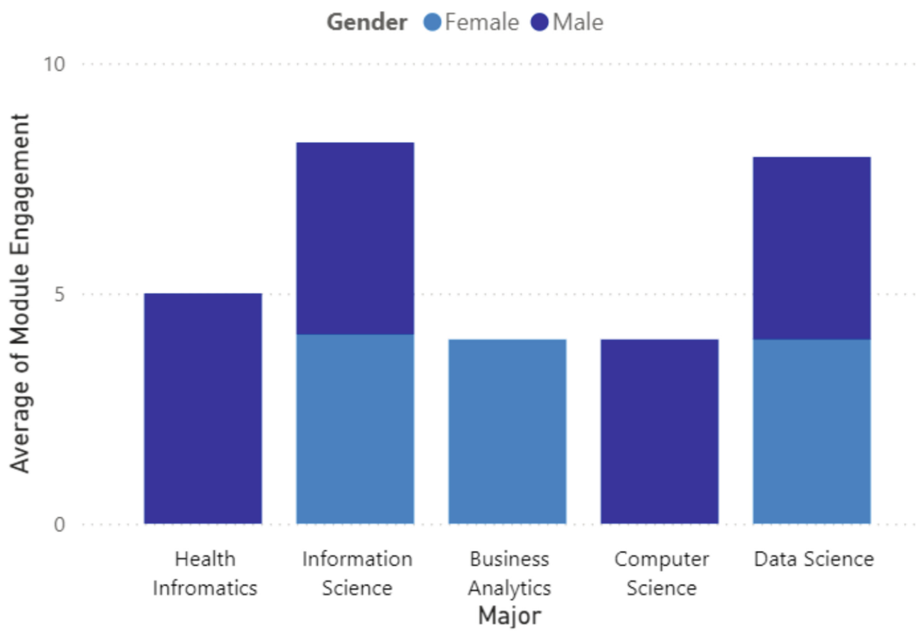


Fig. 7. Changes in Attitude Scores by Age Group.

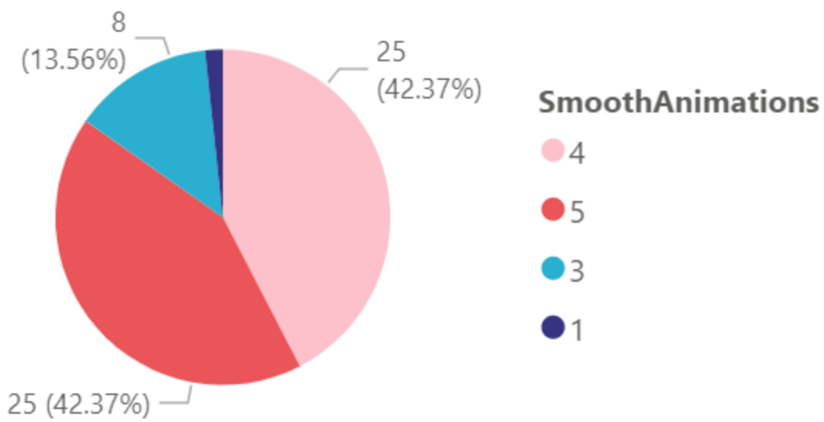


Fig. 8. Smooth Animations Needed.

Figure 8 shows that 43% of respondents believe smooth animations are necessary. This means a little less than half of the survey participants found smooth animations to be an important feature.

As shown in Fig. 9, respondents rated the VR environment highest for its ability to help them achieve safe evacuation (around 6.5 on a 7-point scale). Finding an evacuation route quickly using the active shooter response training environment also received a high

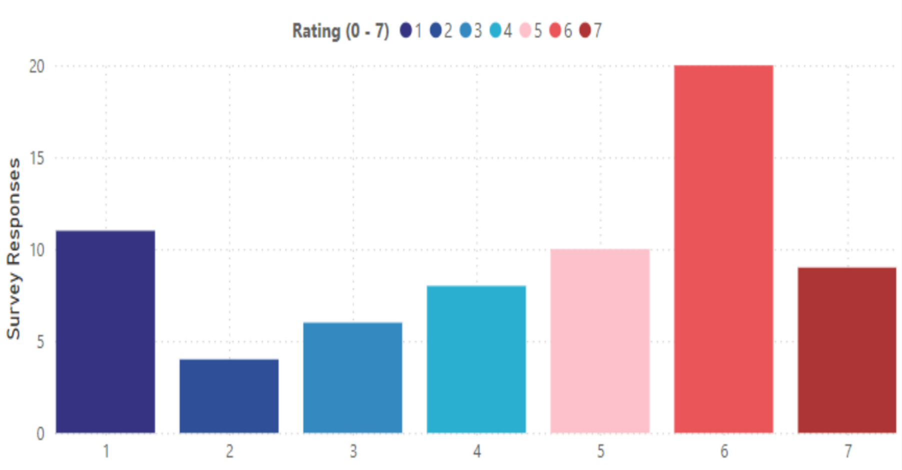


Fig. 9. Usefulness of VR environment to accomplish evacuation safety.

average rating (around 6). These findings suggest the active shooter response training environment effectively addresses core user needs in emergency situations.

Fig. 10. Overall usefulness in multi-user VR environment or collaborative VR environment.

Figure 10, shows the active shooter response training environment's usefulness varied based on gender, age, major, and academic classification. Both male and female respondents found it useful, with males rating it slightly higher. Older respondents found it more useful, while the Computer Science and Health Informatics majors rated it most favorably. Juniors and seniors rated it more positively.

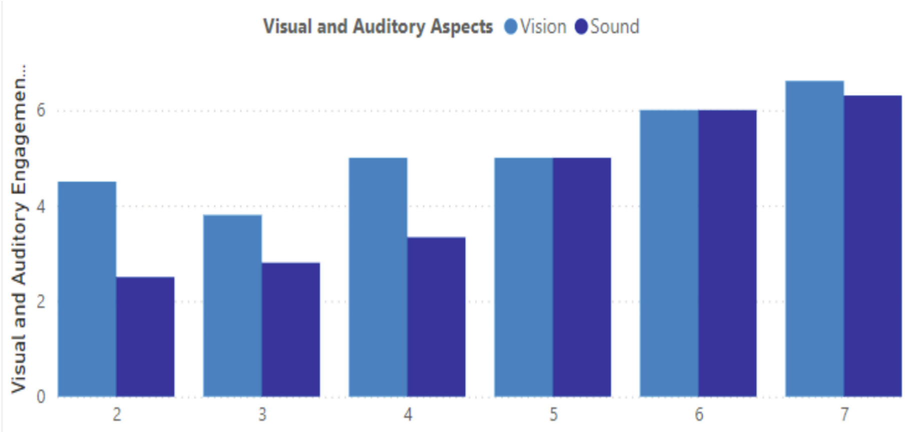


Fig. 11. Visual and auditory engagement aspects

Figure 11, shows that respondents found auditory aspects much more engaging than visual ones. The virtual environment appears to be particularly effective in delivering immersive auditory experiences, while visual engagement falls behind. It was suggested that designers and developers should focus on achieving a balance between both sensory elements to improve the overall user experience. This means ensuring that visuals are as captivating and enriching as the auditory components to create a more immersive and enjoyable virtual environment for users.

5 Conclusions

This paper has presented a multi-user virtual reality (VR) platform where experiments for active shooter response can be conducted using computer-controlled (AI) agents and user-controlled agents. This multi-user VR platform is fully immersive with the use of Meta Quest 3 and touch controllers. It can also be used as non-immersive desktop version through the use of a monitor, mouse and keyboard. The multi-user VR environment is set up using photon unity networking on the cloud and users can participate in the active shooter training drill which leads to considerable cost advantages over large-scale real-life exercises. Studying human behavior during emergencies is often challenging due to the complexity of the scenarios that need to be simulated. Immersive virtual reality provides the opportunity to conduct such human behavior experiments without putting participants at risk. User computer-controlled agents or AI agent's behavior was implemented using behavior trees within the Unity game engine.

A user study was conducted to evaluate the immersive VR active shooter training environment with 195 participants. The evaluation of the immersive VR active shooter training environment included post survey questions from Group Environment Questionnaire (GEQ), Presence Questionnaire (PQ), System Usability Scale (SUS), and Technology Acceptance Model (TAM) Questionnaire. The findings suggest that participants' knowledge, intrinsic motivation, and self-efficacy showed a significant increase immediately following the training. The results indicate that most users felt a stronger sense of presence when engaging with the immersive emergency response training environment designed for an active shooter evacuation scenario. By utilizing the immersive VR platform, trainees enhance their spatial awareness and gain a better understanding of how to navigate the building during high-stress situations, ultimately improving their chances of survival and successful evacuation. The results from the user studies indicate that participants experienced a notable increase in their knowledge, intrinsic motivation, and self-efficacy right after the training. The results show that most users felt that the sense of presence increased when using the immersive emergency response training active shooter environment.

Acknowledgments. This work is funded by the NSF award No. 2319752 and NSF award No. 2321539. This research was approved by the Institutional Review Board (IRB) at the University of North Texas, under IRB protocol number # IRB-23-170, and all participants provided informed consent in accordance with ethical guidelines.

References

1. Sharma, S., Park, J.W., Morris, B.T.: Immersive security personnel training module for active shooter events. In: Proceedings of the IS&T International Symposium on Electronic Imaging (EI 2023) in the Engineering Reality of Virtual Reality Conference, 15–19 January 2023. https://doi.org/10.2352/EI.2023.35.12.ERVR-217
2. Sharma, S., Bodempudi: Immersive virtual reality training module for active shooter events. In: Proceedings of the IS&T International Symposium on Electronic Imaging (EI 2022), in the Engineering Reality of Virtual Reality, pp. 299-1–299-6, January 2022. https://doi.org/10.2352/EI.2022.34.12.ERVR-299
3. Sharma, S., Bodempudi, S.T., Scribner, D., Grazaitis, P.: Active shooter response training environment for a building evacuation in a collaborative virtual environment. In: IS&T International Symposium on Electronic Imaging (EI 2020), in the Engineering Reality of Virtual Reality, Burlingame, California, 26 January–30 January (2020). https://doi.org/10.2352/ISSN.2470-1173.2020.13.ERVR-223
4. Sharma, S., Ali, S.: Multi-agent crowd simulation in an active shooter environment. In: Chen, J.Y.C., Fragomeni, G. (eds.) Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry. HCII 2022. Lecture Notes in Computer Science, vol. 13318. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-06015-1_8
5. Harris, D., et al.: Exploring the role of virtual reality in military decision training. Front. Virtual Real. **4** (2023). https://doi.org/10.3389/frvir.2023.1165030
6. Zhu, R., Lucas, G., Becerik-Gerber, B., Southers, E., Landicho, E.: The impact of security countermeasures on human behavior during active shooter incidents. Sci. Rep. **12**(1) (2022). https://doi.org/10.1038/s41598-022-04922-8
7. Dailey, S., Laskey, K.: Student safety and casualty mitigation during an active school shooter simulation: an exploratory study. Safer Commun. **22**(4), 217–234 (2023). https://doi.org/10.1108/sc-08-2022-0036

8. Dillard, C.: Slow breathing reduces biomarkers of stress in response to a virtual reality active shooter training drill. Healthcare **11**(16), 2351 (2023). https://doi.org/10.3390/healthcare11162351
9. Martaindale, M., Blair, J.: The evolution of active shooter response training protocols since columbine: lessons from the advanced law enforcement rapid response training center. J. Contemp. Crim. Justice **35**(3), 342–356 (2019). https://doi.org/10.1177/1043986219840237
10. Sharma, S., Otunba, S.: Collaborative virtual environment to study aircraft evacuation for training and education. In: Proceedings of IEEE, International Workshop on Collaboration in Virtual Environments (CoVE - 2012), as part of The International Conference on Collaboration Technologies and Systems (CTS 2012), Denver, Colorado, USA, pp. 569–574, 21–25 May 2012
11. Sharma, S., Vadali, H.: Simulation and modeling of a virtual library for navigation and evacuation. In: MSV 2008 - The International Conference on Modeling, Simulation and Visualization Methods, Monte Carlo Resort, Las Vegas, Nevada, USA, 14–17 July 2008
12. Sharma, S., Jerripothula, S., Mackey, S., Soumare, O.: Immersive virtual reality environment of a subway evacuation on a cloud for disaster preparedness and response training. In: Proceedings of IEEE Symposium Series on Computational Intelligence (IEEE SSCI), Orlando, Florida, USA, pp. 1–6, 9–12 December (2014). https://doi.org/10.1109/CIHLI.2014.7013380
13. Sharma, S., Devreaux, P., Scribner, P., Grynovicki, J., Grazaitis, P.: Megacity: a collaborative virtual reality environment for emergency response, training, and decision making. In: IS&T International Symposium on Electronic Imaging (EI 2017), in the Visualization and Data Analysis, Proceedings Papers, Burlingame, California, pp. 70–77(8), 29 January–2 February (2017). https://doi.org/10.2352/ISSN.2470-1173.2017.1.VDA-390
14. Sharma, S., Frempong, I.A., Scribner, D., Grynovicki, J., Grazaitis, P.: Collaborative virtual reality environment for a real-time emergency evacuation of a nightclub disaster. In: IS&T International Symposium on Electronic Imaging (EI 2019), in the Engineering Reality of Virtual Reality, Hyatt Regency San Francisco Airport, Burlingame, California, pp. 181-1–181-10(10), 13 January–17 January 2019
15. Sharma, S., Jerripothula, P., Devreaux, P.: An immersive collaborative virtual environment of a university campus for performing virtual campus evacuation drills and tours for campus safety. In: Proceedings of IEEE International Conference on Collaboration Technologies and Systems (CTS), Atlanta, Georgia, USA, pp. 84–89, 01–05 June 2015. https://doi.org/10.1109/CTS.2015.7210404
16. Liu, W., Zeng, N., Pope, Z., McDonough, D., Gao, Z.: Acute effects of immersive virtual reality exercise on young adults' situational motivation. J. Clin. Med. **8**(11), 2019 (1947). https://doi.org/10.3390/jcm8111947
17. Phillips, S.: Active shooter incidents: training, safety, culture and officers' support for prioritizing victims' lives above their own. Policing Int. J. **47**(4), 529–544 (2024). https://doi.org/10.1108/pijpsm-11-2023-0151
18. Whitton, S., Fletcher, R.: The group environment questionnaire: a multilevel confirmatory factor analysis. Small Group Res. **45**, 68–88 (2013). https://doi.org/10.1177/1046496413511121
19. Witmer, B.G., Jerome, C.J., Singer, M.J.: The factor structure of the presence questionnaire. Presence **14**(3), 298–312 (2005). https://doi.org/10.1162/105474605323384654
20. Lewis, J.: The System Usability Scale: Past, Present, and Future. Int. J. Hum.-Comput. Interact. 1–14 (2018). https://doi.org/10.1080/10447318.2018.1455307
21. Ma, Q., Liu, L.: The Technology Acceptance Model (2005). https://doi.org/10.4018/9781591404743.ch006.ch000

Procedures Training in VR and The Role of Episodic Memory: Literature Review and Synthesis

Nathan A. Sonnenfeld(✉), Vera Daniliv, and Florian G. Jentsch

University of Central Florida, Orlando, FL 32816, USA
Nathan.Sonnenfeld@ucf.edu

Abstract. There has been optimism that virtual reality (VR) may be a suitable medium for training procedures in aviation. However, the mechanisms by which immersion may facilitate or hinder the training of flightcrew-relevant procedures using VR remain underexplored. One important concept is that of episodic memory (EM). Through representational and non-representational mechanisms of cognition, the EM system may have a central role in grounding and framing VR training experiences, thereby facilitating pilots' acquisition and use of knowledge, skills, and attitudes (KSAs) to support flight operations. The current review studies frameworks and concepts related to the cognitive processes underpinning the learning of procedures in VR simulations. Specifically, it focuses on the role, function, and value of the EM system as an explanatory mechanism for a range of pertinent theories and findings, and bridging gaps among constructs and perspectives between and among scales of cognition. We posit here that interventions before, during, and after VR experiences that target the mechanisms of EM may have downstream benefits on spatial cognition, tacit knowledge, and decision-making during flight deck operations. Furthermore, whereas extant training frameworks do not delineate EM or prescribe its measurement, we provide a framework of methods and techniques for assessing the unique contribution of the EM system to learners' memory of training events. Consideration of the role, function, and value of the EM system provides a more precise understanding of how constructs such as immersion may facilitate or hinder the training of procedures using VR simulations.

Keywords: Episodic Memory · Virtual Reality · Procedures Training

1 Introduction

Extended reality ('XR') technologies, particularly virtual reality (VR) simulations, have been viewed favorably for their potential role in flightcrew training, for example, to supplement classroom instruction, to rehearse tasks before simulator training, or to provide additional practice following simulator training [15, 68]. However, VR may not be suitable for all training objectives. Available evidence suggests that VR may be suitable for training procedures. However, there seems to be a lot of variability in the

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 233–251, 2025.
https://doi.org/10.1007/978-3-031-93715-6_15

effectiveness of VR interventions for procedures training. Meta-analyses have shown, for example, that VR is more effective for physical versus cognitive tasks [4, 13, 32, 35, 36]. Findings also tend to support VR as effective for achieving psychomotor and affective outcomes, including important variables such as engagement and self-efficacy. Indeed, anecdotal data suggest that VR is particularly suitable for, and involved processes of embodied cognition and muscle memory [35, 44, 57]. Conversely, VR tends to have negative effects concerning cognitive load and cybersickness, which are higher under VR [5, 11, 61], and there is mixed evidence of VR's effectiveness for training knowledge and cognitive skills, or for their retention.

Effect sizes vary greatly across effects and dimensions, especially when looking beyond simple pre-post gains. Indeed, the interactive nature of VR simulations may have a greater impact on training outcomes than VR's typical hallmark, i.e., immersion. Thus, there are questions related to what added value VR provides over other forms of interactive simulation. Further, VR may be another case where the training medium itself may be of less importance than all the other instructional design considerations embedded in the training [12, 73]. Such other considerations might include the type of training objectives that are targeted, the appropriateness of instructional elements embedded in the system, the use of simulation, the time-on-task spent by trainees, and the actual content of training.

While a diverse range of factors and instructional strategies have been empirically verified to impact learning in VR—at the cognitive level, the functions of the declarative, procedural, and working memory systems fail to fully account for these observed effects. With these gaps driving our current program of research, we suggest that the role and value of episodic memory (EM) have been overlooked as explanatory mechanisms for the efficacy of VR for learning and training procedural tasks.

1.1 Purpose and Approach

Following a synthesis framework by Jaakkola [33], this review identifies similitudes among theoretical models related to the cognitive processes that underpin the learning of procedures in VR simulations. We establish a basis for investigating the role, function, and value of the EM system as an explanatory mechanism for the efficacy of VR simulations for training procedures. Our goal is to structure the fragmented topic of learning procedures using VR simulations. We specifically provide theoretical grounding for the notion that interventions targeting EM before, during, and after VR simulations may have downstream benefits on important outcomes, such as spatial cognition, tacit knowledge, and decision-making, especially for flight deck operations. Furthermore, whereas extant training frameworks do not delineate EM or prescribe its measurement, we outline a framework for assessing the unique contribution of EM to training outcomes. We hope to provide a more precise understanding of how VR simulations may facilitate or hinder the training of procedures.

2 Literature Review and Synthesis

2.1 What Is a Procedure?

Procedures, or procedural tasks, may be defined as ordered sets of goal-directed actions performed in a specific environmental and situational context [9, 39]. They are a universal and critical element of work, particularly in aviation where adherence to procedures is the foundation for safety during flight operations [16, 18, 56, 66]. Designation of a procedure into manageable steps increases its comprehensibility, particularly when those tasks are open-ended, temporally restrictive, or complex to execute [56]. When thoughtfully designed, externalized (e.g., checklists), and internalized (i.e., as a script), they serve as cognitive aids to scaffold performance [56, cf. 60]. However, designated procedures often lack information about system states or environmental factors—the procedure does not unburden the pilot from the responsibilities of decision-making—for example, of adherence, of ensuring the procedure is appropriate to the context, of the expertise to execute actions, or of resolving the underlying situation [56]. That is, the procedure itself is not sufficient for ensuring flight safety—flightcrew members must be trained to perform it across operational contexts.

Performance of a procedure relies upon the acquisition of a range of subordinate knowledge, skills, and attitudes (KSAs). Pilots must acquire declarative knowledge of the designated steps of the procedure and explicit conditional rules for its execution; mental models integrating this knowledge with that of aircraft systems and functional interdependences, and cognitive skills to diagnose situations, appraise actions, and decide how procedures must be applied or amended in each situation [52]. Domain-specific procedural skills ("procedural knowledge") and requisite psychomotor skills translate this declarative knowledge of "how" to execute a procedure into an action response. Task complexity may vary in number of steps, interdependencies, required planning and flexibility, decision points, constraints, and involved personnel [19, 91], requiring other perceptual, cognitive, or affective skills [17, 77]. Even simpler procedures may involve multiple domains and levels of KSAs. Executing procedures often goes beyond what may be explicitly documented or recited, and in this respect, the concept of "procedural knowledge" may disserve the training sciences. Deconstruction of the declarative knowledge of documented steps/rules of a procedure versus the unverbalized skill of its execution may be sufficient for routine situations, yet in non-normal situations, pilots must also exhibit adaptive expertise, drawing on prior experiences and transfer those experiences to novel situations [31]. As supported by expert decision-making literature, this adaptive expertise relies on the functions of the EM system, and knowledge not clearly attributable to either declarative or procedural memory [31, 37, 74]. Sonnenfeld et al. [74] suggested that we might define this as tacit knowledge—imperfectly articulable knowledge of situations and spatiotemporal relations accessible through analogical reasoning, metaphor, narrative, or situated interactions as manifest in actions, outcomes, or shared understanding [23, 43, 58, 82].

2.2 How Are Procedures Trained?

Experiential Learning and Simulation Training. Experiential learning may be viewed as a generalizable and effective approach for training of actions and effective the development of procedural skills [10, 27], In most uses, it refers to the concept of praxis—of learning from real-world experience and trial and error [27]. Kolb's synthesis of Experiential Learning Theory [38] describes this process as an adaptation to conflicts within the environment, grounded in concrete experience, reflective observation, abstract conceptualization, and active experimentation. While advocates originally dismissed the value of technology-based implementations in delivering the uncertainty and conflict of real-world experiences [27], instructional designs of training simulations are often grounded in experiential learning as a justification and guiding framework for their efficacy. As discussed by Sonnenfeld et al. [72], learning at the task- and cognitive scales may be viewed in alignment with the activities described by experiential learning theory at the social scale [53], in VR simulation training contexts.

Simulation training is one of the primary approaches used to implement experiential learning for the training of procedures and other skills, with the literature supporting its appropriateness and comparative effectiveness versus non-simulation approaches for the training of procedures and other skills across safety critical domains [16, 73]. Simulation training broadly refers to the application of virtual environments for practicing tasks and skills, and can "augment, replace, create, and/or manage a learner's experience with the world by providing realistic content and embedded instructional features" [7, p. 1417]. For brevity, we merely acknowledge here, but do not delimit, the rich bodies of literature on simulations and on the evaluation of their training effectiveness which has informed our program of research [73, 75]. While the implementation of simulation training may vary (e.g., instructor guidance, fidelity, adaptivity), the use of extended reality (generally) and VR (specifically) have shaped conversations about the value of certain instructional and experiential factors, and how the effectiveness of such training should be evaluated.

Virtual and Extended Reality. The flightcrew training community has long been interested extent that lower-fidelity simulations may be used to supplement or otherwise optimize traditional procedures training, using flight simulation devices and paper tigers [15, 68, 73]. Historically, the aviation industry has classified these simulation technologies with respect to the degree of physical fidelity they afford (e.g., 14 CFR Part 60 Appendices A-B). Permeating the academic literature, however, are concepts for a spectrum of configurations in which users' perceptions are extended beyond their physical ambient environment. Drawing on early concepts of this space [51], the literature goes into great depth in differentiating and defining simulation technologies and associated experiences based on a range of considerations, including hardware used, interactions, and affordances [46, 51, 64, 69]. For this research, we operationalize these media by their affordances in terms of the stimuli being perceived and the ambient environmental context:

- An experience that primarily affords perception of physical stimuli within an ambient physical environment denotes **physical reality** as the delivery medium.

- An experience that primarily affords perception of virtual stimuli within an ambient physical environment denotes **augmented reality** (AR) as the delivery medium.
- An experience that primarily affords perception of physical stimuli within an ambient virtual environment denotes **mixed reality** (MR) as the delivery medium.
- An experience that primarily affords perception of virtual stimuli within an ambient virtual environment denotes **virtual reality** (VR) as the delivery medium.

This spectrum of reality-to-virtuality is known as the extended reality (XR) spectrum [64, 78, cf. 51]. Such experiences provide a range of affordances, such as immersion, interactivity, involvement, and presence [45, 49], They also afford the acquisition and practice of KSAs in an operationally relevant context. Several factors impact on learning and training using XR and simulation technologies. The type (e.g., reactions vs. learning vs. behaviors) and domain of outcomes (e.g., knowledge vs. skills), the task-technology fit (e.g., physical vs. cognitive), experiential factors (e.g., cognitive demand, cybersickness), and instructional factors (e.g., simulation, feedback) have all been identified in meta-analyses and supporting empirical work as impacting the effectiveness of VR [4, 13, 15, 32, 35, 36, 44, 57].

Actionable and generalizable information from these studies remains limited, particularly for flightcrew training contexts, and several questions arise from the consistencies and discrepancies among these studies—why is VR more effective for physical/psychomotor tasks than cognitive skills/knowledge? Why is there such variance in the efficacy of different VR simulations? Despite decades of research, there remains no singular theory that provides a rationale for these findings and discrepancies, at a sufficient level of scope to account for interactions across cognitive, task, and social scales of learning and training using VR simulations. Relevant theories either focus on a single scale of cognition [38, 41, 45, 87], prescribe principles without a unifying explanatory theory [79], or are generalized from studies without immersion or simulation [47, 50]. While our prior work [74] has generally delineated the role of EM in experiential learning via simulations and advocated for consideration of this concept in training, our current research explores how EM may provide an explanatory mechanism for VR simulation training across scales of cognition. In the next section, we narrow our granularity from the scale of training to that of learning.

2.3 How Are Procedures Learned?

Theoretical frameworks and findings across the cognitive sciences have largely converged on several relevant features of the cognitive architecture underlying learning [40, 41], at least from a traditional computational-representational view [cf. 20]. This consensus generally extends to the functions of several cognitive systems and processes related to learning—here defined as a change in long-term memory [59], with long-term memory (LTM) subsequently defined as a higher-order cognitive system with a neural basis that functions to consolidate, maintain, and retrieve information at an indefinite scale, subject to decay/loss [65, 81, 84]. The major features of this architecture are outlined below.

Learning and Long-Term Memory: Common Model of Cognition. We ground this synthesis in a cognitive architecture, as learning exhibited over the social scale of human action may be assumed to arise from the accumulation of learning occurring incrementally across cognitive- and rational-scale experiences [40, 41, 53], accepting Anderson's [2] decomposition and relevance theses. That is, learning across the scale of hundreds of hours is driven by that occurring at milliseconds; similarly, the microstructure of memory impacts the measurable processes of experience and KSA acquisition. The Common Model of Cognition (CMC; 40; 41] provides a succinct coverage of the relevant subsystems and processes involved in learning, including perception, representation and chunking, declarative and procedural memory, and motor response [40, 41]. As a product of the comparison and synthesis of other prominent cognitive architectures, the CMC is neither prescriptive nor comprehensive, intended as a basis for a shared ontology and iterative refinement [40, 41]. This makes the CMC a suitable grounding for synthesizing of the concepts of interest here.

It also has its limitations; citing a lack of clear consensus, the authors [40, 41] acknowledge omission of episodic memory, metacognition, social cognition, and other learning processes—including the function of goals in working memory (WM) and action as addressed by other architectures [14, 89]. Delimited to the level of deliberate action at a cognitive scale, the CMC does not address neural, task, or social scale processes as addressed within other frameworks [cf. 2, 34]. Despite these and other limitations, the CMC provides a common frame of reference for the basic structures, functions, and relationships among the perception, WM, LTM, and motor systems—for how information relevant to the instruction and execution of procedures may be perceived, encoded, retrieved, and acted upon. It also provides a basis for discussing the strengths and limitations of other models with implications for learning procedures using VR, as we review and synthesize in the following sections.

Attention, Working Memory, and Cognitive Demand. Several aspects of the CMC pertinent to the learning of procedures using VR are clarified within models of the phenomena of interest. In this section, we provide a basis for discussing the role of EM in VR by focusing on the mechanisms of attention within the Model of Human Information Processing (HIP) [e.g., 87], the multi-component Model of Working Memory (M-WM) [e.g., 6], and the Cognitive Theory of Multimedia Learning (CTML) [47].

Attention: Human Information Processing Model. The HIP [87] details an architecture sharing components with the CMC, but from the perspective of information processing during performance. Attention, here, is more than a generic source of constraint [87]. Selective attention influences which environmental stimuli are perceived. Focused attention impacts what information is maintained and encoded via WM. Divided attention affects response selection and execution through the prioritization of information [87]. Through these mechanisms, the HIP details how performance decrements result when concurrent task demand exceeds available attentional resources (i.e., overload), depleting residual capacity for managing performance (e.g., during a learning task).

Working Memory: Multi-Component Model. The M-WM [6, 87] details how modality-specific information may be concurrently processed during learning. WM functions to store and manipulate information in support of goal-directed activity [55], after a series of iterations, the theory depicts WM as composed of four major components [6, 87]:

- A central executive allocates attention toward goal-relevant information for encoding, selects goal-relevant strategies for encoding, facilitates consolidation within LTM, and directs attention to maintain or retrieve information after encoding.
- An episodic buffer passively integrates modality-specific information into chunks and facilities conscious access to information within WM, varying in efficiency and the suppression of goal-irrelevant information by the central executive.
- A visuospatial sketchpad that encodes visual, spatial, and haptic features and functions subordinate to the episodic buffer; and
- A phonological loop that stores and integrates auditory-related information (e.g., sound, speech, nonverbal) with an articulatory loop for vocal rehearsal.

WM & Cognitive Demand: Cognitive Theory of Multimedia Learning. Lastly, CTML [47] has largely subsumed the largest contributions of competing frameworks under its umbrella of principles and supporting guidance [47, 50]. For example, principles of cognitive load theory associated with the optimization of the intrinsic, extrinsic, and germane loads experienced by learners are the essential, extraneous, and generative processing of the CTML [47]. That is, respectively, cognitive demand imposed by the instructional objective, and inefficiencies or efficiencies in the instructional design of the experience [47, 50]. CTML delineates how the modality of information delivery impacts the modality-specific processes of the M-WM a range of instructional principles have been introduced under CTML, with recent additions addressing factors such as self-regulation, immersion, and embodiment [47], albeit without as much explanatory value as frameworks targeting those individual phenomena [cf. 48, 88, 89].

System and Experiential Factors: Cognitive-Affective Model of Immersive Learning. These models of attention, working memory, and cognitive demand provide a rich but incomplete basis for understanding the function and value of the EM system for procedures training using VR, however other sources of literature (e.g., 4E cognition, VR research) are available to address such gaps [50]. In this section, we reiterate a few constructs key to the use of VR as grounded by Makransky and Petersen's [45] Cognitive Affective Model of Immersive Learning (CAMIL).

Immersion. Following Slater and Wilbur [71], immersion within an experience may be defined as an objective property of a mediating system descriptive of the extent to which the system provides an inclusive (e.g., the extent to which non-diegetic sensory information is excluded), extensive (e.g., the extent of sensory modalities accommodated), surrounding (e.g., the extent to which the system matches the natural sensory field), and vivid (e.g., the extent of resolution and richness of sensory information) illusion of reality to the users' senses. The degree of immersion afforded by a system is continuous rather than binary—discrimination between head-mounted displays (HMDs) as "Immersive VR' and monitors as "Non-Immersive VR" obfuscates these factors. While various hardware and software factors contribute to the degree of immersion afforded by a system, our focus I is not on the sources, bur rather the effects of immersion on procedural training [48].

Presence. Presence, the subjective sense of being within an environment irrespective of physical location, is a core concept in the analysis and evaluation of VR simulations

[70, 90]. Descriptive of a state of consciousness, presence (physical/spatial) may be conceptualized as the attribution of attention to the stimuli from an ambient environment [72, 86], a function of the immersion afforded by a media [70, 71] and the subjective involvement of the user in the experience [90]. That is, presence can be experienced from non-immersive media such as narrative provided sufficient involvement. attention as being requisite or strongly associated with presence [50, 72], and parallels may be drawn between the functions of attention in the HIP [87] and the dimensions of virtual experience—such that divided attention may have similitude with focus (i.e., presence-absence), selective attention with locus (i.e., reality-virtuality), and focused attention with sensus (i.e., conscious-unconscious) [86].

Other Affective Factors. Several other experiential factors and outcomes have been found to moderate the learning of procedures under varying conditions of immersion, including embodiment, motivation, self-regulation, and self-efficacy [45]. For brevity, we identify these as factors to be accounted for investigations of the effects of immersion on memory formation among the declarative, procedural, and episodic systems, to be explored through further refinement.

Cognition in the (Virtual) Environment: 4E Cognition. Despite general consensuses in the computational-representational paradigm, these notions of cognition and learning are incomplete. They are targets of refinement by alternative perspectives such as 4E cognition, ecological psychology, and dynamical systems theory, which contribute among other concepts an articulation of the dynamic coupling of the brain-body-environment system [20, 49, 54]. Generally, these bodies of literature emphasize: (1) the situated and ecological role of affordances (i.e., embedded cognition) [24, 28, 54], the body as a constraint, distributor, and regulator of cognitive processes (i.e., embodied cognition) [49, 88], (3) the role of sensorimotor activity, volition, and re-enactment in memory and learning (i.e., enactive cognition) [25, 54], and (4) the functions of interactivity and externalization in the distribution of cognitive processes during learning and performance (i.e., extended cognition) [60, 80]. While our research does not attempt to isolate these facets of cognition in the use of VR, a review of the scope of literature on 4E cognition is sufficient to acknowledge its premises from a pluralistic perspective [20] such that non-representational and representational cognitive phenomena provide viable explanatory principles for the processes involved in learning procedures using VR simulations [49, 74].

From a pluralistic perspective, the CMC provides one side of the account for how the cognitive system may process and respond to ecological information—environmental stimuli in the ambient array, kinematically designating the spatiotemporal dynamics of objects and events [29]. In one view, representation is the form by which ecological information may be communicated across a medium (e.g., energy, neural activity, cognitive structures), irrespective of whether it remains coupled (i.e., direct perception) or decoupled (i.e., abstraction) with the environment [29]; the structures of the CMC designating the cognitive-scale synergies of information processing which functionally constrain degrees of freedom within this architecture, enabling learning. For our framework, we consider that the ecologically rich, instructionally relevant, modality-specific information within the VR simulation may be processed such that representations are grounded, even when decoupled, by embedded affordances. As the CMC does not prescribe the representation of motor systems [40, 41], concepts of brain-body-environment coupling in 4E cognition literature [e.g., 49, 54, 88] provide sufficient grounding for both representational and non-representational mechanisms for behavior in VR simulations, to account for learners' responses to affordances as invariances across the ambient array of information in the virtual training environment [28, 30].

Despite their complementary perspectives, another limitation of the reviewed models (e.g., CMC, HIP, M-WM, CTML, CAMIL) is that they do not explicitly account for the relationship of their mechanisms to concepts typically associated with the use of VR (e.g., immersion, interactivity, presence) [45]. Furthermore, the relationships between these processes and memory formation within a distinct EM system have not been sufficiently explored, despite the occasional nod to such mechanisms (e.g., the episodic buffer). Concepts from 4E cognition may help address such gaps—here, the role of attention in procedures training via VR.

Selective attention pertains to which environmental stimuli are perceived [87]. The ecological perspective (i.e., embedded cognition) offers that a learner will attend to invariances in the ambient array of the virtual environment—invariances containing rich information regarding the plausible actions that the learner may perform within that environment—affordances—given their KSAs and the constraints of their embodiment [28]. Learners' embodiment and available affordances are often perceptibly different in VR than in physical reality [49, 50]. VR, often by technical or instructional necessity, constrains the possible actions the learner may perform, changes the psychomotor inputs associated with certain actions (e.g., locomotion), and provides sensory information at magnitudes lower degree of fidelity than the real world, thus changing the very nature of that ambient array. However, simulations—particularly VR—may also afford actions that would be otherwise impossible to enact within the real world [49, 50]. Use of any VR simulation involves familiarization, with similitude to ecological concepts of attunement (i.e., exploratory actions enacted to determine invariances within the ambient array) and calibration (i.e., exploratory actions enacted to fine-tune perception-action) [30, 50]. Learners may become sufficiently attuned in the locus of their virtual experience, as to selectively attend to available affordances in VR, and to recalibrate in compensation for inadvertent physiological changes in the embodied experience (e.g., cyber sickness, breaks-in-presence) [30, 86]. With a corresponding increase in germane load/processing, balancing that imposed by the intrinsic and extraneous aspects of the learning task [72], a variety of instructional elements (e.g., strategies, features) may be leveraged to scaffold the learners' selective attention to important cues and content [79]. Selective attention manifests in learners' enactment of exploratory behaviors in VR, providing the foundation for the enactment of performatory behaviors [30] in executing procedural skills.

Focused attention influences what information is maintained in WM for encoding [87] with similitude to the essential processing and intrinsic load of CTML [47] and the sensus of experience [86]. As discussed later, this aligns with processes by which episodic elements of goal-relevant experiential information and affordances within VR become integrated and consolidated into the progressively higher-order structures of the

EM system, in adherence to the mechanisms detailed by the CMC [40, 41] and self-memory system [e.g., 14]. Maintaining VR stimuli to the extent of awareness, implies that the information is perceived to be of sufficient salience and relevance for the bottom-up experiential processing of EM [14]. Lastly, the mechanisms of divided attention may be considered to align with how extraneous load impacts the prioritization of information for (internal and external) response selection [87], with concurrence to how invariances across multiple modalities in VR simulations may lead to an excess of extraneous load, constraining on learning outcomes [5, 11, 61, 72]. Without appropriate instructional elements to facilitate generative processing [1, 72], learners may experience a diminished sense of presence (i.e., absence) along the focus dimension of virtual experience [86] with corresponding effects on memory formation [72].

Episodic Memory: Self-Memory System. While the reviewed literature has outlined the basic processes for memory formation relevant to learning procedures, a substantial body of research from the cognitive sciences details the role of a third LTM subsystem, episodic memory (EM), which functions interdependently with these other cognitive systems to facilitate LTM [21, 85]. While a range of recent work has highlighted the role of EM in VR, much of this literature remains constrained to the clinical psychology domain [62, cf. 69], despite a rich body of training science and aviation training literature on cognitive functions which cannot be solely accounted for by the functions of the declarative and procedural memory systems (e.g., vividly recalling prior experiences, mental simulation of future events and action consequences). Given its experiential aspects, consideration of procedures training using VR simulations necessitates a re-examination of the EM system in learning.

Characteristics and Functions of the EM System. As described by Sonnenfeld et al. [74] Episodic Memory (EM) may be concisely defined as "a neurocognitive memory system that enables people to remember past happenings" [85, p. 69], which "receives and stores information about temporally dated episodes or events, and temporal-spatial relations between them [83, p. 223]. The two characteristics of EM most relevant to training are time and context [22]. Regarding time, the EM system allows for the reconstruction of personal experiences from the past (i.e., retrospection), and mental simulation of possible events in the future (i.e., prospection) [22, 84]. Regarding context, the EM system encodes the spatiotemporal characteristics and relations of events [84], due to our sensorimotor embodiment [88], such that spatial cognition may be a feature of EM [3, 21]. It is also often implicated in the structuring of higher-order schemata including mental models, narratives, and self-concepts (e.g., attitudes, beliefs, identity) [3, 14, 74].

Properties and Mechanisms of the EM System. Conway [14] conceptualized EM as being composed of experiential summaries of perceptual and cognitive processing—representative of situated and embodied experiences, such that sensory details may be inhibited (e.g., if not meaningful via relevance to goals) or activated (e.g., environmentally or semantically primed) during reconstruction/recollection with a variable degree of accuracy. Linking 4E cognition to EM theory, our prior work [74] suggested that this property of the EM system results from learners being embedded in a particular environment and context, from embodiment in sensorimotor processing facilitating the encoding and retrieval of EM, and from enaction such that learners with active control over an experience develop richer experiential summaries [49, 67, 72]. These experiential summaries, or episodic elements [14, p. 2308], are interpreted and successively chunked into memories of episodes and events through the frames provided by higher-order structures (e.g., goals, but also mental models, narratives, etc.) [3, 14, 37]. Without the attribution of value (e.g., goal relevance) from these frames, precise details of the experience beyond relative time and context become subject to inhibition and loss. goals provide the context maintaining those memories within a broader frame of reference [14, 22, 84]. Through this process, the EM system generates models of the meaningful sensory, spatiotemporal, and affective elements of the simulated environment and task [3]. These episodic structures have visual representation and perspective, relative to the salience and fidelity of these elements—here, to the fidelity afforded by a given configuration [14, 22]. Propositional knowledge and condition-action structures are assimilated into their respective declarative and procedural systems [22]. Without the attribution of value (e.g., goal relevance) from these frames, precise details of experience beyond relative time and context become subject to inhibition and loss; goals provide the context for maintaining those memories within a broader frame of reference and for facilitating integration in higher-order structures (e.g., mental models, narratives, goals, attitudes [3, 14, 22].

Tacit Knowledge. What remains of EM as the result of these processes may be defined here as tacit knowledge—imperfectly articulable knowledge of situations and spatiotemporal relations accessible through implicit means (e.g., direct perception, intuition, analogy, metaphor, and narrative) as manifest in situated actions, outcomes, or shared understanding [23, 43, 58, 82]. This operationalization arises as classical views of tacit knowledge akin to implicit and procedural memory [63] have since been accounted for within other models [41] and provides an explanandum for a range of observed phenomena otherwise attributable to an alignment of theories of EM and experiential processing [74, cf. 58]. These include not just retrospection and prospection, as discussed here, but single-trial and social learning [e.g., 24, 43, 80]. We adopt the notion that memory formation associated with the EM system may be accounted for through the recall of spatiotemporal relations, associated with the given procedure. Furthermore, we anticipate these relations may be expressed through the visualizations and narratives produced throughout the learning and assessment periods, and manifest as tacit knowledge, which may be differentiated from memory formation and knowledge associated with other cognitive structures.

Assessing EM & Tacit Knowledge. Building on our prior work [74], we have started to develop a framework for the measurement of EM and tacit knowledge, and tested its application in a recent study [76]. The episodic recall tests included (1) a delayed free recall task; (2) a what-where-when (WWW) task, and (3) a spatiotemporal mapping task. These measures were selected based on a prior scoping review and framework for measuring EM and tacit knowledge [74]. The episodic recall test was untimed, but participants were informed of an expected duration of 15 min. This measurement framework seeks to triangulate tacit knowledge through the use of three complementary assessments, including (1) a delayed free-recall task, accounting for top-down EM processes, measuring a composite of features of episodic representation (e.g., specificity, vividness, coherence); (2) a what-where-when task adapted from Laurent et al. [42], accounting for feature binding in bottom-up EM processing, which integrates measures for item/object memory (what) with the spatial (where) and temporal (when) context of the training content [42] paired with items derived from the remember/know/guess paradigm [e.g., 26, 62]; and (3) a spatiotemporal mapping task adapted from a point/route configuration knowledge test protocol [8], intended to account for learners' cognitive map as an allocentric representation of the virtual environment [8]. Results provided preliminary validation that the framework was successful in differentiating learning between the episodic and declarative/semantic systems, albeit we continue its refinement. These results also lent credibility to the concept of a distinct EM system, which enables instructional features (e.g., narrative, visual cues) and system factors (e.g., immersion, interactivity) to affect learners' memory of spatiotemporal relations differently than that of the declarative steps of a procedure.

Episodic Facilitation. Episodic facilitation as an instructional approach may be described as an application of instructional elements considered to be aligned with the properties of EM and associated learning and design principles as presented in our framework for EM in experiential learning [74]. The framework presents a series of learning and design principles for experiential learning in VR simulations grounded in the properties of the EM system [14]. The learning and design principles provide the foundation for interventions across the training cycle. Before training, for example, episodic facilitation may be implemented through advance organizers and narratives to support EM properties associated with framing, visual representation, and perspective. During training, EM may be supported through enhancing immersion to facilitate experiential processing and providing cues to increase the salience of meaningful information leveraging differential activation and inhabitation in the EM system. After training, episodic facilitation may occur through narrative reflection and mental simulation to facilitate recollective experience and the iterative grounding/framing of self-concepts to support training objectives [14, 22, 74]. That is, we suggest that the use of episodic facilitation may improve procedures training using VR simulations, due to the functions and processes of the EM system. The acquisition of tacit knowledge of executing a procedure may be facilitated through immersion, as experiences are grounded through the direct perception of ecological information as affordances (i.e., via bottom-up EM processes). The acquisition of tacit knowledge of executing a procedure may also be facilitated through narrative, as experiences are framed to influence the organization and salience of experience (i.e., via top-down EM processes).

2.4 A Framework for Episodic Memory in Procedures Training Using VR

This literature review and synthesis furthers the theoretical foundations for a program of research on the function and value of the EM system, and interventions targeting these functions, in memory formation during training of procedures using VR. In prior work [74], we outlined a framework for simulation training that aligned mechanisms of EM with experiential learning and 4E cognition. In this review and synthesis, we further aligned that EM framework with relevant models of learning across scales [53], to refine a theoretical basis for investigations and interventions involving the use of VR simulations for procedures training. Having identified similitudes and limitations among the reviewed theories here, we suggest that they may be aligned with and mutually informative to this EM framework [74]:

- **HIP & EM.** For bottom-up EM processing, the HIP specifies what stimuli may be perceived as episodic elements (selective attention) and maintained for encoding (focused attention). For top-down processing, the HIP specifies what episodic structures are prioritized for retrieval. Complementarily, EM theory offers frames as an explanatory mechanism for directing focused and divided attention.
- **M-WM & EM.** For bottom-up processing, M-WM details the mechanisms of how modality-specific episodic elements are integrated into higher-order structures and clarifies the interdependence of EM and the procedural system for enacting goals via the episodic buffer. Considerations for 4E cognition in EM clarify the function and value of modality-specific information for tacit knowledge and clarify how encoding and consolidation are driven by (goal-defining) frames.
- **CTML & EM.** CTML highlights how the stimuli and affordances perceived by a learner in VR simulation are largely at the discretion of simulation designers, and that EM—as an LTM subsystem—may be similarly impacted by the effects of intrinsic, extraneous, and generative load/processing. Consideration of the EM concept contributes that the top-down processing EM system may affect which stimuli contribute to these types of cognitive load/processing, impacting outcomes. Considerations for 4E cognition in EM contribute that some stimuli and affordances may be generative or even not contribute load, being embedded within the VR simulation, in that they are directly perceived and acted upon due to perception-action coupling or are otherwise offloaded to bodily & environmental systems.
- **CAMIL & EM.** CAMIL specifies system factors (e.g., immersion, interactivity) that may affect what episodic elements are perceived and encoded, and experiential factors (e.g., presence, agency) that may influence how episodic structures are grounded and framed by the EM system. The model also specifies other constructs and processes (e.g., self-efficacy, self-regulation) that may affect EM formation. Our EM framework implies that learning outcomes in VR cannot be expressed through a singular metric or type of knowledge; and that these may differentially affect memory formation across different systems.
- **CMC & EM.** As the cognitive architecture grounding our synthesis, the CMC provides a detailed specification of processes and interdependencies involved in representation, WM, procedural memory, and declarative memory which we view as interfacing with the EM system across bottom-up and top-down processes. The function and specification of non-symbolic metadata (e.g., frequency, recency, co-occurrence, similarity, utility) and constituent work on EM as a WM archive may be useful concepts for further refinement of EM theory. In turn, this EM framework could address gaps in the CMC in accounting for retrospection, prospection, embodiment, and direct perception in offloading representation to the (virtual) environment.

A pluralistic view of the EM system, accepting both representational and non-representational mechanisms [20], grounds the functions of perception-action systems in attunement and calibration to invariant stimuli as affordances for embodied actions within the embedded virtual environment [28–30]. Episodic elements, derived from experiential information, are selectively attended to through the frames of EM which prioritize goal-related affordances [14, 87]. Through mechanisms of focused attention, the WM system facilitates the selection and integration of episodic elements into successively higher-order episodic structures, bounded by goals and encoded within LTM through the episodic buffer [3, 6, 14, 22, 87]. With the allocation of attention to processing experiential information from the VR simulation, corresponding changes may emerge within higher-order cognitive processes associated with virtual experience (e.g., engagement, presence) [72, 74, 86]. Facilitated in part through non-symbolic metadata, detailed by the CMC [40, 41], these mechanisms provide a basis for time and context—spatiotemporal relations—associated with these episodes [3, 14, 42, 84]. Propositional knowledge (e.g., steps of a procedure) and condition-action structures (e.g., cognitive scripts) are assimilated into their respective systems [22. 40, 41]. Procedural memory regulates internal and external actions in support of goals [22, 55, 87, 89], driven by higher-order EM structures and self-concepts (e.g., frames, values, identity) [14]. Given deliberate practice, procedural skills are acquired and executed with reduced cognitive demand through compilation [9, 41, 47], supported by mechanisms of 4E cognition such as embodiment and offloading [49, 60]. Over time, adaptive expertise may be acquired as the learner further attunes to patterns of spatiotemporal relations among the embedded environmental cues, expressed through situated actions and as imperfectly articulable tacit knowledge [74].

These and other processes outlined within this review provide the foundation for an explanatory account of the mechanisms of learning procedures using VR simulations, which address concepts (e.g., presence, embodiment) largely unaccounted for by the functions of the declarative/semantic and procedural memory systems. We synthesize a nascent framework for simulation training aligning EM with experiential learning and 4E cognition [74] with complementary models of learning across scales [53]. Delineating the nature of procedures within their operational context, we discussed how procedural skills may be trained, and briefly reviewed cognitive underpinnings of the learning of procedures in VR simulations, with a focus on conceptual intersections underlying these theories with the unique characteristics and functions of the EM system. Our synthesis illustrates how the concept of EM from the cognitive sciences [3, 14, 21, 22, 84] may inform investigations, assessments, and interventions concerning procedures training using VR simulations, and provides the foundation for a theoretical framework from which research questions may be derived to guide our current program of research.

3 Conclusions

In support of a nascent program of research, we identified similitudes among frameworks and concepts concerning the cognitive processes involved in the learning of procedures in VR simulations. Specifically, our review focused on the role and function of the EM system as an explanatory mechanism bridging gaps among theories and concepts across scales of cognition. Through our synthesis, we provide the basis to justify a position that interventions before, during, and after VR experiences targeting the mechanisms of EM may have downstream benefits on spatial cognition, tacit knowledge, and decision-making during flight deck operations. Furthermore, whereas extant training frameworks do not delineate EM or prescribe its measurement, we provide a framework for assessing the unique contribution of the EM system to learners' memory of training events. Such consideration of the role, function, and value of the EM system provides a more precise understanding of how constructs such as immersion may facilitate or hinder the training of procedures using VR simulations.

Acknowledgments. The views expressed herein are those of the authors and do not reflect the views of the institutions to which they are employed, any sponsors, or the University of Central Florida.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Albus, P., Seufert, T.: The modality effect reverses in a virtual reality learning environment and influences cognitive load. Instr. Sci. **51**, 545–570 (2023)
2. Anderson, J.R.: Spanning seven orders of magnitude: a challenge for cognitive modeling. Cogn. Sci. **26**(1), 85–112 (2002)
3. Andonovski, N.: Episodic representation: a mental models account. Front. Psychol. **13**, 899371 (2022)
4. Angel-Urdinola, D.F., Castillo-Castro, C., Hoyos, A.: Meta-analysis assessing the effects of virtual reality training on student learning and skills development (Policy Research Working Paper Series 9587). The World Bank (2021)
5. Auer, S., Gerken, J., Reiterer, H., Jetter, H.C.: Comparison between virtual reality and physical flight simulators for cockpit familiarization. In: Schneegass, S., Pfleging, B., Kern, D. (Eds.), Proceedings of Mensch und Computer 2021, pp. 378–392. Association for Computing Machinery, New York (2021)
6. Baddeley, A.D., Hitch, G.J., Allen, R.J.: A multicomponent model of working memory. In: Logie, R., Camos, V., Cowan, N. (eds.) Working Memory: State of the Science, pp. 10–43. Oxford University Press, United Kingdom (2021)
7. Bell, B.S., Kanar, A.M., Kozlowski, S.W.: Current issues and future directions in simulation-based training in North America. Int. J. Hum. Resour. Manage. **19**(8), 1416–1434 (2008)
8. Bendell, R., Williams, J.: Assessing spatial knowledge and mental map development under virtual training conditions. Proc. Hum. Fact. Ergon. Soc. Ann. Meet. **67**(1), 1611–1616 (2023)
9. van den Bosch, K.: Durable competence in procedural tasks through appropriate instruction and training. In: Harris, D. (Ed.), Engineering Psychology and Cognitive Ergonomics: Transportation Systems, Medical Ergonomics and Training, (1st ed., pp. 431–438). Aldershot, Ashgate (1999)
10. Burch, G.F., Giambatista, R., Batchelor, J.H., Burch, J.J., Hoover, J.D., Heller, N.A.: A meta-analysis of the relationship between experiential learning and learning outcomes. Decis. Sci. J. Innov. Educ. **17**(3), 239–273 (2019)
11. Choi, H.H., Van Merriënboer, J.J., Paas, F.: Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. Educ. Psychol. Rev. **26**, 225–244 (2014)
12. Clark, R.E.: Media will never influence learning. Education Tech. Research Dev. **42**(2), 21–29 (1994)
13. Coban, M., Bolat, Y.I., Goksu, I.: The potential of immersive virtual reality to enhance learning: a meta-analysis. Educ. Res. Rev. **36**, 100452 (2022)

14. Conway, M.A.: Episodic memories. Neuropsychologia **47**(11), 2305–2313 (2009)
15. Cross, J.I., Boag-Hodgson, C., Ryley, T., Mavin, T., Potter, L.E.: Using extended reality in flight simulators: a literature review. IEEE Trans. Visual Comput. Graphics **29**(9), 3961–3975 (2022)
16. Dahlstrom, N., Dekker, S., van Winsen, R., Nyce, J.: Fidelity and validity of simulator training. Theor. Issues Ergon. Sci. **10**(4), 305–314 (2009)
17. Dattel, A., Karunratanakul, K., Crockett, S., Fabbri, J.: How procedural and conceptual training affect flight performance for learning traffic patterns. Proc. Hum. Fact. Ergon. Soc. Ann. Meet. **59**(1), 855–858 (2015)
18. Drury, C.G., Johnson, W.B.: Writing aviation maintenance procedures that people can/will follow. Proc. Hum. Fact. Ergonom. Soc. Ann. Meet. **57**(1), 997–1001 (2013)
19. Farr, M.J.: The long-term retention of knowledge and skills: A cognitive and instructional perspective (IDA Memorandum Report M-205). Office for the Under Secretary of Defense for Research and Engineering, Institute for Defense Analyses: Alexandria, VA, September 1986
20. Favela, L.H., Martin, J.: “Cognition” and dynamical cognitive science. Mind. Mach. **27**, 331–355 (2017)
21. Ferbinteanu, J.: Memory systems 2018–towards a new paradigm. Neurobiol. Learn. Mem. **157**, 61–78 (2019)
22. Fiore, S.M.: Making time for memory and remembering time in motivation theory. In: Kanfer, R., Chen, G., Pritchard, R.D. (Eds.), Work Motivation: Past, Present and Future, pp. 541–553 (2008)
23. Forbus, K.D., Hinrichs, E.T., Crouse, E.M., Blass, J.: Analogies versus rules in cognitive architecture. Adv. Cogn. Syst. **9**, 13 (2020)
24. Gallagher, S.: The 4Es and the 4As (affect, agency, affordance, autonomy) in the meshed architecture of social cognition. In: Robinson, M.D., Thomas, L.E. (eds.) Handbook of Embodied Psychology: Thinking, Feeling, and Acting, pp. 357–370. Springer, Cham (2021)
25. Gallagher, S., Lindgren, R.: Enactive metaphors: Learning through full-body engagement. Educ. Psychol. Rev. **27**(3), 391–404 (2015)
26. Gardiner, J.M., Ramponi, C., Richardson-Klavehn, A.: Recognition memory and decision processes: a meta-analysis of remember, know, and guess responses. Memory **10**(2), 83–98 (2002)
27. Gentry, J.W.: What is experiential learning? In: Guide to Business Gaming and Experiential Learning, pp. 9–20. Nichols Pub, New York (1990)
28. Gibson, J.J.: The theory of affordances. In: Shaw, R.E., Bransford, J. (eds.) Perceiving, Acting, and Knowing: Toward an Ecological Psychology, pp. 67–82. Lawrence Erlbaum Associates, Hillsdale, N.J. (1977)
29. Golonka, S., Wilson, A.D.: Ecological representations. Ecol. Psychol. **31**(3), 235–253 (2019)
30. Hacques, G., Komar, J., Dicks, M., Seifert, L.: Exploring to learn and learning to explore. Psychol. Res. **85**(4), 1367–1379 (2021)
31. Hoffman, R.R., Ward, P., Feltovich, P.J., DiBello, L., Fiore, S.M., Andrews, D.H.: Accelerated Expertise: Training for High Proficiency in a Complex World. Taylor & Francis (2014)
32. Howard, M.C., Gutworth, M.B., Jacobs, R.R.: A meta-analysis of virtual reality training programs. Comput. Hum. Behav. **121**, 106808 (2021)
33. Jaakkola, E.: Designing conceptual articles: four approaches. AMS Rev. **10**(1–2), 18–26 (2020)
34. Jackson, P.C., Jr.: Thoughts on bands of action. Proc. Comput. Sci. **145**, 710–716 (2018)
35. Jongbloed, J., Chaker, R., Lavoué, E.: Immersive procedural training in virtual reality: a systematic literature review. Comput. Educ. 105124 (2024)

36. Kaplan, A.D., Cruit, J., Endsley, M., Beers, S.M., Sawyer, B.D., Hancock, P.A.: The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. Hum. Factors **63**(4), 706–726 (2021)
37. Klein, G., Phillips, J.K., Rall, E.L., Peluso, D.A.: A data–frame theory of sensemaking. In: Hoffman, R.R. (ed.) Expertise Out of Context, pp. 113–155. Psychology Press (2007)
38. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development (2nd ed.) Pearson Education, Inc., NJ (Originally published 1984) (2015)
39. Konoske, P.J., Ellis, J.A.: Cognitive factors in learning and retention of procedural tasks (NPRDC TR 87-14). Navy Personnel Research and Development Center, San Diego, California, December 1986
40. Laird, J.E.: An analysis and comparison of ACT-R and Soar. In: Proceedings of the Ninth Annual Conference on Advances in Cognitive Systems (2021)
41. Laird, J.E., Lebiere, C., Rosenbloom, P.S.: A standard model of the mind: toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. AI Mag. **38**(4), 13–26 (2017)
42. Laurent, X., Ensslin, A., Marí-Beffa, P.: An action to an object does not improve its episodic encoding but removes distraction. J. Exp. Psychol. Hum. Percept. Perform. **42**(4), 494–507 (2016)
43. Linde, C.: Narrative and social tacit knowledge. J. Knowl. Manag. **5**(2), 160–171 (2001)
44. Longo, U.G., et al.: Augmented reality, virtual reality and artificial intelligence in orthopedic surgery: a systematic review. Appl. Sci. **11** (2021)
45. Makransky, G., Petersen, G.B.: The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. Educ. Psychol. Rev. **33**(3), 937–958 (2021)
46. Mann, S., Furness, T., Yuan, Y., Iorio, J., Wang, Z.: All reality: virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. In: Proceedings of the 12th Annual ACM International Conference on Multimedia, pp. 620–627 (2018)
47. Mayer, R.E.: Multimedia Learning, (3rd ed.). Cambridge University Press (2021)
48. McGowin, G., Fiore, S.M.: Mind the gap! Advancing immersion in virtual reality—factors, measurement, and research opportunities. Proc. Hum. Fact. Ergon. Soc. Ann. Meet. **68**(1), 1648–1654 (2024)
49. McGowin, M., Fiore, S.M., Oden, K.: Towards a theory of learning in immersive virtual reality: Designing learning affordances with embodied, enactive, embedded, and extended cognition. In: Cherner, T., Fegely, A. (eds.) Bridging the XR Technology-to-Practice Gap: Methods and Strategies for Blending Extended Realities into Classroom Instruction, pp. 35–53. Association for the Advancement of Computing in Education (2023)
50. McGowin, G., Sonnenfeld, N.A., Fiore, S.M.: Navigating cognitive demand in virtual reality: implications for education and training. Proc. Hum. Fact. Ergon. Soc. Ann. Meet. **68**(1), 1668–1673 (2024)
51. Milgram, P., Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inform. Syst. **E-77D**(12), 1321–1329 (1994)
52. Munro, A., Surmon, D., Pizzini, Q.: Teaching procedural knowledge in distance learning environments. In: O'Niel, H.F., Perez, R.S. (eds.) Web-Based Learning: Theory, Research, and Practice, pp. 255–278. Lawrence Erlbaum Associates, Inc., New Jersey, USA (2006)
53. Newell, A.: Unified Theories of Cognition. Harvard University Press, Cambridge (1990)
54. Newen, A., Gallagher, S., De Bruin, L.: 4E cognition: Historical roots, key concepts, and central issues. In: Newen, A., De Bruin, L., Gallagher, S. (eds.) The Oxford Handbook of 4E Cognition, pp. 3–16. Oxford University Press (2018)
55. Oberauer, K.: Design for a working memory. Psychol. Learn. Motiv. **51**, 45–100 (2009)
56. Ockerman, J., Pritchett, A.: A review and reappraisal of task guidance: aiding workers in procedure following. Int. J. Cogn. Ergon. **4**(3), 191–212 (2000)

57. Ong, C.W., Tan, M.C.J., Lam, M., Koh, V.T.C.: Applications of extended reality in ophthalmology: systematic review. J. Med. Internet Res. **23**(8), e24152 (2021)
58. Owens, K.: Employing artificial intelligence to increase occupational tacit-knowledge through competency-based experiential learning. In: Goldberg, B., Robson, R., Proceedings of the 1st Workshop on Artificial Intelligence in Support of Guided Experiential Learning, pp. 58–67 (2023)
59. Paas, F., Sweller, J.: Implications of cognitive load theory for multimedia learning. In: Mayer, R.E., (Ed.) The Cambridge Handbook of Multimedia Learning (2nd ed., pp. 27–42) (2014)
60. Pande, P.: Learning and expertise with scientific external representations: an embodied and extended cognition model. Phenomenol. Cogn. Sci. **20**(3), 463–482 (2021)
61. Parong, J., Mayer, R.E.: Cognitive and affective processes for learning science in immersive virtual reality. J. Comput. Assist. Learn. **37**(1), 226–241 (2021)
62. Pause, B.M., Zlomuzica, A., Kinugawa, K., Mariani, J., Pietrowsky, R., Dere, E.: Perspectives on episodic-like and episodic memory. Front. Behav. Neurosci. **7**(33) (2013)
63. Polanyi, M.: 2009. The Tacit Dimension. The University of Chicago Oress, Original work published (1966)
64. Rauschnabel, P.A., Felix, R., Hinsch, C., Shahab, H., Alt, F.: What is XR? towards a framework for augmented and virtual reality. Comput. Hum. Behav. **133**, 107289 (2022)
65. Richards, B.A., Frankland, P.W.: The persistence and transience of memory. Neuron **94**(6), 1071–1084 (2017)
66. Rodríguez, J., Gutiérrez, T., Sánchez, E.J., Casado, S., Aguinaga, I.: Training of procedural tasks through the use of virtual reality and direct aids. In: Lányi, C.S. (Ed.), Virtual Reality and Environments, pp. 43–68 (2012)
67. Sauzéon, H., et al.: The use of virtual reality for episodic memory assessment: the effect of active navigation. Exp. Psychol. **59**(2), 99–108 (2012)
68. Schaffernak, H., et al.: Novel mixed reality use cases for pilot training. Educ. Sci. **12**(5), 345 (2022)
69. Skarbez, R., Smith, M., Whitton, M.C.: Revisiting Milgram and Kishino's reality-virtuality continuum. Front. Virtual Real. **2**, 647997 (2021)
70. Slater, M.: A note on presence terminology. Pres. Conn. **3**(3), 1–5 (2003)
71. Slater, M., Wilbur, S.: A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Pres. Tele. Virtual Environ. **6**(6), 603–616 (1997)
72. Smith, S.A.: Virtual reality in episodic memory research: a review. Psychon. Bull. Rev. **26**, 1213–1237 (2019)
73. Sonnenfeld, N.A., Nguyen, B., Boesser, C.T., Duruaku, F., Jentsch, F.: Considerations for Electronic & Distance Learning of Procedures in Flight Crew Training [Technical report]. Team Performance Laboratory, Institute for Simulation & Training, University of Central Florida, 21 October 2021
74. Sonnenfeld, N.A., Nguyen, B., Gomez, C., Jentsch, F.G., Fiore, S.M.: On episodic memory in experiential learning for flightcrew training. In: Interservice/Industry Training, Simulation, and Education Conference, p. 23396 (2023a)
75. Sonnenfeld, N.A., Nguyen, B., Alonso, A., Jentsch, F.: Training effectiveness evaluation: frameworks and considerations for flightcrew training review & approval. Proc. Hum. Fact. Ergon. Soc. Ann. Meet. **67**(1), 2002–2004 (2023)
76. Sonnenfeld, N.A., et al.: Crafting recall: Impacts of narrative on semantic vs. episodic memory & perceptions for an aviation procedure. In: Proceedings of the International Conference on Applied Human Factors and Ergonomics (in press)
77. Stadler, M.: On learning complex procedural knowledge. J. Exp. Psychol. Learn. Mem. Cogn. **15**(6), 1061–1069 (1989)

78. Stanney, K.M., Nye, H., Haddad, S., Hale, K.S., Padron, C.K., Cohn, J.V.: Extended reality (XR) environments. In: Salvendy, G., Karwowsi, W. (Eds.), Handbook of Human Factors and Ergonomics, 5th edn., pp. 782–815. Wiley (2021)
79. Stanney, K.M., Skinner, A., Hughes, C.: Exercisable learning-theory and evidence-based andragogy for training effectiveness using XR (ELEVATE-XR): Elevating the ROI of immersive technologies. Int. J. Hum.-Comput. Interact. **39**(11), 2177–2198 (2023)
80. Sutton, J., Harris, C.B., Keil, P.G., Barnier, A.J.: The psychology of memory, extended cognition, and socially distributed remembering. Phenomenol. Cogn. Sci. **9**(4), 521–560 (2010)
81. Sweller, J.: Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. Instr. Sci. **32**(1), 9–31 (2004)
82. Taylor, H.: Tacit knowledge: conceptualizations and operationalizations. Int. J. Knowl. Manag. **3**(3), 60–73 (2007)
83. Tulving, E.: Précis of elements of episodic memory. Behav. Brain Sci. **7**(2), 223–268 (1984)
84. Tulving, E.: How many memory systems are there? Am. Psychol. **40**(4), 385–398 (1985)
85. Tulving, E.: What is episodic memory? Curr. Dir. Psychol. Sci. **2**(3), 67–70 (1993)
86. Waterworth, E.L., Waterworth, J.A.: Focus, locus, and sensus: the three dimensions of virtual experience. Cyberpsychol. Behav. **4**(2), 203–213 (2001)
87. Wickens, C.D., Carswell, C.M.: Information processing. In: Salvendy, G., Karwowski, W. (Eds.), Handbook of Human Factors and Ergonomics (5th ed., pp. 114–158) (2021)
88. Wilson, M.: Six views of embodied cognition. Psychon. Bull. Rev. **9**, 625–636 (2002)
89. Winne, P.H.: Self-regulated learning viewed from models of information processing. In: Zimmerman, B.J., Schunk, D.H. (eds.) Self-Regulated Learning and Academic Achievement, 2nd edn., pp. 145–178. Routledge, New York (2001)
90. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence **7**(3), 225–240 (1998)
91. Wood, R.E.: Task complexity: definition of the construct. Organ. Behav. Hum. Decis. Process.Decis. Process. **37**(1), 60–82 (1986)

Sales Skills Training in Virtual Reality: An Evaluation Utilizing CAVE and Virtual Avatars

Francesco Vona(✉), Michael Stern, Navid Ashrafi, Julia Schorlemmer, Jessica Stemann, and Jan-Niklas Voigt-Antons

University of Applied Sciences Hamm-Lippstadt, Hamm, Germany
{francesco.vona,michael.stern,navid.ashrafi,julia.schorlemmer,
jessica.stemann,jan-niklas.voigt-antons}@hshl.de

Abstract. This study investigates the potential of virtual reality (VR) for enhancing sales skills training using a Cave Automatic Virtual Environment (CAVE). VR technology enables users to practice interpersonal and negotiation skills in controlled, immersive environments that mimic real-world scenarios. In this study, participants engaged in sales simulations set in a virtual dealership, interacting with avatars in different work settings and with various communication styles. The research employed a within-subjects experimental design involving 20 university students. Each participant experienced four distinct sales scenarios randomized for environmental and customer conditions. Training effectiveness was assessed using validated metrics alongside custom experience questions. Findings revealed consistent user experience and presence across all scenarios, with no significant differences detected based on communication styles or environmental conditions. The study highlights the advantages of semi-immersive VR systems for collaborative learning, peer feedback, and realistic training environments. However, further research is recommended to refine VR designs, improve engagement, and maximize skills transfer to real-world applications.

Keywords: Sales Skills Training · Virtual Reality · CAVE

1 Introduction

Sales and negotiation skills are fundamental to business success [10], and their importance is particularly important in the context of a highly competitive global market and always rising customer expectations [19]. Traditionally, sales training relies heavily on the development of interpersonal skills through role-playing exercises. In these exercises, the trainers act as customers, allowing the trainees to practice in a low-stakes environment. This approach, known as face-to-face training, requires significant resources, as it depends on the expertise of training specialists. Consequently, face-to-face training is often the most expensive form of traditional training [19]. Despite its high cost, face-to-face training remains

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 252–267, 2025.
https://doi.org/10.1007/978-3-031-93715-6_16

widely used, often supplemented with paper-based or video-based materials [24]. However, traditional methods like role-plays and business simulations depend heavily on participant performance and adaptability, which can limit their overall effectiveness. These limitations highlight the need for innovative training approaches to provide experiential, interactive, and scalable learning opportunities in sales and negotiation. Immersive media, including virtual reality (VR) and, more generally, extended reality (XR), has emerged as a promising alternative to traditional methods. Technologies like head-mounted displays and cave automatic virtual environments (CAVE) create realistic, engaging, and interactive training scenarios, addressing many limitations of conventional techniques [19]. Unlike static training materials, VR can replace traditional paper-based questionnaires with more immersive VR assessments, offering a more engaging way to measure trainee performance and learning outcomes [14]. Additionally, VR and virtual environments can induce emotional states more effectively than other media types, which is particularly beneficial for realistic and impactful training scenarios [22].

In particular, VR offers unique advantages for sales training. It allows trainees to practice in lifelike simulations that closely mimic real-world scenarios, leading to improved skill retention and application in actual job settings [19]. Unlike traditional role-play exercises, VR enables risk-free learning environments where trainees can make mistakes and refine their skills without real-world consequences [13]. Moreover, VR-based training can standardize learning experiences, ensuring consistent quality across participants, while providing immediate feedback.

CAVEs, as an advanced XR tool, provide additional benefits for immersive training. They create collaborative, large-scale virtual environments where trainees can practice negotiation and interpersonal skills in dynamic, team-based settings. These technologies expand the scope of experiential learning, offering structured, interactive simulations that facilitate the development of critical sales and negotiation competencies [4]. Despite the growing interest in immersive media for education, there is a lack of empirical studies focusing on the development, testing, and evaluation of immersive training tools tailored specifically for sales and negotiation contexts. Current research has largely overlooked the potential of technologies like CAVEs in addressing the shortcomings of traditional training approaches.

This study seeks to fill this gap by exploring the application of CAVEs in simulations for sales and negotiation training. By focusing on the unique capabilities of CAVE environments, this research aims to create a scalable, engaging, and effective platform for immersive learning. The primary goal of this study is to harness the potential of immersive media, particularly CAVEs, to enhance sales training. Specifically, the objectives include: i) Designing a user-centered sales and negotiation simulation tailored for CAVE environments, ii) implementing this simulation to create an interactive and engaging learning experience, and iii) conducting an initial usability study to evaluate the effectiveness of the CAVE-based training approach.

2 Related Work

Virtual reality training programs have demonstrated success across diverse domains, including gamified shooting simulations [15], sommelier training [13], and technical skill development [2]. Increasingly, VR training is being adopted in educational contexts, offering immersive and engaging experiences that improve outcomes compared to traditional methods [6,8,21]. The immersive nature of VR enhances realism and engagement, providing trainees with dynamic environments to practice and refine skills. These features are particularly valuable in sales training, where interpersonal skills are critical [19,21]. VR enables realistic simulations of job scenarios, allowing trainees to practice in a risk-free environment. This immersive experience has been shown to improve performance in actual workplace settings, producing outcomes superior to those achieved through traditional alternatives [8]. For instance, research [21] highlights the application of VR in sales training, using case studies and cutting-edge research to explore its implications for practice. However, despite these advancements, there remains a significant gap in the development of VR systems dedicated specifically to sales training, emphasizing the need for further research in this area [19].

CAVE systems have also been successfully applied in various fields, including education, safety training, and engineering [11,12,18,20]. Their interactive and immersive nature has been shown to enhance learning experiences, making them versatile tools in modern training environments [5,9,25]. However, the effectiveness of CAVE varies depending on the application and individual learner abilities, underscoring the importance of ongoing research and development in this domain. In education, CAVE systems have been used to teach fire safety skills to children through game-like interactions, increasing engagement and motivation. These systems make standard safety information more engaging and enjoyable, leading to improved learning outcomes [18]. Similarly, serious games for school fire prevention have leveraged CAVE to provide realistic, interactive simulations, enhancing learning through hands-on discovery [12]. CAVE has also been explored in the context of emotional intelligence training. While it can simulate emotional scenarios effectively, studies indicate that training success often depends on trainees' spatial abilities rather than emotional intelligence itself. This suggests that, while VR provides valuable simulations, it may not fully replicate complex interpersonal communication [11]. In sports, CAVE systems have been used to train athletes for high-pressure scenarios by inducing controlled anxiety. While promising, further research is needed to evaluate the long-term benefits of these applications [20]. Last, CAVE was also explored in engineering education, where it was compared with other VR setups, demonstrating superior outcomes in student achievement. The immersive and interactive nature of CAVE provides a more engaging learning experience, resulting in better educational outcomes than traditional methods [1].

This work builds upon the authors' earlier research on user-centered simulations for leadership development in CAVE environments [23]. In their previous study, the authors designed and evaluated leadership training scenarios tailored to simulate realistic workplace situations, such as providing critical feedback or addressing health concerns. These scenarios enabled participants to interact with virtual characters in dynamic, context-rich environments, delivering high levels of user presence and interactivity. The findings underscored the potential of CAVE as an effective tool for experiential learning, particularly in domains requiring interpersonal skill development, such as leadership and sales training.

3 Methods

3.1 Study Design

The study employed a within-subjects experimental design to ensure that all participants experienced each of the testing conditions. The goal was to observe and assess how participants adapted their communication strategies and perceived their experiences across different contexts. The order of the conditions was randomized, minimizing potential biases related to individual differences. The experimental conditions were designed to simulate real-world sales scenarios in a controlled virtual environment. Participants engaged in role-playing exercises, interacting with virtual customers who displayed varying personalities and communication styles in different working environments. The four distinct scenarios combined two factors: Customer (Avatar) Personality (Friendly vs. Unfriendly) and Environmental Atmosphere (Friendly vs. Unfriendly). Conditions were defined as follows: Condition 1 = Friendly User x Friendly Environment (FUxFE), Condition 2 = Friendly User x Unfriendly Environment (FUxUE), Condition 3 = Unfriendly User x Friendly Environment (UUxFE), and Condition 4 = Unfriendly User x Unfriendly Environment (UUxUE). This factorial design allowed researchers to systematically explore how these variables influenced participants' user experiences and interactions. In each scenario, the participants had to follow a pre-configured script. Figure 1 provides a visual representation of the conditions tested during the study.

3.2 Participants

Participants were recruited from the student population of our university. The final sample comprised 20 individuals, with an average age of 24.65 years (SD = 4.20). The sample included 75% male participants (n = 15) and 25% female participants (n = 5). Participants' affinity for technology interaction (ATI) was assessed using the ATI scale [7], with an overall mean score of 3.65 ($SD = 1.09$), indicating moderate comfort and familiarity with technology within the sample. Recruitment efforts ensured diversity in educational backgrounds, though all participants had some level of familiarity with sales concepts, either through coursework or extracurricular activities. Inclusion criteria required participants to have no prior experience with the specific VR system used in the study. To encourage participation and ensure adequate representation, participants who were not employed by the host institution received a monetary compensation of 15 Euros per hour for their time and effort. The study was conducted in

Fig. 1. The four conditions that were tested during the experiment: Condition 1 = Friendly User x Friendly Environment (bottom left), Condition 2 = Friendly User x Unfriendly Environment (top left), Condition 3 = Unfriendly User x Friendly Environment (bottom right), and Condition 4 = Unfriendly User x Unfriendly Environment (top right).

compliance with ethical guidelines and received approval from the university's local ethics commission.

3.3 Measures

A combination of quantitative and qualitative measures was used to evaluate participants' experiences:

– Presence: The Igroup Presence Questionnaire (IPQ) was administered to measure participants' sense of being "present" in the virtual environment. This scale assessed factors such as spatial presence, involvement, and realism [17].
– Affinity for Technology Interaction: The ATI (Affinity for Technology Interaction) scale was used to profile participants' general attitude and comfort with technology. This measure provided insights into individual differences that could influence user experience in a VR setting [7].
– User Experience: The User Experience Questionnaire - Short Version (UEQ-S) was employed to evaluate participants' overall satisfaction with the VR environment. The scale captured dimensions such as pragmatic and hedonic quality, and usability [16].
– Social Presence: The Social Presence Questionnaire (SPQ) measured participants' perceived social presence, focusing on the sense of being with and interacting with others in virtual environments. The SPQ evaluates dimensions such as mutual awareness, co-presence, and interaction quality [3].
– Custom Questions: A set of custom questions was included after each scenario to gather additional feedback on the specific interaction and perceived

challenges. These questions were designed to identify contextual nuances not captured by standardized measures. The custom questions were assessed on a 7-point Likert scale. The items are as follows: CUSQ1: "How realistic did the interview situation feel?" (1 = Not realistic at all, 7 = Very realistic); CUSQ2: "How well did you feel during the simulated interview?" (1 = Bad, 7 = Very good); CUSQ3: "Did you experience any discomfort or pain during the interview?" (1 = No discomfort at all, 7 = A lot of discomfort); CUSQ4: "Did you achieve your goal for this interview?" (1 = Not achieved at all, 7 = Fully achieved); CUSQ5: "Did you experience any challenges during the interview?" (1 = No challenges at all, 7 = A lot of challenges).

At the end of the session, participants were allowed to leave further feedback about their overall experience.

3.4 Procedure

Upon arrival, participants were welcomed and briefed on the study's purpose and procedure. After signing an informed consent form, they were asked to complete a demographics questionnaire to collect information about their age, gender, educational background, and prior experience with VR technology or sales scenarios. The CAVE system was initialized, and the physical room was arranged to resemble a dealership office. This setup included realistic props like a desk and chairs to create a contextually relevant and engaging environment. Participants were then introduced to the sales task and given a general briefing about the sales scenarios. Each participant completed four scenarios, with each scenario lasting approximately five minutes. In these scenarios, participants assumed the role of a salesperson tasked with selling either a used car or motorcycle to a virtual customer. The customer's personality (friendly vs. unfriendly) and environmental atmosphere (friendly vs. unfriendly) varied across scenarios, ensuring exposure to all conditions. To maintain consistency, participants were provided with conversation guidelines to structure their interactions (Fig. 2). After completing each scenario, participants filled out a post-run questionnaire to evaluate their experience and assess their interaction with the virtual customer. Upon completing all scenarios, participants were asked to fill out a final questionnaire summarizing their overall experience. This questionnaire included both quantitative measures (e.g., standardized scales) and open-ended questions to capture qualitative insights. The entire study session, including briefing, scenarios, and debriefing, lasted approximately 60 min per participant.

4 Results

A series of analyses were conducted to evaluate the participants' experiences and interactions within the simulated sales scenarios. Initially, descriptive statistics were computed to summarize the data and provide an overview of the participants' responses across the different conditions. Subsequently, repeated-measures

Fig. 2. A participant was photographed during the study, sitting on a chair in the CAVE system. Between him and the avatar, a table was placed to enhance the realism of the setting.

analyses of variance (RM-ANOVAs) were performed to examine potential differences between the experimental conditions concerning the user experience, the sense of presence, and social presence. However, these analyses did not reveal any statistically significant effects. Following this, additional qualitative analyses were carried out to explore the quantitative data in greater depth, aiming to identify trends and insights that could further inform the study's findings.

4.1 Descriptive Statistics

Descriptive statistics were calculated for all measured variables to provide an overview of participants' experiences across different conditions. These include the User Experience Questionnaire - Short Version (UEQ-S), the Igroup Presence Questionnaire (IPQ), the Social Presence Questionnaire (SPQ), and a set of custom questions tailored to assess specific aspects of the simulated interviews. The descriptive analysis highlights the average (M) and variability (SD) within each condition, offering insights into participants' perceptions of usability, presence, and various qualitative aspects of the interview process.

User Experience Questionnaire (UEQ-S). The descriptive analysis of the UEQ-S revealed variations in both pragmatic and hedonic quality across the four conditions. For Pragmatic Quality, mean scores were highest in Condition 2 (FUxUE) ($M = 4.65$, $SD = 1.22$), followed closely by Condition 1 (FUxFE) ($M = 4.60$, $SD = 1.42$) and Condition 3 (UUxFE) ($M = 4.60$, $SD = 1.35$).

Condition 4 (UUxUE) showed the lowest mean score (M = 4.43, SD = 1.52), indicating reduced pragmatic usability in an unfriendly user and environment context. For Hedonic Quality, the highest mean score was observed in Condition 2 (FUxUE) (M = 4.43, SD = 1.13), followed by Condition 1 (FUxFE) (M = 4.39, SD = 1.38). Scores decreased in Condition 3 (UUxFE) (M = 4.26, SD = 1.46) and Condition 4 (UUxUE) (M = 4.14, SD = 1.52), reflecting a diminished sense of enjoyment and engagement in less favorable scenarios.

Igroup Presence Questionnaire (IPQ). The total scores of the IPQ demonstrated consistent perceptions of presence across conditions, with mean values spanning from 3.17 to 3.33. The highest presence score was recorded in Condition 4 (UUxUE) (M = 3.33, SD = 1.15), while the lowest was observed in Condition 2 (FUxUE) (M = 3.17, SD = 1.10). These findings suggest that participants' sense of presence remained relatively stable across scenarios, with minor variations. It is interesting to note that the feeling of presence does not decrease with "unfriendly" conditions.

Social Presence Questionnaire (SPQ). Descriptive statistics for the SPQ showed minor differences in perceived social presence. The highest mean score was observed in Condition 2 (FUxUE) (M = 3.29, SD = 1.06), while the lowest was in Condition 1 (FUxFE) (M = 3.18, SD = 1.14). Condition 4 (UUxUE) (M = 3.26, SD = 1.21) and Condition 3 (UUxFE) (M = 3.19, SD = 1.12) demonstrated slightly higher perceived social presence scores compared to Condition 1.

Table 1 presents the descriptive statistics, including the means and standard deviations, for the User Experience Questionnaire, the Igroup Presence Questionnaire, and the Social Presence Questionnaire across all four conditions.

Table 1. Descriptive Statistics for UEQ-S, IPQ, and SPQ across conditions. The total values represent aggregated mean scores across all items within each respective questionnaire.

Variable	Condition 1 *M (SD)*	Condition 2 *M (SD)*	Condition 3 *M (SD)*	Condition 4 *M (SD)*
Pragmatic Quality (UEQ-S)	4.60 (1.42)	4.65 (1.22)	4.60 (1.35)	4.43 (1.52)
Hedonic Quality (UEQ-S)	4.39 (1.38)	4.43 (1.13)	4.26 (1.46)	4.14 (1.52)
UEQ-S total value	4.50 (1.30)	4.54 (1.18)	4.43 (1.40)	4.28 (1.50)
IPQ total value	3.25 (1.06)	3.17 (1.10)	3.31 (1.02)	3.33 (1.15)
SPQ total value	3.18 (1.14)	3.29 (1.06)	3.19 (1.12)	3.26 (1.21)

Custom Questions (CUSQ). The custom questions assessed participants' perceptions of the interview situation across four conditions, focusing on realism, well-being, discomfort, goal achievement, and challenges experienced.

CUSQ1 (Realism): Participants rated how realistic the interview situation felt. Across the four conditions, ratings ranged from moderately to highly realistic. The highest realism score was observed in Condition 4 (UUxUE) (M = 3.45, SD = 1.36), while Condition 1 (FUxFE) scored the lowest (M = 2.90, SD = 1.33).

CUSQ2 (Well-being): Ratings for participants' well-being during the simulated interview varied slightly across conditions. Condition 4 (UUxUE) yielded the highest average score (M = 4.75, SD = 1.62), suggesting participants felt best in this scenario. Conversely, Condition 1 (FUxFE) showed the lowest mean score (M = 4.00, SD = 1.62).

CUSQ3 (Discomfort): Participants reported their levels of discomfort or pain during the interview. Scores remained relatively low across all conditions, with Condition 3 (UUxFE) showing the lowest average discomfort (M = 3.25, SD = 1.48) and Condition 4 (UUxUE) the highest (M = 3.45, SD = 1.36).

CUSQ4 (Goal Achievement): Participants assessed whether they achieved their goals during the interview. Scores were fairly consistent, with Condition 2 (FUxUE) having the highest mean score (M = 4.75, SD = 1.62), indicating greater perceived goal achievement. The lowest score was observed in Condition 1 (FUxFE) (M = 4.00, SD = 1.62).

CUSQ5 (Challenges): Finally, participants rated the extent of challenges experienced during the interview. Scores were comparable across conditions, with Condition 3 (UUxFE) showing slightly higher levels of challenges (M = 3.25, SD = 1.48) compared to Condition 4 (UU x UE), which had the lowest (M = 3.00, SD = 1.62).

Table 2 presents the descriptive statistics, including the means and standard deviations, for the custom questions.

Table 2. Descriptive Statistics for Custom Questions Across Conditions. The custom questions were assessed on a 7-point Likert scale.

Custom Question	Condition 1 *M (SD)*	Condition 2 *M (SD)*	Condition 3 *M (SD)*	Condition 4 *M (SD)*
CUSQ1 (Realism)	2.90 (1.33)	3.20 (1.40)	3.25 (1.48)	3.45 (1.36)
CUSQ2 (Well-being)	4.00 (1.62)	4.50 (1.40)	4.60 (1.48)	4.75 (1.62)
CUSQ3 (Discomfort)	3.00 (1.33)	3.20 (1.40)	3.25 (1.48)	3.45 (1.36)
CUSQ4 (Goal Achievement)	4.00 (1.62)	4.75 (1.62)	4.60 (1.48)	4.50 (1.40)
CUSQ5 (Challenges)	3.00 (1.33)	3.20 (1.40)	3.25 (1.48)	3.00 (1.62)

4.2 Repeated-Measures ANOVA

To examine differences across the experimental conditions, repeated-measures ANOVAs were conducted for each dependent variable: Pragmatic Quality (UEQ-S), Hedonic Quality (UEQ-S), UEQ-S Total Value, IPQ Total Value, and SPQ

Total Value. While none of the analyses revealed statistically significant differences between conditions, these results are presented to ensure transparency and methodological rigor.

Including these findings allows a comprehensive understanding of the data and ensures that even non-significant outcomes are documented. Furthermore, the effect size measures (η^2_G) provide insights into the magnitude of the observed effects, which may inform future research or guide experimental design adjustments.

Repeated-Measures ANOVA: Pragmatic Quality (UEQ-S). A repeated-measures ANOVA was conducted to examine the effect of condition on pragmatic quality, as assessed by the UEQ-S. The analysis revealed no significant main effect of condition, $F(3, 57) = 0.24$, $p = 0.865$, $\eta^2_G = 0.004$. Mauchly's test indicated that the assumption of sphericity was met, $W = 0.81$, $p = 0.588$. Consequently, no sphericity corrections were applied.

Post-hoc pairwise comparisons using Bonferroni adjustments revealed no significant differences between any pair of conditions ($p > 0.05$).

Repeated-Measures ANOVA: Hedonic Quality (UEQ-S). The analysis for hedonic quality, measured by the UEQ-S, indicated no significant differences between conditions, $F(3, 57) = 0.81$, $p = 0.495$, $\eta^2_G = 0.007$. Mauchly's test confirmed that the sphericity assumption was not violated ($W = 0.74$, $p = 0.375$), and sphericity corrections were therefore unnecessary.

Pairwise comparisons with Bonferroni adjustments did not reveal significant differences across conditions ($p > 0.05$).

Repeated-Measures ANOVA: Total Value of UEQ-S. A repeated-measures ANOVA was conducted to evaluate the effect of the condition on the total value of the UEQ-S. The analysis revealed no significant main effect of condition, $F(3, 57) = 0.64$, $p = 0.594$, $\eta^2_G = 0.007$. Mauchly's test indicated that the assumption of sphericity was met ($W = 0.72$, $p = 0.319$), and sphericity corrections were therefore unnecessary.

Post-hoc pairwise comparisons using Bonferroni adjustments showed no significant differences between any pair of conditions ($p > 0.05$).

Repeated-Measures ANOVA: Total Value of IPQ. The repeated-measures ANOVA assessing the impact of condition on the total value of the IPQ revealed no significant main effect of condition, $F(3, 57) = 0.22$, $p = 0.879$, $\eta^2_G = 0.004$. Mauchly's test indicated a violation of the sphericity assumption ($W = 0.39W = 0.39$, $p = 0.005$). However, sphericity corrections using Greenhouse-Geisser ($p = 0.815$) and Huynh-Feldt ($p = 0.840$) estimates did not alter the non-significant result.

Post-hoc pairwise comparisons with Bonferroni adjustments revealed no significant differences between any pair of conditions ($p > 0.05$).

Repeated-Measures ANOVA: Total Value of SPQ. The analysis for the total value of the SPQ showed no significant main effect of condition, $F(3, 57) = 0.21$, $p = 0.887$, $\eta^2_G = 0.002$. Mauchly's test confirmed that the assumption of sphericity was not violated (W = 0.69W = 0.69, $p = 0.246$), so no corrections were applied.

Post-hoc pairwise comparisons using Bonferroni adjustments revealed no significant differences between any pair of conditions ($p > 0.05$).

4.3 Qualitative Results

Qualitative feedback was collected to provide deeper insights into participants' experiences during the simulations. Participants were asked open-ended questions about the virtual environment, interactions with avatars, challenges faced, and areas for improvement. The responses were analyzed thematically, highlighting the critical aspects of their experiences.

Environment. Participants' perceptions of the simulated environment varied significantly. Many appreciated the immersive quality of certain conditions, with one participant noting, *"The friendly environment felt welcoming and made the tasks more manageable."* However, unfriendly conditions were described as "cold" and "distracting," with participants reporting that the visual and auditory elements were sometimes exaggerated, reducing realism. Some mentioned inconsistencies, such as static objects or limited interactivity, which detracted from the overall experience.

Interaction with Avatars. Feedback on avatar interactions was similarly mixed. Participants appreciated that avatars introduced a human element into the simulation, with one stating, *"The avatars helped simulate real conversations, which was engaging."* However, others noted that the avatars' behavior sometimes felt "robotic" or "repetitive," with limited adaptability to user input. Participants suggested enhancing the avatars' responsiveness and increasing the variety in their communication styles to make interactions more dynamic and realistic.

Challenges Encountered. Participants reported a range of challenges, primarily related to technical issues and task complexity. For example, lag and delayed responses occasionally disrupted the flow of tasks, with one participant mentioning, *"The system froze briefly, which broke my concentration."* Additionally, some found the cognitive load overwhelming, particularly in conditions with both unfriendly users and environments, describing it as "stressful to the point of distraction."

Suggestions for Improvement. Participants proposed several improvements to enhance the simulation experience. Many emphasized the need for greater realism in environmental and avatar interactions. For example, one participant recommended, *"Making objects in the environment respond to user actions would make the scenarios more realistic."* Others suggested optimizing system performance to minimize technical issues and introducing more varied scenarios to increase engagement and reflect real-world complexity.

5 Discussion

The present study aimed to evaluate participants' experiences and interactions within simulated sales scenarios by examining user experience, presence, and social presence across different experimental conditions. Although the results did not yield statistically significant differences between conditions, the findings provide valuable insights into the nuances of participants' experiences and suggest potential areas for improvement in future simulations.

User Experience and Perceived Quality. The descriptive statistics for the User Experience Questionnaire indicated that pragmatic and hedonic quality ratings varied slightly across conditions. Conditions with friendly users (FUxUE and FUxFE) tended to yield higher scores for both pragmatic and hedonic quality, suggesting that a more welcoming and supportive context enhances usability and enjoyment. Interestingly, the lowest scores for both dimensions were observed in the condition with unfriendly users and an unfriendly environment (UU x UE), reinforcing the importance of fostering a positive interactional and environmental context to optimize user experience.

Despite these trends, the repeated-measures ANOVAs revealed no statistically significant effects of condition on pragmatic or hedonic quality. This finding may suggest that while users are sensitive to contextual variations, the overall impact of these variations on their experience may not be strong enough to produce measurable differences within the scope of this study. Alternatively, the measures employed or the sample size might not have been sufficient to detect small but meaningful effects.

Presence and Social Presence. The Igroup Presence Questionnaire (IPQ) and Social Presence Questionnaire (SPQ) results showed stable perceptions of presence across conditions. Notably, the condition with unfriendly users and an unfriendly environment (UUxUE) did not result in diminished presence, contrary to expectations. This finding suggests that participants' sense of "being there" and their perception of social presence may be more resilient to adverse contextual factors than previously assumed. However, qualitative feedback revealed that unfriendly conditions were often perceived as less immersive or realistic, highlighting a potential disconnect between quantitative measures of presence and participants' subjective experiences.

Custom Questions and Qualitative Insights. Custom questions provided additional insights into participants' perceptions of realism, well-being, discomfort, goal achievement, and challenges across conditions. Realism and well-being scores were highest in the UUxUE condition, an unexpected finding given the unfriendly context. This result may reflect participants' adaptation to challenging scenarios or a heightened sense of accomplishment in overcoming adversity. However, qualitative feedback highlighted that unfriendly conditions were sometimes described as "cold" or "distracting," suggesting that while participants adapted, their experiences were not uniformly positive. Participants also identified challenges such as technical issues, high cognitive load, and limited interactivity in the simulations. These challenges were particularly pronounced in conditions involving both unfriendly users and environments, where participants reported feeling overwhelmed or distracted.

Lesson Learned. Finally, the set of descriptive results and qualitative feedback was distilled into a set of lessons learned that can be useful for the development of similar scenarios:

- **Context Matters:** Friendly environments and interactions appear to enhance user experience, suggesting that incorporating supportive and engaging elements can improve usability and enjoyment. However, designers should also consider how to make unfriendly scenarios realistic yet manageable, as they are often necessary for training purposes.
- **Enhancing Realism:** Participants emphasized the need for greater realism in both environmental and avatar interactions. Features such as dynamic object behavior, more responsive avatars, and varied communication styles could make simulations more engaging and reflective of real-world scenarios.
- **Minimizing Technical Issues:** Technical disruptions, such as lag or system freezes, were reported to disrupt participants' focus and immersion. Optimizing system performance should be a priority to ensure a seamless experience.
- **Balancing Cognitive Load:** High task complexity and simultaneous challenges were reported as overwhelming in some conditions. Future simulations should aim to balance cognitive demand to maintain user engagement without inducing excessive stress.

6 Conclusion

To conclude, immersive media holds immense potential for experiential learning and training in sales and negotiation contexts. By offering realistic, and immersive experiences, platforms such as the VR CAVE provide learners with state-of-the-art training opportunities, overcoming barriers related to geographical, and time constraints. The experiment demonstrated that students were highly engaged, curious, and motivated to test new sales scenarios, particularly by gaining initial experience in handling challenging interactions with unfriendly customers. While the study did not reveal statistically significant differences between conditions, it identified important trends and insights that can guide the design of future simulations. By prioritizing positive user interactions, improving realism, and addressing technical challenges, virtual environments can be enhanced to better meet user needs. Future research should delve deeper into the complex interactions between contextual factors, user perceptions, and system design to further refine the development of effective and engaging virtual simulations.

6.1 Limitations and Future Work

This study has several limitations that should be addressed in future research. First, the lack of statistically significant differences across conditions may be due to a limited sample size, which could reduce statistical power. Future studies should consider larger sample sizes to better detect subtle effects. Second, while the quantitative measures provided valuable insights, they may not fully capture participants' subjective experiences. Integrating more qualitative methods, such as in-depth interviews or focus groups, could provide a richer understanding of user perceptions. Third, the conversation with the avatars was relatively narrow, following a premade script. In the near future, the integration of LLM will be explored to foster the training effectiveness. Our hypothesis is that it will allow participants to engage in dynamic, context-dependent dialogues within realistic settings, enabling them to respond flexibly to customer inquiries and objections. Additionally, the integration of multimodal feedback mechanisms, which extend beyond visual and auditory feedback to include emotional response tracking, could further enhance the training's effectiveness.

Acknowledgments. This work was supported by the European Union's Horizon Europe programme under grant number 101092875 "DIDYMOS-XR" (https://www.didymos-xr.eu).

In this paper, we used Overleaf's built-in spell checker, the current version of ChatGPT (GPT 4.0), and Grammarly. These tools helped us fix spelling mistakes and get suggestions to improve our writing. If not noted otherwise in a specific section, these tools were not used in other forms.

References

1. Alhalabi, W.: Virtual reality systems enhance students' achievements in engineering education. Behav. Inf. Technol. **35**, 919–925 (2016). https://doi.org/10.1080/0144929X.2016.1212931
2. Nassar, A.K., Al-Manaseer, F., Knowlton, L.M., Tuma, F.: Virtual reality (VR) as a simulation modality for technical skills acquisition. Ann. Med. Surgery **71**(2021), 102945 (2021). https://doi.org/10.1016/j.amsu.2021.102945
3. Biocca, F., Harms, C., Burgoon, J.K.: Toward a more robust theory and measure of social presence: review and suggested criteria. Pres. Teleoperators Virt. Environ. **12**(5), 456–480 (2003). https://doi.org/10.1162/105474603322761270

4. Blascovich, J., Bailenson, J.: Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn Of The Virtual Revolution. Harper Collins, New York (2011)
5. Chen, X., Chen, Z., Li, Y., He, T., Hou, J., Liu, S., He, Y.: ImmerTai: immersive motion learning in VR environments. J. Vis. Commun. Image Represent. **58**, 416–427 (2019). https://doi.org/10.1016/j.jvcir.2018.11.039
6. Familoni, B.T., Onyebuchi, N.C.: Augmented and virtual reality in U.S. education: a review: analyzing the impact, effectiveness, and future prospects of AR/VR tools in enhancing learning experiences. Int. J. Adv. Res. Soc. Sci. **6**(4) (2024). https://doi.org/10.51594/ijarss.v6i4.1043
7. Franke, T., Attig, C., Wessel, D.: A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. Int. J. Hum.-Comput. Interact. **35**(6), 456-467 (2019). https://doi.org/10.1080/10447318.2018.1456150
8. Howard, M., Gutworth, M., Jacobs, R.: A meta-analysis of virtual reality training programs. Comput. Hum. Behav. **121**, 106808 (2021). https://doi.org/10.1016/J.CHB.2021.106808
9. Kyan, M., et al.: An approach to ballet dance training through MS kinect and visualization in a cave virtual reality environment. ACM Trans. Intell. Syst. Technol. (TIST) **6**, 1–37 (2015). https://doi.org/10.1145/2735951
10. Lim, J.: A conceptual framework on the adoption of negotiation support systems. Inf. Softw. Technol. **45**(8), 469–477 (2003)
11. Maslova, K., Gasimov, A., Konovalova, A.: Using virtual reality to develop emotional intelligence. Eur. Psychiatry **65**(S1), S244–S245 (2022). https://doi.org/10.1192/j.eurpsy.2022.631
12. Mystakidis, S., et al.: Design, development, and evaluation of a virtual reality serious game for school fire preparedness training. Educ. Sci. (2022). https://doi.org/10.3390/educsci12040281
13. Moonen, N., Heller, J., Hilken, T., Danny Han, D.I., Mahr, D.: Immersion or social presence? Investigating the effect of virtual reality immersive environments on sommelier learning experiences. J. Wine Res. **35**(2), 101–118 (2024). https://doi.org/10.1080/09571264.2024.2310297
14. Regal, G., et al.: Questionnaires embedded in virtual environments: reliability and positioning of rating scales in virtual environments. Qual. User Experience **4**(1), 1–13 (2019). https://doi.org/10.1007/s41233-019-0029-1
15. Rifdi, N.I.A., Sunar, M.S., Sa'Adon, M.N.I.: Gamifying shooting training in cave automatic virtual environment (CAVE). In: 2022 2nd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Bandung, Indonesia, pp. 198–203 (2022). https://doi.org/10.1109/ICICyTA57421.2022.10037856
16. Schrepp, M., Hinderks, A., Thomaschewski, J.: Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). Int. J. Interact. Multimedia Artif. Intell. **4**(6), 103–108 (2017). https://doi.org/10.9781/ijimai.2017.09.001
17. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. Pres. Teleoperators Virt. Environ. **10**(3), 266–281 (2001). https://doi.org/10.1162/105474601300343603
18. Smith, S., Ericson, E.: Using immersive game-based virtual reality to teach fire-safety skills to children. Virtual Reality **13**, 87–99 (2009). https://doi.org/10.1007/s10055-009-0113-6
19. Stephens, R., Awasthi, A., Crowley, K., Boyle, F., Walsh, J.: A literature review of virtual reality interpersonal training for salespeople. In: 2021 32nd Irish Signals and Systems Conference, ISSC 2021, IEEE (2021). https://doi.org/10.1109/ISSC52156.2021.9467845

20. Stinson, C., Bowman, D.: Feasibility of training athletes for high-pressure situations using virtual reality. IEEE Trans. Vis. Comput. Graph. **20**, 606–615 (2014). https://doi.org/10.1109/TVCG.2014.23
21. Upadhyay, A., Khandelwal, K.: Virtual reality: adding immersive dimension to sales training. Hum. Resour. Manage. Int. Digest **26**, 42–45 (2018). https://doi.org/10.1108/HRMID-01-2018-0014
22. Voigt-Antons, J.N., Lehtonen, E., Palacios, A.P., Ali, D., Kojic, T., Möller, S.: Comparing emotional states induced by 360 videos via head-mounted display and computer screen. In: 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–6). IEEE, May 2020
23. Vona, F., Ćeranić, M., Rybnikova, I., Voigt-Antons, J.N.: Designing user-centered simulations of leadership situations for cave automatic virtual environments: development and usability study. In: International Conference on Human-Computer Interaction, pp. 324–331. Springer, Cham, July 2023
24. Werrlich, S., Lorber, C., Nguyen, P.-A., Yanez, C., Notni, G.: Assembly training: comparing the effects of head-mounted displays and face-to-face training. In: Chen, J., Fragomeni, G. (eds.) VAMR 2018. LNCS, vol. 10909, pp. 462–476. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91581-4_35
25. Xiu, Y., Liu, X., Yuan, Y., Zhao, H., Zhang, C.: A new method of dance rhythm training based on an immersive cave automatic virtual environment. IEEE Access **11**, 109422–109434 (2023). https://doi.org/10.1109/ACCESS.2023.3317886

Research on the Application of Tangible Interaction in Mixed Reality for Dental Implant Teaching

Zengyu Xiong(✉), Yuxuan Li, and Xing Fang

Wuhan University of Technology, 122 Luoshi Road, Wuhan, Hubei, People's Republic of China
871513062@qq.com

Abstract. Objective: This study aims to evaluate medical students' acceptance of a den-tal implant education system that integrates Mixed Reality (MR) technology with Tangible Interaction (TI). Methods: In a secure and comfortable environment, 250 medical students completed a series of dental implant tasks, including delicate tooth extraction operations and precise implant placement, using the Meta Quest 3 Head-Mounted Display (HMD) and wireless handheld controllers. Following this, participants completed a questionnaire regarding their perceptions and acceptance of the MR tangible interaction educational system. The study design is grounded in the Technology Ac-ceptance Model (TAM), incorporating tangible interaction as an additional variable to assess its role in the learning process. Results: Results indicate that 83.7% of respondents held positive attitudes towards utilizing MR tech-nology. The study hypotheses were confirmed, demonstrating that tangible interaction within mixed reality significantly enhances perceived usefulness and perceived ease of use, thereby strengthening user attitudes and ultimate usage intention. Conclusion: Mixed Reality technology effectively improves the teaching outcomes of complex skills such as dental implant surgery by providing high-quality educational resources, enhancing the realism of operational practice, and increasing the precision of guidance, while overcoming the limitations of traditional physical models.

Keywords: Mixed Reality (MR) · Tangible Interaction · Dental Implant Education · Technology Acceptance Model (TAM) · Medical Education

1 Introduction

With the rapid development of information technology, medical education is gradually shifting towards a reliance on technological and digital resources. Electronic devices and online platforms are being utilized for information retrieval, simulation training, and distance learning, while computer simulations and virtual reality (VR) technologies have enhanced students' clinical skills and decision-making abilities [1, 2]. Teaching in dental implant surgery encompasses multiple stages of precise operations such as implantation and restoration, emphasizing stringent quality control. The quality of educational resources [3], the realism of operational practice [4], and the accuracy of guidance

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 268–285, 2025.
https://doi.org/10.1007/978-3-031-93715-6_17

[5] directly impact the effectiveness of dental implant surgery teaching. Besides using physical models for instruction, virtual technologies are commonly employed to ensure the efficacy of dental implant surgical education [6]; these technologies aid students in understanding procedural details but lack tangible interaction [7].

Moreover, although VR technology can be used for simulating experiments and exploring procedures [8], students still need to engage in actual operations to gain direct experience with material properties and physical laws. Therefore, combining virtual technology with practical operations in dental implant surgery education can better promote the mastery of theoretical knowledge and practical skills by students. Insufficient or unstructured practical operation training can affect the cultivation of clinical skills and overall teaching quality.

By providing high-quality educational resources, enhancing the realism of operational practice, and improving the accuracy of guidance, interactive 3D virtual models can increase the precision of surgical training while eliminating the physical limitations imposed by physical models. In traditional VR-based surgical training, the absence of true haptic feedback makes it difficult for trainees to accurately master the manipulation force and feel of surgical instruments, limiting training effectiveness [9, 10]. Medical students need to transfer from a virtual simulation environment to physical models for validation, which may impact their spatial perception ability, causing distraction and extending the learning curve, thus affecting teaching quality and skill acquisition.

To address this issue, Augmented Reality (AR) technology has been introduced into surgical education to improve spatial relationship perception and overall teaching outcomes. AR technology enhances students' understanding and mastery of complex surgical techniques and spatial layouts [11] by overlaying virtual information onto the real environment, increasing satisfaction and content retention [12], and enhancing learning effects through reduced monotony, increased interactivity, and provision of immediate feedback [13]. However, current AR educational technologies cannot freely adjust the size and position of 3D models or provide multi-angle perspectives, potentially limiting medical students' comprehensive understanding of spatial relationships.

Mixed Reality (MR) [14] integrates elements of both the real world and virtual world's digital content, combining the advantages of AR [15] and VR [16]. By reconstructing 3D models of real objects, MR technology generates intuitive and accurate three-dimensional images that enable medical students to clearly identify key structures of the surgical site, thereby creating realistic surgical training scenarios. Under the guidance of the teaching system, an interactive feedback loop between the real and virtual worlds is formed. MR technology supports dental implant surgery teaching by offering a highly immersive experience through the interaction of virtual hands and models in a simulated 3D digital environment.

Previous studies on dental implant teaching have mainly involved physical models [17], VR [18], and AR [19] instruction; however, research based on MR technology remains limited. Therefore, this study aims to propose a dental implant teaching system based on MR technology, grounded in the Technology Acceptance Model (TAM), analyzing user behavior and intentions, and employing Structural Equation Modeling (SEM) to quantify path impacts, providing support and guidance for medical education.

2 Theoretical Models and Research Hypotheses

2.1 TAM Model

Davis [20] introduced the Technology Acceptance Model (TAM) in 1989, which has since been widely applied to explain the degree of acceptance and usage behavior towards new technologies. Grounded in the Theory of Reasoned Action (TRA), TAM elaborates that the use behavior of information technology is primarily influenced by individuals' inherent intention to use, highlighting the importance of Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) in this process [20, 21]. Therefore, this paper employs TAM to analyze users' acceptance and usage behavior regarding the implant teaching method based on tangible interaction within Mixed Reality (MR).

2.2 Research Hypotheses

Figure 1 illustrates the research model for this study, which is constructed based on the Technology Acceptance Model (TAM) and integrates Tangible Interaction (TI) as an external factor. This integration aims to examine the impact of TI on students' acceptance in the context of learning dental implant surgery.

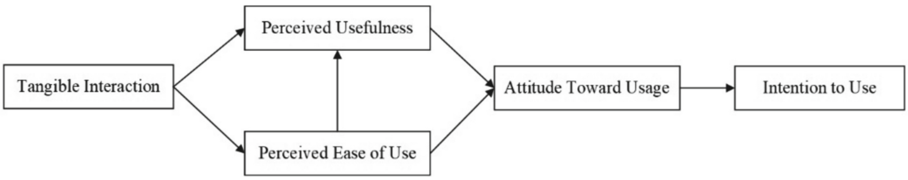


Fig. 1. Theoretical research model

Tangible Interaction. Tangible Interaction (TI) refers to the natural and intuitive interaction between users and virtual objects as well as physical objects in a mixed reality environment [22, 23]. In this study, TI refers to the ability of medical students to interact with virtual models such as teeth and jawbone models using gestures. According to research, increasing sensory input modalities in a virtual environment can significantly enhance the user's sense of presence and memory of the environment and its objects [24]. Based on this, it is reasonable to infer that tangible interaction in mixed reality will improve the user's perceived usefulness due to the multisensory experience it provides.

Hypothesis 1 (H1): Tangible Interaction in mixed reality has a positive effect on its perceived usefulness.

Real-time computation of mechanical interactions between real and virtual objects can enhance the natural experience for users in mixed reality applications. This approach combines tools from machine learning, computer vision, and computer graphics, allowing users to naturally interact with deformable virtual objects [22]. Therefore, in this study, Tangible Interaction (TI) in mixed reality may affect medical students' perceived ease of use in learning dental implant surgery.

Hypothesis 2 (H2): Tangible Interaction in mixed reality has a positive effect on its perceived ease of use.

Perceived Usefulness. Perceived usefulness is one of the core variables in the Technology Acceptance Model (TAM) [20], referring to the degree to which a user believes that using an information technology will enhance their job or life performance [22]. Research indicates that perceived usefulness is highly correlated with both current and future usage behavior of users [20]. In this study, perceived usefulness refers to the learning effectiveness and improvement in clinical skills that medical students experience when using mixed reality technology for learning dental implant procedures. If medical students believe that this technology significantly improves learning efficiency and practical ability, they are more likely to use this technology actively and continuously.

Hypothesis 3 (H3): Perceived usefulness has a positive effect on attitude toward usage.

Perceived Ease of Use. Perceived ease of use is another core variable in the Technology Acceptance Model (TAM) [20], referring to the degree to which a user believes that using an information technology will be free of effort [22]. Simple and easy-to-use technology can reduce learning costs and operational complexity, allowing users to more easily enjoy the benefits brought by the technology, thereby enhancing positive evaluations and the intention to use it. In this study, perceived ease of use refers to the operational simplicity and learning fluidity that medical students experience when using mixed reality technology for learning dental implant procedures. If they find the technology easy to use, they are more likely to develop a positive attitude and continue using it.

Hypothesis 4 (H4): Perceived ease of use has a positive effect on perceived usefulness.

Hypothesis 5 (H5): Perceived ease of use has a positive effect on attitude toward usage.

Attitude Toward Usage and Intention to Use. In this study, attitude toward usage and intention to use refer to the medical students' subjective evaluation of tangible interaction in mixed reality and their subjective likelihood of actively engaging with and utilizing this technology. Similar to how DAVIS established the relationship between these two variables when he initially developed the Technology Acceptance Model (TAM) [25], in the context of learning dental implant surgery, the higher the level of acceptance by medical students of tangible interaction in mixed reality, the more likely they are to actively participate and effectively leverage these technologies for learning and practice.

Hypothesis 6 (H6): Attitude toward usage has a positive effect on intention to use.

3 Experimental Methods

3.1 Participant

Participants were recruited through contacts with students and faculty members of local universities in Wuhan, as well as via local communities and online recruitment methods. This study conducted research using an online questionnaire system, employing a random sampling method to collect sample data, resulting in the collection of 250 questionnaires. After screening, a final total of 232 valid questionnaires were obtained, with an effective response rate of 92.8%. Regarding the basic characteristics of the sample, among the respondents, there were 136 females and 96 males; the highest proportion of respondents was aged between 18–25 years old (43.5%, 101 individuals), followed by those aged 26–30 years old (33.6%, 78 individuals). Respondents under 18 years old accounted for 3.9% (9 individuals), while those over 30 years old accounted for 19.0% (44 individuals).

3.2 Facility

The Head-Mounted Display (HMD) - Meta Quest 3 (Fig. 2) was utilized to assist participants in navigating the virtual environment and conducting the experiment. Wireless handheld controllers were used for manipulating virtual objects, such as pointing, clicking, and grabbing. The built-in tracking system of the Meta Quest 3 accurately tracks the position of the HMD and handheld controllers, capturing the participant's movements in real-time to ensure immersion and interaction accuracy. For safety and comfort, the experiment was conducted in a relatively empty room, and virtual boundaries were set by identifying the ground space to prevent collisions due to insufficient space.

Fig. 2. Meta Quest 3 (Source: Internet)

3.3 Experimental Planning

Experimental Process Design. Before the experiment began, the researchers provided participants with a detailed explanation of the study protocol and obtained their written informed consent. Following this, the Meta Quest 3 HMD and handheld controllers were calibrated to ensure accuracy and reliability. Researchers then thoroughly instructed the participants on how to wear and use the equipment, familiarizing them with the operation methods through an immersive virtual reality dental implant education system (Fig. 3). Several trial runs were conducted until the researchers confirmed that the participants could skillfully operate the equipment.

In the formal experiment, participants followed instructions to complete a series of dental implant tasks, including precise tooth extraction, accurate implant placement, secure screwing in of the implant, firm locking of the implant, snug installation of the abutment, and seamless connection of the crown. After completing all tasks, participants filled out a questionnaire collecting basic information such as age, gender, education level, and experience with MR technology, as well as their opinions and acceptance of the MR-based tangible interaction dental implant education system. The entire experiment took approximately 10 min to complete.

Fig. 3. Participants wear and use the device

Specific Experimental Procedure. As shown in Fig. 4, the specific steps for mixed reality dental implant surgery instruction are as follows:

(a) Fine tooth extraction: Using specialized forceps, accurately align with the tooth root and apply gentle yet steady force to ensure the tooth is removed smoothly and intact.
(b) Implant with precision: Employing a precision holding tool, place the implant into the predetermined position with exactness, laying a solid foundation for subsequent steps.
(c) Firmly screw into the implant: Utilize a specially designed drill bit to carefully rotate and embed the implant into the prepared site, ensuring a tight and stable integration with bone tissue. It's important to maintain even pressure throughout this process for optimal fit.
(d) Secure the implant: Carefully adjust the implant screw until it meshes perfectly with the implant, achieving secure positioning that ensures long-term stability.
(e) Close mounting abutment: Insert the abutment precisely into the implant and make fine adjustments to ensure there are no gaps between the two, establishing a stable platform for crown installation.
(f) Seamless crown: Install the crown with meticulous care, making micro-adjustments to ensure perfect alignment with the abutment. This not only restores tooth function but also maintains aesthetics, achieving an optimal restoration effect.

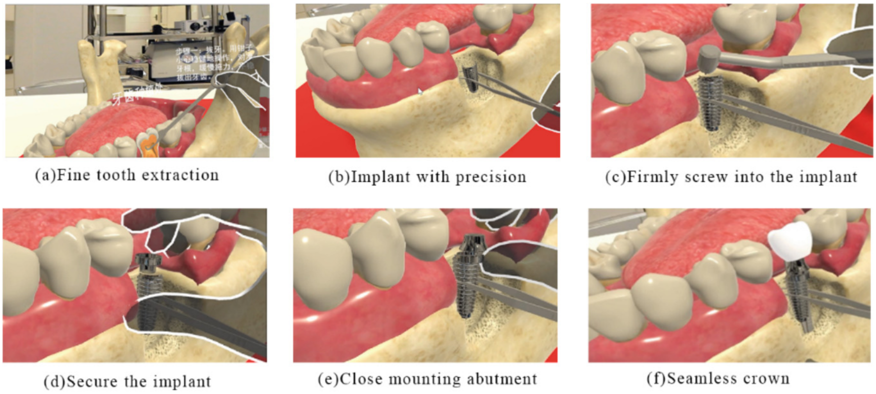


Fig. 4. Experimental Procedure Diagram

3.4 Outcome Measurement

This study utilized online questionnaires to collect data on medical students' learning. During the questionnaire design phase, classic scales from relevant research literature both domestically and internationally were referenced, with appropriate modifications made according to tangible interaction in mixed reality. The questionnaire is divided into two sections: the first section covers the basic information of the respondents, including gender, age, education level, and their familiarity with tangible interaction in mixed reality; the second section encompasses measurement variables for each latent variable within the research model. Each variable was estimated using multi-dimensional scales sourced from related reference literature. A Likert 5-point scale was used, where 1 point indicates "Strongly Disagree," and 5 points indicate "Strongly Agree." The questionnaire consists of five sections with a total of 13 items, the specific content of which can be found in Table 1.

Table 1. The indicator system of model variables and references

Variable	Measurement Indicators	Scale Items	References
Tangible Interaction(TI)	TI1	In mixed reality learning activities, I feel that my interaction with virtual objects is very natural	Badía et al. [22]

(*continued*)

Table 1. (*continued*)

Variable	Measurement Indicators	Scale Items	References
	TI2	I can easily manipulate tools through tangible interaction in the mixed reality environment to practice skills	
	TI3	Tangible interactions make me feel as if I am really performing complex skill operations	
Perceived Usefulness(PU)	PU1	Using tangible interactions in mixed reality allows me to understand abstract learning content in a more intuitive and vivid way	Davi et al. [20]
	PU2	Tangible interactions in mixed reality increase my excitement and engagement in learning, making me learn more proactively	
	PU3	Through practical simulations using tangible interactions in mixed reality, I feel that my ability to solve real-world problems has improved	
	PU4	When learning through tangible interactions in mixed reality, I can interact with virtual objects effortlessly as if they really exist	
Perceived Ease of Use(PEOU)	PEOU1	When learning through tangible interactions in mixed reality, I can interact with virtual objects effortlessly as if they really exist	Davi et al. [20]

(*continued*)

Table 1. (*continued*)

Variable	Measurement Indicators	Scale Items	References
	PEOU2	Navigation and guidance features in the mixed reality environment allow me to move freely in virtual space without easily losing direction	
	PEOU3	The frequency of technical issues I encounter during the use of tangible interactions in mixed reality learning, such as lag or system stuttering, is low and does not affect the overall learning experience	
Attitude Toward Usage(ATU)	ATU1	I am very interested in using tangible interactions in mixed reality for learning because it provides a novel learning experience	Malatj et al. [26] Andye et al. [27]
	ATU2	I believe that tangible interactions in mixed reality can help me understand complex concepts more effectively because it allows me to interactively learn between the virtual and real worlds	
	ATU3	I look forward to mixed reality learning environments immersing me in the learning content through tangible interactions as if I were there, thereby increasing my engagement and memory retention	

(*continued*)

Table 1. (*continued*)

Variable	Measurement Indicators	Scale Items	References
Intention to Use(IU)	IU1	I am very excited about integrating tangible interactions in mixed reality into learning because it offers unique interactive and immersive learning experiences	Malatji et al. [26] Andy et al. [27]
	IU2	I find it easier to concentrate in the immersive learning environment created by mixed reality, which helps me deeply understand and remember information	
	IU3	Even after the first attempt, I am willing to frequently use tangible interactions in mixed reality for learning because I believe it will continuously enhance my learning outcomes	

4 Outcome

4.1 Sample Characteristics

This study conducted a survey using WJX (Questionnaire Star), an online questionnaire system, and collected sample data through random sampling. A total of 250 questionnaires were collected. After screening the questionnaires, 232 valid responses were obtained, resulting in a response validity rate of 92.8%.

From the basic characteristics of the sample (Table 2), it can be seen that among the respondents, 58.6% are female, which is higher than the proportion of males; the age group with the highest representation is 18 to 25 years old (43.5%), followed by those aged 26 to 30 (31.2%); educational levels are primarily concentrated at the bachelor's degree level (33.6%). Regarding the familiarity with tangible interaction, 54.3% of the surveyed individuals indicated that they have heard of it but do not understand it well, while 6.9% reported a higher level of understanding. Moreover, 83.7% of the respondents expressed a willingness to use tangible interaction.

Table 2. The basic characteristics of samples

Question	Item	Number of people (persons)	Proportion (%)
Gender	Male	96	41.4%
	Female	136	58.6%
Age	Under 18 years old	9	3.9%
	18 to 25 years old	101	43.5%
	26 to 30 years old	78	33.6%
	Over 30 years old	44	19.0%
Education Background	Associate Degree or Below	32	13.8%
	Bachelor's Degree	122	52.6%
	Master's Degree or Above	78	33.6%
Level of Familiarity with Tangible Interaction	Not Heard Of	50	21.6%
	Heard Of, But Not Familiar	126	54.3%
	Used, Somewhat Familiar	40	17.2%
	Very Familiar	16	6.9%
Willingness to Use Tangible Interaction	Willing	194	83.7%
	Unwilling	38	16.4%

4.2 Data Analysis and Hypothesis Testing

Reliability Testing. Reliability testing was conducted on the questionnaire data (Table 3), and Cronbach's α for all five latent variables exceeded 0.7. The overall Cronbach's α for the questionnaire was 0.94, indicating that the scales designed in the questionnaire have good reliability and a high level of data credibility.

The confirmatory factor analysis (CFA) of the scale was performed using AMOS 27.0 software. The unstandardized estimates were all significant at the 0.05 level (all less than 0.001). As shown in Table 3, in this study, the factor loadings and average variance extracted (AVE) for each observed variable were greater than 0.500, and the composite reliability (CR) was greater than 0.700, meeting the standard requirements for convergent validity. This indicates that the measurement scales used in this study have good convergent validity.

Validity Testing. This study utilized SPSS 27.0 to conduct the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity on the sample. As shown in Table 4, the overall KMO value for the scale is 0.881, and Bartlett's test of sphericity is significant at the 0.05 level. This indicates that the sample data are suitable for factor analysis.

Table 3. Reliability and convergent validity

Latent Variables	Parameter Significance Estimates				Item Reliability		Cronbach's Alpha	Composite Reliability	Convergent Validity
	Unstd.	S.E.	Z	P	Std.	SMC		CR	AVE
TI	1				0.923	0.878	0.865	0.824	0.934
	0.776	0.035	22.267	***	0.907	0.852			
	0.761	0.036	21.376	***	0.893	0.851			
PU	1				0.923	0.822	0.876	0.808	0.944
	0.784	0.037	21.332	***	0.884	0.822			
	0.806	0.037	22.063	***	0.895	0.806			
	0.779	0.035	22.138	***	0.893	0.805			
PEOU	1				0.907	0.805	0.891	0.811	0.928
	0.868	0.041	21.019	***	0.898	0.802			
	0.834	0.042	19.908	***	0.897	0.801			
ATU	1				0.895	0.798	0.79	0.790	0.919
	0.762	0.041	18.701	***	0.874	0.797			
	0.807	0.041	19.45	***	0.897	0.785			
IU	1				0.937	0.782	0.812	0.749	0.899
	0.774	0.043	18.006	***	0.886	0.764			
	0.668	0.043	15.599	***	0.765	0.585			

Table 4. Results of KMO and Bartlett's test of sphericity

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy		0.881
Bartlett's Test of Sphericity	Approximate Chi-Square	3602.017
	Degrees of Freedom	120
	Significance (P)	0

4.3 Structural Equation Modeling Analysis

Structural Equation Modeling (SEM) is a multivariate statistical technique primarily used to test hypothesized relationships among variables. It represents theories through a system of theoretical linear equations and can handle multiple causal relationships as well as unobservable (latent) variables [28]. SEM is particularly suited for exploring complex variable relationships, such as mediation effects, moderation effects, and the measurement of latent variables [29, 30]. In this study, SEM was utilized to analyze and validate the hypothesized relationships within the Technology Acceptance Model (TAM).

Table 5. Model suitability test

Indicators	Reference Standards	Observed Results
CMIN/DF	1–3 is excellent, 3–5 is good	1.658
RMSEA	<0.05 is excellent, <0.08 is good	0.053
IFI	>0.9 is excellent, > 0.8 is good	0.983
TLI	>0.9 is excellent, > 0.8 is good	0.978
CFI	>0.9 is excellent, > 0.8 is good	0.983
GFI	>0.9 is excellent, > 0.8 is good	0.958

Table 6. The test results of the path relationship of the SEM model

Path Relationships			Estimate	S.E.	C.R.	*P*	**Significance**
PEOU	< ---	TI	0.545	0.05	8.066	***	Significant
PU	< ---	TI	0.34	0.067	4.742	***	Significant
PU	< ---	PEOU	0.369	0.09	5.153	***	Significant
AT	< ---	PU	0.23	0.053	3.355	***	Significant
AT	< ---	PEOU	0.548	0.069	7.739	***	Significant
BI	< ---	AT	0.499	0.088	7.772	***	Significant

Note: *** indicates $p < 0.001$, ** indicates $p < 0.01$, * indicates $p < 0.05$.

Goodness-of-Fit. To ensure that the data conform to the theoretical model, it is necessary for the model fit indices to meet established criteria. The proposed model was tested using AMOS software. The fit indices for each model in the initial theoretical model are shown in Table 5. By randomly drawing observations from the original dataset, 2000 bootstrap subsamples were generated to assess the variability of statistics and enhance the reliability of the model. After model modifications, the fit indices showed a Chi-Square to Degrees of Freedom ratio (CMIN/DF) of 1.658 (<3), RMSEA < 0.08, GFI, CFI, and TLI all greater than 0.9. These results meet the reference standards, indicating that the theoretical model proposed in this study fits the sample data very well.

Hypothesis Testing and Result Analysis. Based on the results obtained from AMOS software, we conducted tests on the standardized path coefficients and the significance probability (p-values) of the unstandardized estimates. Table 6 lists the standardized path coefficients of the model along with the verification status of each hypothesis, indicating that all proposed hypotheses were supported. The standardized correlations of the modified model are shown in Fig. 5.

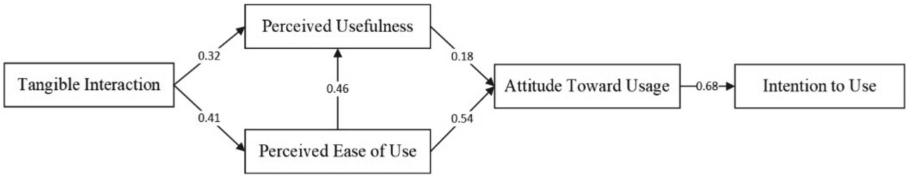


Fig. 5. The standardized correlations of the modified model

5 Discussion

To evaluate the acceptance of a dental implant education system based on Mixed Reality (MR) technology among medical students, particularly focusing on operations conducted through Tangible Interaction (TI), we conducted a survey study. This research examined the roles of tangible interaction, perceived usefulness, perceived ease of use, and attitudes toward usage in predicting the intention to use these systems during the learning process. The following sections discuss each factor that contributes to predicting the acceptance of the MR-based dental implant education system.

5.1 Tangible Interaction

Our study found that the tangible interaction teaching method is more popular among medical students compared to traditional teaching methods. This is primarily because tangible interaction significantly enhances students' practical skills and surgical operation abilities [31, 32], allowing them to practice complex dental implant procedures in a safe and controlled environment, which boosts their confidence and reduces potential errors when operating on real patients in the future.

Although our developed tangible interaction system is relatively basic and focuses on fundamental dental implant techniques, it provides students with an unprecedented immersive experience. However, since some medical students are accustomed to using more diverse and complex traditional educational materials (such as textbooks and video lectures), their excitement and acceptance of the new MR-based system did not meet expectations.

Therefore, in designing future MR-based tangible interaction educational platforms, it is essential to fully consider the needs of medical students at different stages, offering customized content and personalized learning objectives. Additionally, to reduce learning difficulties, the system should feature high visual clarity and intuitive operation guides to help students quickly master new technologies, thereby effectively improving professional proficiency. By continuously optimizing these aspects, we can anticipate the development of next-generation dental education solutions that are both efficient and widely accepted.

5.2 Perceived Usefulness

Perceived Usefulness (PU) has a positive impact on Behavioral Intention (BI), but this effect is mediated by Attitude Toward Usage (ATU). When medical students recognize that Mixed Reality (MR) technology can effectively enhance dental implant procedures, they are more likely to view its use as a good idea, thereby stimulating their intention to use it. The study specifically focused on the effectiveness, performance, and efficiency of MR-based dental education systems to improve their practicality and technology acceptance.Firstly, enhancing teaching content can improve the effectiveness of the MR system [33, 34]. For example, adhering to the latest dental implant guidelines and incorporating expert opinions ensures that training aligns with clinical needs. Introducing case studies and real-case simulations increases the authenticity and applicability of learning, helping students better prepare for future challenges.Secondly, providing clear operational guidance and real-time feedback mechanisms can enhance system performance. Integrating sensors and algorithms to monitor the accuracy of student operations and offering immediate visual or auditory corrections helps improve skill levels and boost confidence [35, 36]. Thirdly, improving the user interface and interaction design can increase learning efficiency. Simplifying setup and calibration processes optimizes the user experience, allowing students to get started quickly and focus on actual learning. An intuitive graphical user interface (GUI), video tutorials, and detailed manuals help beginners familiarize themselves with the system rapidly. Automated calibration tools and troubleshooting guides reduce initial setup time [37].

5.3 Perceived Ease of Use

In the current study, Perceived Ease of Use (PEOU) has a positive impact on Perceived Usefulness (PU) and Attitude Toward Usage (ATU). Individuals seem to find the technology useful and develop a positive attitude toward it, regardless of their technical skills or prior experience. However, it should be noted that the sample in this study primarily consisted of medical students, who may possess certain foundational skills and openness in adopting new technologies. Therefore, these findings may not fully generalize to all user groups, especially those less familiar with or resistant to new technologies.

To ensure that MR-based dental education systems are widely accepted, future research should consider a broader range of users, including medical professionals from diverse backgrounds and varying levels of technical proficiency. Additionally, further exploration of other potential factors, such as the availability of training and support resources, and how these factors influence users' acceptance of MR technology, is needed. By integrating these factors, MR systems can be better designed and optimized to be useful not only for medical students but also for a wider range of healthcare practitioners.

5.4 Attitude Toward Using

We found that when medical students have a positive Attitude Toward Using (ATU) Mixed Reality (MR) technology for dental implant procedures, they are more likely to express an intention to use this technology. Research indicates that their experience with the technology is crucial in fostering such a positive attitude. Therefore, educational institutions and relevant stakeholders should consider providing more opportunities for medical students to familiarize themselves or experience this technology. For instance, before the formal curriculum begins, demonstrations and presentations, expert explanations, short-term trial opportunities, sharing of success stories, and interactive discussion sessions can be used to showcase the advantages of MR-based technologies [38, 39]. These measures not only enhance medical students' awareness and interest in MR technology but also help them develop a positive attitude, thereby increasing their willingness and enthusiasm to use MR technology in their actual learning and practice.

6 Conclusion

In this study, we developed an immersive Mixed Reality (MR) teaching system for dental implant education and explored the role of Tangible Interaction and the Technology Acceptance Model (TAM) in promoting students' acceptance of this technology. The research shows that integrating the physical world with virtual information can significantly enhance students' understanding and mastery of complex surgical procedures while increasing their affinity for and willingness to use the technology. Future work can leverage these findings to improve the design, development, and implementation of MR-based educational systems. Educators and technology developers should focus on enhancing learning experiences through enriched tangible interactions and promote MR solutions with advanced interactive features to help professionals more effectively acquire high-demand complex skills such as dental surgery.

References

1. Vivekanantham, S., Ravindran, R.P.: Technology: changing the focus of medical education? Adv. Med. Educ. Pract. **2014**, 25–26 (2014)
2. Patel, V.L.: Recent advances in computer technologies and medical education. Yearb. Med. Inform. **5**(01), 521–524 (1996)
3. Hosseini, M., et al.: Dental student's satisfaction with the video-assisted educational approach in teaching oral and maxillofacial surgery principles. Open Access Macedonian J. Med. Sci. **10**(D), 435–440 (2022)
4. Dudley, J.: The use of artificial teeth for post-core techniques in a pre-clinical fixed prosthodontics undergraduate teaching program: an evaluation of student experience. J. Int. Oral Health **13**(2), 144–150 (2021)
5. Paladino, J.R., et al.: The benefits of expert instruction in microsurgery courses. J. Reconstr. Microsurg. **37**(02), 143–153 (2021)
6. Liu, L., et al.: Simulation training for ceramic crown preparation in the dental setting using a virtual educational system. Eur. J. Dent. Educ. **24**(2), 199–206 (2020)
7. Yoshimoto, R., Sasakura, M.: Using real objects for interaction in virtual reality. In: Proceedings of the 21st International Conference Information Visualisation (IV), pp. 1–8. IEEE (2017)
8. Grivokostopoulou, F., et al.: Utilizing virtual reality to assist students in learning physics. In: 2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 1–5. IEEE (2017)
9. Zhang, L., et al.: The added value of virtual reality technology and force feedback for surgical training simulators. Work **41**(Supplement 1), 2288–2292 (2012)
10. Chahal, B., Aydin, A., Ahmed, K.: Virtual reality vs. physical models in surgical skills training. An update of the evidence. Curr. Opin. Urol. **34**(1), 32–36 (2024)

11. Shafarenko, M.S., et al.: The role of augmented reality in the next phase of surgical education. Plast. Reconstruct. Surgery-Global Open **10**(11), e4656 (2022)
12. Abarghouie, M.H.G., Omid, A., Ghadami, A.: Effects of virtual and lecture-based instruction on learning, content retention, and satisfaction from these instruction methods among surgical technology students: a comparative study. J. Educ. Health Promo. **9**(1), 296 (2020)
13. Espejo-Trung, L.C., Elian, S.N., Luz, M.A.A.D.C.: Development and application of a new learning object for teaching operative dentistry using augmented reality. J. Dent. Educ. **79**(11), 1356–1362 (2015)
14. Speicher, M., Hall, B.D., Nebeling, M.: What is mixed reality?. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Article No. 1, pp. 1–12. ACM, New York (2019)
15. Cooper, E.A.: The perceptual science of augmented reality. Ann. Rev. Vis. Sci. **9**, 455–478 (2023)
16. Burdea, G.C., Coiffet, P.: Virtual Reality Technology, 2nd edn. Wiley, Hoboken (2024)
17. Mahrous, A., et al.: A comparison of pre-clinical instructional technologies: natural teeth, 3D models, 3D printing, and augmented reality. J. Dent. Educ. **85**(11), 1795–1801 (2021)
18. Li, Y., et al.: The current situation and future prospects of simulators in dental education. J. Med. Internet Res. **23**(4), e23635 (2021)
19. LeBlanc, V.R., et al.: A preliminary study in using virtual reality to train dental students. J. Dent. Educ. **68**(3), 378–383 (2004)
20. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
21. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: a comparison of two theoretical models. Manage. Sci. **35**(8), 982–1003 (1989)
22. Badías, A., et al.: Real-time interaction of virtual and physical objects in mixed reality applications. Int. J. Numer. Meth. Eng. **121**(17), 3849–3868 (2020)
23. Anagnostopoulos, A., Pnevmatikakis, A.: A realtime mixed reality system for seamless interaction between real and virtual objects. In: Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts, pp. 1–7. ACM, New York (2008)
24. Dinh, H.Q., et al.: 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. 1–7. IEEE (1999)
25. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**(3), 425–478 (2003)
26. Malatji, W.R., Van Eck, R., Zuva, T.: Understanding the usage, modifications, limitations and criticisms of technology acceptance model (TAM). Adv. Sci. Technol. Eng. Syst. J. **5**(6), 113–117 (2020)
27. Raditya, A., Dewi, A.C., As'adi, M.: An empirical study to validate the Technology Acceptance Model (TAM) in evaluating "Desa Digital" applications. In: IOP Conference Series: Materials Science and Engineering, vol. 1125. No. 1. IOP Publishing, Bristol (2021)
28. Cao, X.: The application of structural equation model in psychological research. CNS Spectr. **28**(S1), S17–S19 (2023)
29. Stephenson, M.T., Holbert, R.L., Zimmerman, R.S.: On the use of structural equation modeling in health communication research. Health Commun. **20**(2), 159–167 (2006)
30. Rahman, W., Shah, F.A., Rasli, A.: Use of structural equation modeling in social science research. Asian Soc. Sci. **11**(4), 371–380 (2015)
31. Towers, A., et al.: Combining virtual reality and 3D-printed models to simulate patient-specific dental operative procedures—a study exploring student perceptions. Eur. J. Dent. Educ. **26**(2), 393–403 (2022)

32. Wu, J.: Remote oral implant teaching system based on VR virtual simulation technology. In: 2023 2nd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS), pp. 1–5. IEEE (2023)
33. Mowla, M.M.: Morning report: a tool for improving medical education. J. Bangladesh Coll. Phys. Surg. **30**(2), 71–73 (2012)
34. McGaghie, W.C., et al.: Medical education featuring mastery learning with deliberate practice can lead to better health for individuals and populations. Acad. Med. **86**(11), e8–e9 (2011)
35. Ando, A., et al.: Development of a skill learning system using sensors in a smart Phone for vocational education. In: International Conference on Computer Supported Education, vol. 2, pp. 1–6. SCITEPRESS (2013)
36. Dhinesh, R., Preejith, S.P., Sivaprakasam, M.: Understanding the effects of a real-time motion capture system with voice feedback for tennis toss training. In: 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–6. IEEE (2023)
37. UNIX System Laboratories: OPEN LOOK Graphical User Interface: User's Guide. Prentice Hall, Upper Saddle River (1991)
38. Fischetti, J., et al.: Practice before practicum: simulation in initial teacher education. Teach. Educ. Q. **57**(2), 155–174 (2022)
39. Sinha, U., Sinha, S.: WE-G-BRC-06: magnetic resonance imaging: enhancing concepts through desktop models and supplementing clinical rotations with video presentations. Med. Phys. **38**(6Part33), 3830–3830 (2011)

Multimodal Interaction in Virtual Environments

The Intelligent Car Seat Adjustment System Based on a Multimodal Driving Fatigue Detection Method

Yunpeng Bai[1,2,3], Min Zhao[1,2,3], Wanming Zhong[1,2,3], Wenzhe Cun[2], Yuanjun Li[2], Mengya Zhu[4], Chenjie Zhao[5], Bingjun Liu[6], Yuan Feng[1,3](✉), and Dengkai Chen[1,2,3](✉)

[1] School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China
y.feng@nwpu.edu.cn
[2] Ningbo Institute of Northwestern Polytechnical University, Ningbo 315103, China
[3] Key Laboratory of Industrial Design and Ergonomics Ministry of Industry and Information Technology (Northwestern Polytechnical University), Xi'an 710072, China
[4] Ningbo University, Ningbo 315211, China
[5] Hebei Expressway Ecological Construction Co., LTD, Shijiazhuang 050035, China
[6] North Information Control Research Institute Group Co. LTD, No.528 General Avenue, Nanjing 210000, China

Abstract. Fatigue driving is a crucial factor in causing traffic accidents. To enhance the accuracy and reliability of determining the driver's fatigue state, automatically adjust the seat based on the judgment result, and further stimulate the driver's fatigue state to achieve the goal of safe driving, this paper focuses on the extraction of drivers' facial and physiological characteristic data and the construction of a multimodal fusion model. Firstly, it deeply analyzes the basic theories related to the face, heart rate, and electroencephalogram (EEG), elaborates on the extraction methods of various features and their associations with the fatigue state, and introduces the applicable recognition methods. A driving simulation platform is utilized to conduct fatigue driving experiments, collect facial video, heart rate signal, and EEG signal data, and construct a fatigue driving dataset. Subsequently, a multimodal fatigue state recognition model based on BCL-SVM is proposed. The facial, heart rate, and EEG features are respectively input into Back Propagation Neural Network (BP), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) networks for preliminary prediction, then the decision fusion is carried out through the Support Vector Machine (SVM) to determine the driver's fatigue state. Finally, based on the determined result, a seat adaptive adjustment method model is proposed, providing ideas for alleviating driver fatigue and improving driving safety.

Keywords: fatigue driving · facial features · physiological features · multimodal fusion · adaptive adjustment

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 289–305, 2025.
https://doi.org/10.1007/978-3-031-93715-6_18

1 Introduction

Fatigue driving has become a key factor in causing traffic accidents, resulting in huge losses to society and families. Therefore, effectively detecting driver fatigue and taking appropriate measures are of crucial importance for preventing traffic accidents and ensuring road safety [17].

In the field of fatigue driving detection, scholars have proposed various detection methods. Sha et al. [1, 16]. Extracted grip signal features closely related to fatigue from the collected steering wheel grip force data and combined them with a BP to establish a fatigue detection model, with an accuracy rate of 87%. However, such methods are greatly affected by factors such as driving style, vehicle conditions, and road conditions, resulting in poor accuracy and stability. Zhu et al. [2]. Proposed a fatigue detection method based on electroencephalogram (EEG) signals. By extracting the energy values of drivers in the δ, θ, α, and β bands and comparing the fatigue characteristics of different time periods, they found that the degree of driver fatigue was positively correlated with the fatigue characteristic values. Wang et al. [3]. Used the electrode-frequency distribution map of EEG signals to construct an emotion recognition model based on a deep convolutional neural network, achieving an accuracy rate of 90.59% in identifying the driver's fatigue state. However, these methods require drivers to wear professional equipment, which may cause discomfort to the drivers and have high costs, limiting their practical applications. Ma et al. [4]. Proposed an algorithm for yawning detection using a CNN. They input facial images into the network and combined them with a Softmax classifier to determine whether the driver was yawning, thus identifying fatigue driving. Zhu et al. [5]. Proposed an algorithm based on a boosting tree to detect facial key points and extract features such as PERCLOS, the longest continuous eye closure time, and the number of yawns. The experimental results showed that the accuracy rate of this model was as high as 92.5%. However, facial feature detection methods are easily affected by environmental factors such as light and occlusion, resulting in unstable detection accuracy. Multimodal feature fusion detection methods combine various types of feature information such as facial features, EEG features, electrooculogram signals, and electrocardiogram features to comprehensively analyze and judge the driver's fatigue state, aiming to improve the accuracy and reliability of the fatigue driving detection system. Cao et al. [6]. Proposed a fatigue driving monitoring system based on the fusion of electrooculogram signals and image information. By combining the features of eye images with the electrooculogram fatigue feature monitoring method with less human intervention, the monitoring accuracy was improved. Wang et al. [7]. Took the breathing signals, eye movement signals, and steering wheel signals of normal drivers as the research objects, collected and separated the features of fatigue detection signals. The experimental results showed that the accuracy rate of the non-uniform signal fusion method in fatigue detection was as high as 80%. Although the multimodal feature fusion method has achieved certain results, its adaptability and robustness in specific environments need to be further improved. In addition, the adjustment functions of traditional car seats are relatively basic, mostly manual operations, and the adjustment range is limited, unable to perform personalized adjustments according to the real-time state of the driver. During long hours of driving, drivers are prone to fatigue, and traditional seats are difficult to provide effective fatigue relief functions.

This research focuses on solving the existing problems in fatigue driving detection. It mainly carries out the work of extracting the facial and physiological characteristic data of drivers and constructs a multimodal fusion model [15]. The facial, heart rate and EEG features are respectively input into Back Propagation Neural Network (BP), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) networks for preliminary prediction, and then the fatigue state is judged through the decision fusion of Support Vector Machines(SVM). In addition, an innovative seat adaptive adjustment strategy is formulated. When the system detects fatigue, the intelligent seat automatically provides auditory and olfactory stimuli to relieve the driver's fatigue and improve driving safety. This research is expected to open up new directions for related fields and is of great significance for reducing the traffic accident rate and ensuring road safety.

2 Related Work

2.1 Multimodal Feature Extraction Methods for Drivers' Fatigue Detection

1. Fatigue Detection Using Facial Features

Eye Fatigue Feature Extraction. The opening and closing state of the eyes is judged by calculating the Eye Aspect Ratio (EAR), and the formula is shown in (1) [8]. When a person opens their eyes, the width-to-length ratio is fixed, while when they close their eyes, the length remains unchanged but the width narrows rapidly, and the aspect ratio changes. The eye feature points are shown in Fig. (1), and P1 to P6 are six feature points. When the eyes are open, the eye openness is large, the denominator remains unchanged, and the numerator increases, resulting in an increase in the EAR value; when the eyes are closed, the eye openness decreases, the denominator remains unchanged, and the numerator becomes smaller, resulting in a decrease in the EAR value. Thus, the opening and closing state of the driver's eyes can be detected, and the eye fatigue state can be judged based on the eye closing frequency.

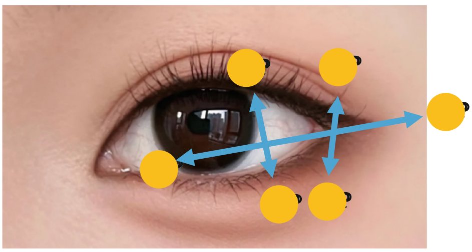


Fig. 1. Eye Feature Points.

$$EAR = \frac{\|P_3 - P_4\| + \|P_5 - P_6\|}{2\,\|P_1 - P_2\|} \tag{1}$$

(a) PERCLOS Criterion.

PERCLOS refers to the percentage of the time when the eyes are closed within a unit time over the total time, and the calculation formula is shown as follows (2) [9].

$$f = \frac{t_3 - t_2}{t_4 - t_1} \times 100\% \tag{2}$$

Among them, t3 – t2 represents the duration of eye closure, and t4 – t1 represents the duration of eye opening. When the driver is fatigued, the duration of eye closure will increase. The proportion of eye closure duration is often used as a parameter for the eye fatigue state, that is, the proportion of the duration when the eyes are closed exceeding a certain threshold within a specific time period, and P80 is often used as the criterion for judgment.

The PERCLOS value can also be represented by the proportion of the number of frames of eye fatigue within a certain period of time to the total number of video frames, and its calculation formula is shown as follows (3).

$$f = \frac{M}{N} \times 100\% \tag{3}$$

In formula (3), M represents the sum of the number of frames with closed eyes within a specific time, N represents the sum of all frames within a specific time, and f represents the PERCLOS value.

(b) Duration of Continuous Eye Closure.

When the driver's fatigue level increases, the duration of continuous eye closure becomes longer. Therefore, this paper uses the maximum value of the duration of eye closure during the awake state as a threshold to determine whether the driver is fatigued.

(c) Blink Frequency.

Usually, the blink frequency increases when the human body is fatigued. To judge the fatigue state, the blink frequency during the awake state can be used as a reference threshold. If the detected frequency is lower than the threshold, the driver may be awake; otherwise, the driver is in a fatigue state.

2. Mouth Fatigue Feature Extraction

The same as the method of extracting eye fatigue features, the mouth aspect ratio (MAR) is calculated to judge the opening and closing of the mouth [10]. Under normal circumstances, the driver's mouth is closed or slightly open. If they feel tired, they may yawn, and by counting the frequency of yawning over a period of time, it can be judged whether they are fatigued.

Physical Fatigue Detection Using Heart Rate Signals. In practical applications, a heart rate monitoring device (such as a heart rate belt) can be used to obtain signals, and relevant software or programming languages can be used to extract heart rate features. This study extracts driving fatigue features from heart rate data, including average heart rate (AVGHR), standard deviation of heart rate (SDNN), and peak heart rate (MAX). The specific mathematical expressions are as follows:

$$AVGHR = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

Formula (4) represents the average heart rate (AVGHR), where xi represents the *i* heart rate sample value and *N* represents the total value of the heart rate samples.

$$SDNN = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{5}$$

Formula (5) represents the standard deviation of heart rate (SDNN), where *xi* represents the *i* heart rate sample value, *N* represents the total number of heart rate sample values, and $\overline{x}$ represents the average value of heart rate sample values.

$$MAX = \max_{i=1}^{N} x_i \tag{6}$$

Formula (6) represents the peak heart rate (MAX), where *xi* represents the *i* heart rate sample value and *N* represents the total value of the heart rate samples.

EEG Fatigue Feature Extraction. The extraction of EEG fatigue features is a process of analyzing EEG signals to identify and quantify driving fatigue, and the commonly used methods include time domain analysis and frequency domain analysis [11].

Assume that $x(n)$ is a random discrete signal, its autocorrelation function is represented as $r(k)$, and its power spectral density function is:

$$P(w) = \sum_{k=-\infty}^{\infty} r(k)e^{-jwk} \tag{7}$$

In formula (7), $r(k) = E[x(n)^*(n + k)]$ represents the autocorrelation function, where E stands for mathematical expectation and * represents conjugation. When its autocorrelation function satisfies the following conditions:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{K=-\infty}^{+\infty}|k\|r(k)| = 0 \tag{8}$$

Formula (8) can also be expressed as:

$$P(w) = limE\left[\frac{1}{N}\left|\sum_{N=1}^{N\to\infty} x(n)e^{-jwn}\right|^2\right] \tag{9}$$

The power spectral density of the electroencephalogram signals in the four rhythmic waves of θ (4~8 Hz), α (8~13 Hz), β (13~30 Hz), and γ (30~45 Hz) is shown as follows.

$$\begin{bmatrix} E_\theta = \sum P(w), & 4 < w \leq 8 \\ E_\alpha = \sum P(w), & 8 < w \leq 13 \\ E_\beta = \sum P(w), & 13 < w \leq 30 \\ E_\gamma = \sum P(w), & 30 < w <= 45 \end{bmatrix} \tag{10}$$

Because electroencephalogram (EEG) signals have randomness and instability, relying solely on the energy intensity in a specific frequency band is difficult to accurately describe the brain state, nor can it cope with the uncertainty brought about by individual differences. This study uses the power spectral density ratio between bands as the EEG signal feature, and the method for calculating the power spectral density ratio between different frequency bands is as follows:

$$R_{(\theta/\beta)} = \frac{E_\theta}{E_\beta} \tag{11}$$

In formula (11), $R(\theta/\beta)$ represents the power spectral density ratio between the θ wave band and the β wave band.

After extracting the power spectral density ratio of wave frequency bands, this paper selects R (α/β), R (θ/β), and R ($\alpha + \theta/\beta$) [12] as electroencephalogram fatigue indicators to analyze the changes in the driver's fatigue state.

2.2 Preliminary Construction of a Multimodal Fatigue Driving Detection Model

The proposed BCL-SVM multimodal fatigue recognition model, the core of which is to conduct preliminary testing on different modal features using appropriate networks and then perform decision fusion using SVM. This model first inputs facial, heart rate, and EEG features into the BP, CNN, and LSTM networks respectively. The output values of the three networks represent the driver's fatigue level (0 represents awake and 1 represents fatigue), which are used as the input of the SVM. The SVM finds a hyperplane in the high-dimensional space and maximizes the geometric margin to complete the final judgment of the driver's fatigue state.

3 Method

3.1 Facial Feature Fatigue Recognition Method

When identifying driver fatigue through facial features, features such as facial expressions, eye and mouth states show complex non-linear relationships, and the BP neural network can effectively capture their associations.

The BP neural network includes input, hidden, and output layers, and each neuron is fully connected to the neurons in the next layer. The network structure is shown in Fig. 2. The vector $X = \left(x_1, x_2, ..., x_i, ..., x_m\right)$ of the input layer represents m input features of the problem, while the vector $H^l = \left(h_1^l, h_2^l, ..., h_j^l, ..., h_{s_l}^l\right)$ of the hidden layer represents the features of the L layer. The representation of the hidden layer usually depends on the weighted sum and bias of the input data as well as the non-linear transformation of the activation function. By adjusting the structure and parameters of the network, the representational ability of the hidden layer can be affected, enabling the neural network to learn the features and patterns of the data better and generate n output layer vectors $Y = \left(y_1, y_2, ..., y_k, ..., y_n\right)$, which represent the results of the model [13].

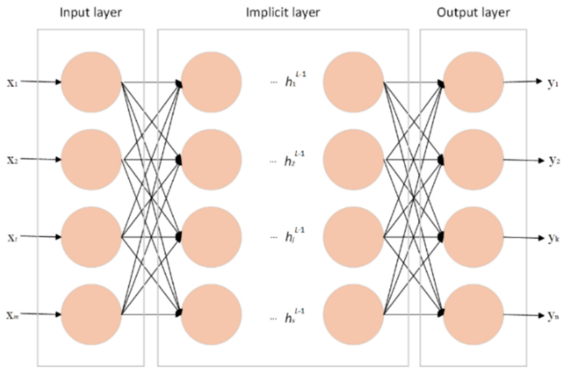


Fig. 2. BP Neural Network Structure Diagram.

According to the above principles, when identifying the driver's fatigue state through facial features, for the three input facial features represented by $\left(x_{perclos}, x_{blink}, x_{\text{yawn}}\right)$ which stand for PERCLOS, blink frequency, and yawn frequency respectively, the output value of the BP neural network prediction model, represented by f_{face}, indicates the driver's fatigue level and can be described as:

$$f_{face} = net_{BP}\left(x_{perclos}, x_{blink}, x_{yanvn}\right) \tag{12}$$

3.2 Heart Rate Feature Fatigue Recognition Method

When identifying the driver's fatigue state through heart rate features, features such as average heart rate, heart rate peak, and heart rate standard deviation show non-linear relationships, and their importance varies at different time periods or frequencies. The CNN can effectively capture local features through convolution and pooling operations.

When recognizing the heart rate features of a driver in a fatigue state through a CNN, for the three input heart rate features, $\left(x_{HR\_average}, x_{HR\_sdnn}, x_{HR\_max}\right)$ are used to represent the average heart rate, heart rate standard deviation, and heart rate peak, and the output value of the convolutional neural network prediction model, represented by fHR, indicates the driver's fatigue level and can be described as:

$$f_{HR} = net_{CNN}\left(x_{HR\_average}, x_{HR\_sdnn}, x_{HR\_max}\right) \tag{13}$$

3.3 EEG Fatigue Detection Method

EEG belongs to time series data, which has temporal and correlative characteristics. The analysis and modeling of sequence data should consider time dependence and temporal relationships in order to complete tasks such as prediction, pattern recognition, or classification. LSTM, as a special type of recurrent neural network, can handle long time series data more effectively when recognizing the fatigue state of drivers. LSTM has memory units and can capture long-term dependencies in time series. The LSTM unit consists of three gating units and one memory unit, including the forget gate, input gate, and output gate, which are used to control information flow and memory operations [14].

When recognizing the fatigue state of drivers based on electroencephalogram features through the Long Short-Term Memory network, for the three input electroencephalogram features, $\left(\mathcal{X}_{R(\alpha/\beta)}, \mathcal{X}_{R(\theta/\beta)}, \mathcal{X}_{R(\alpha+\theta)/\beta}\right)$ are used to represent the ratios of power spectral density R(α/β), R(θ/β), and R(α + θ/β). The output value of the Long Short-Term Memory network prediction model, represented by f EEG, indicates the driver's fatigue level and can be described as:

$$f_{EEG} = net_{LSTM}\left(x_{R(\alpha/\beta)}, x_{R(\theta/\beta)}, x_{(\alpha+\theta)/\beta)}\right) \tag{14}$$

3.4 Multi-modal Fatigue State Detection Method

First, facial features such as PERCLOS, blink frequency, and yawn frequency are extracted from the facial data collected by the camera and then input into the BP neural network prediction model to output the driver's fatigue state. Secondly, heart rate features such as average heart rate, heart rate standard deviation, and heart rate peak are extracted from the heart rate data collected by the heart rate belt, and they are fed into the CNN network prediction model to output the driver's fatigue state. Next, electroencephalogram features $R(\alpha/\beta)$, $R(\theta/\beta)$, and $R(\alpha + \theta/\beta)$ are extracted from the electroencephalogram data collected by the electroencephalograph and input into the LSTN network prediction model to output the driver's fatigue state. These features are respectively input into their respective models for training, and finally, the output values of the three models are imported into the SVM model for decision fusion, and the final prediction result is output to determine whether the driver is in a state of fatigue. The model structure is shown in Fig. 3 below.

The expression of the multi-feature fusion method based on BCL-SVM is:

$$y_{face} = net_{BP}\left(x_{perclos}, x_{blink}, x_{yawn}\right) \tag{15}$$

In formula (15), xperclos, xblink, xyawn are the values of PERCLOS, blink frequency, and yawn frequency, respectively, and yface is the driver's fatigue level output by the BP neural network, with a value of 0 or 1.

$$y_{HR} = net_{CNN}\left(x_{HR\_average}, x_{HR\_sdnn}, x_{HR\_max}\right) \tag{16}$$

In formula (16), $x_{\mathrm{HR\_average}}$, $x_{\mathrm{HR\_sdnn}}$, and $x_{\mathrm{HR\_max}}$ are the average heart rate, heart rate standard deviation, and heart rate peak, respectively, and y_{HR} is the driver's fatigue level output by the CNN network, with a value of 0 or 1.

$$y_{EEG} = net_{LSTM}\left(x_{R(\alpha/\beta)}, x_{R(\theta/\beta)}, x_{(\alpha+\theta)/\beta)}\right) \tag{17}$$

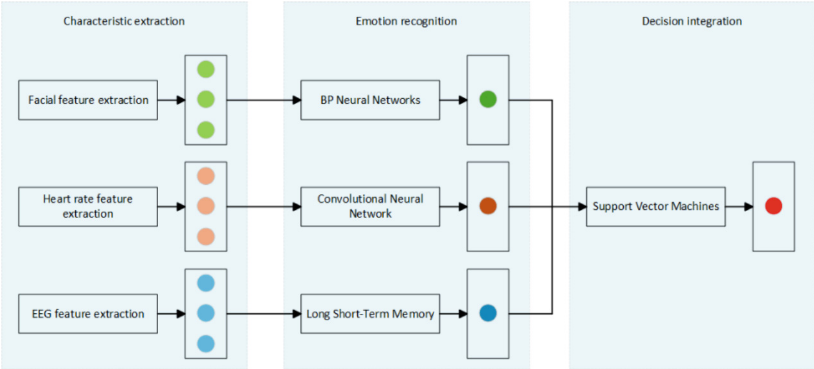


Fig. 3. Model Structure Diagram.

In formula (17), $x_{R(\alpha/\beta)}$, $x_{R(\theta/\beta)}$, and $x_{R(\alpha+\theta/\beta)}$ are the ratios of electroencephalogram power spectral density $R_{(\alpha/\beta)}$, $R_{(\theta/\beta)}$, and $R_{(\alpha+\theta/\beta)}$, and y_{EEG} is the driver's fatigue level output by the LSTM network, with a value of 0 or 1.

$$y_{reasult} = SVM\left(y_{face}, y_{HR}, y_{EEG}\right) \tag{18}$$

In formula (18), $y_{reasult}$ represents the output values of expressions (15), (16), and (17). After decision fusion in the SVM model, the predicted value of the driver's fatigue level is output, with the value being either 0 or 1.

4 Experiments

4.1 Ethical Statement and Participant Recruitment

This research was approved by the Ethics Review Board of Ningbo Institute of Northwestern Polytechnical University. We gained informed consent from all participants. And the data of all participants are stored and processed in accordance with Ningbo Institute of Northwestern Polytechnical University. The data of all participants are processed in accordance with the Declaration of Helsinki. Participants are informed that the driving simulator may cause discomfort such as dizziness, and they are granted the right to pause or withdraw from the experiment at any time. Only after each participant provides a written consent can their experimental data be used for academic research, and then the experiment can officially proceed.

In this study, a total of 24 eligible participants were recruited for the experiment. All participants held valid driving licenses and had no physical or mental illnesses. Within 24 h before the start of the experiment, they refrained from consuming beverages containing stimulating substances such as alcohol or caffeine. To ensure a more diverse representation of the driving population and reflect the gender ratio of Chinese drivers, the participants were divided into two groups: 10 students (5 males and 5 females, with an average age of 27 years and an average driving experience of 3 years) and 14 working adults (7 males and 7 females, with an average age of 41 years and an average driving age of 12 years). After completing the experiment, all participants received small gifts for their participation.

4.2 Settings of the Simulated Driving Environment

The driving simulation platform used in this study is an automatic-transmission vehicle with six degrees of freedom. It has a complete operating system, including a steering wheel, an accelerator pedal, a brake pedal, a handbrake, a real-life dashboard interface, and turn signals, etc. This driving simulation platform can achieve forward-backward tilting and up-down vibration during the driving process, providing drivers with the most realistic driving experience possible.

We simulated real-life road scenes and traffic conditions and designed multiple virtual road scenarios for simulated driving. The specific road scenarios are shown in Fig. 4. These designs are based on Chinese urban road design standards, including the "Code for Design of Urban Road Engineering (2016 Edition)" and the "Code for Planning of Urban Road Intersections (2012 Edition)". The simulated driving road is a two-way single-lane road, with each lane being 3.75 m wide. The road infrastructure, surrounding buildings, trees, and billboards are all constructed by referring to real urban road sections. The settings of road intersections, traffic lights, and traffic flow take into account road congestion. The traffic flow includes various vehicle types such as cars, trucks, and buses, without specifying specific vehicle brands. Each driver can adjust the seat and other equipment according to personal preferences to maintain a comfortable driving experience and a wide field of vision.

Fig. 4. Urban Virtual Road.

4.3 Data Collection

In this fatigue-driving experiment, the fatigue level of drivers is determined by analyzing their facial and physiological characteristics. The experiment uses a Logitech HD webcam C922PRO, an EMOTIV Epoc+electroencephalogram (EEG) device, and a Polar H10 heart-rate monitor to collect the drivers' facial images and physiological signals respectively. To ensure the accuracy of the data, the subjects are required to wear the heart-rate monitor closely against the skin and fix it. Before the experiment, the 14 channels of the EEG device are adjusted to show all-green on the Contact Quality Map, ensuring a 100% connection quality.

The study assesses the mental state and fatigue level of the subjects with the help of a subjective fatigue questionnaire and the Karolinska Sleepiness Scale (KSS). The

KSS has a rating scale from 1 to 9. A score of 1 represents being extremely alert, while a score of 9 represents being extremely fatigued, accompanied by symptoms such as drowsiness, tearing, and eye-stinging. The higher the score, the more severe the fatigue of the subjects. The aim is to verify whether the subjective fatigue level of the subjects changes before and after the driving task.

4.4 Experimental Procedure

Each participant is required to complete two rounds of simulated experiments, one in a normal state and the other in a fatigued state. There is a one-week interval between the two rounds. The experimental data of subjects who are unable to complete the two-round controlled experiments will not be retained. Based on the general biological rhythm, the experiment is scheduled to collect data of the subjects in a sober state from 8:00 am to 12:00 pm and data in a fatigued state from 14:00 pm to 18:00 pm. To reduce subjective differences, the Karolinska Sleepiness Scale (KSS) is adopted to determine the degree of driving fatigue. Before the experiment, the staff will assist the subjects in wearing and debugging the equipment and allow them sufficient time to get familiar with the simulated driving platform.

3. Experiment in Normal State

Subjects are required to ensure sufficient and high-quality sleep the night before the experiment and avoid consuming foods or taking medications that can stimulate the brain nerves. The experiment will be conducted from 8:00 am to 11:00 am, lasting for two hours. To reduce the interference of subjective factors, data recording will start 5 min after the subjects start driving. Due to the long duration of the experiment, according to the subjective fatigue questionnaire, the subjects may experience fatigue halfway through the experiment, so experimental data in the fatigued state can also be collected during this stage.

4. Experimental Procedure during a Fatigue State

Subjects are required to have insufficient sleep the night before the experiment and are not allowed to take additional breaks on the day of the experiment. The experiment will be carried out from 14:00 to 18:00 in the afternoon, which is the most tiring period for the human body, and it will last for two hours. To reduce the influence of subjective factors, data recording will start 5 min after the subjects start driving. Due to the complexity of the experimental environment, referring to the subjective fatigue questionnaire, the subjects may be in a sober state in the first half of the experiment, so experimental data in the sober state can also be collected during this stage.

4.5 Data Analysis

Data Analysis for Collected Facial Data. This study focuses on the analysis of facial features in fatigue driving. Facial video data of 24 drivers in sober and fatigued states were collected. After screening out abnormal segments, with a sampling rate of 25 frames per second and a time window of 60 s as the standard, 36,000 valid samples were obtained.

During data processing, the facial fatigue feature extraction method was used to calculate the Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) of the eyes and mouth, and determine their thresholds. The face was detected with the help of Python, OpenCV, and dlib libraries. The coordinates of 68 facial key points were obtained and marked. Twelve position information of the eyes and twelve of the mouth were input into the EAR and MAR functions respectively, and compared with the preset thresholds to determine blinking and yawning actions.

Figure 5 presents the EAR and MAR data of one driver in different states. The first 150 frames represent the sober state, and the subsequent 250 frames represent the fatigued state. To prevent misjudgment, an EAR of 0.11 and a MAR of 0.6 were set as the thresholds. When the EAR is less than 0.11, it indicates that the eyes are closed, and when the MAR is greater than 0.6, it means the mouth is open.

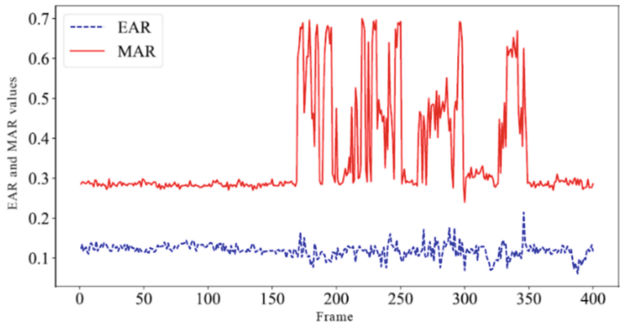


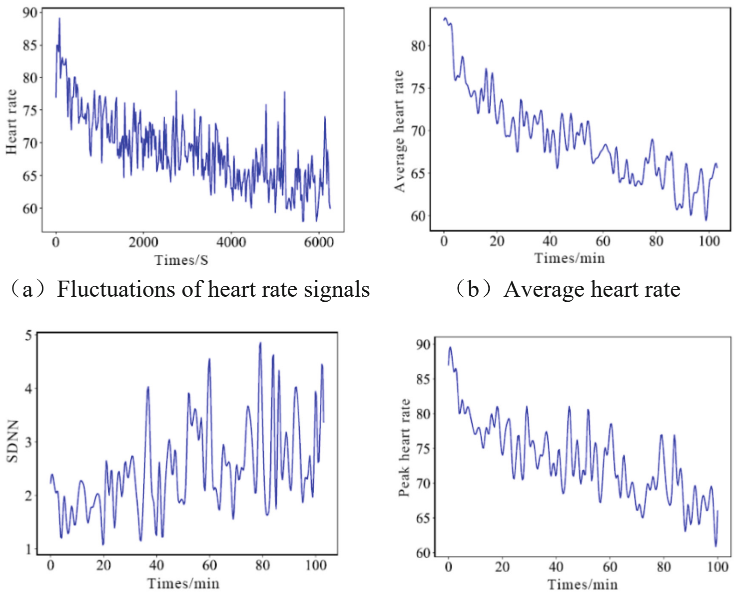
Fig. 5. Analysis Results of EAR and MAR.

Blinking and yawning are continuous actions. When the EAR of the current frame is greater than 0.11 and the EAR of the next frame is less than or equal to 0.11, it is recorded as one blink. When the MAR of the current frame is less than 0.6 and the MAR of the next frame is greater than or equal to 0.6, it is recorded as one yawn. Based on the thresholds of EAR and MAR, the Percentage of Eyelid Closure over Time (PERCLOS), the number of blinks, and the number of yawns for each video sample are calculated. Table 1 shows the calculation results of some samples.

Processing and Analysis of Heart Rate Data. In the simulated driving experiment, the heart rate data of 24 drivers in the awake and fatigued states were collected. After screening out the abnormal data, 4086 fatigue samples were obtained with a 60 – second time window as the standard. The average heart rate, heart rate standard deviation, and heart rate peak were used as indicators to analyze the changes in the driver's fatigue state. The data from the Polar H10 heart rate strap was exported as a CSV file and imported into Pycharm to generate a heart rate signal graph. As shown in Fig. 6, in the fatigue task, with the increase of driving time, the average heart rate, heart rate standard deviation, and heart rate peak fluctuate greatly. The average heart rate and the peak value slowly decline, and the heart rate standard deviation slowly rises.

Table 1. Calculation Results of Facial Features for Some Samples.

Sample Serial Number	PERCLOS	Number of blinks	Number of yawns	Whether fatigued or not
1	0.0053	8	0	No
2	0.0226	32	2	Yes
3	0.0067	10	0	No
4	0.0133	20	1	Yes
5	0.0093	14	0	No
6	0.0320	48	1	Yes
7	0.0100	15	0	No
8	0.0420	63	2	Yes
9	0.0180	27	0	No
10	0.0273	40	4	Yes



（a）Fluctuations of heart rate signals （b）Average heart rate

（c）Standard deviation of heart rate （d）Peak heart rate

Fig. 6. Results of Heart Rate Characteristics

Processing and Analysis of Electroencephalogram (EEG) Data. The electroencephalogram (EEG) data collected from the EMOTIV Epoc+wireless portable electroencephalograph needs to be exported as a CSV file using Emotiv PRO and then imported into Pycharm. The EEG signals are converted into floating-point values by a

14-bit analog-to-digital converter and stored by Emotiv PRO. The DC level is approximately 4200 μV. To extract the EEG features, a 0.16Hz first-order high-pass filter is used to eliminate the DC offset and long-term drift of the original EEG signals.

After noise reduction, calculate the average value of the average power spectral density of the θ, α, β, and γ waves of the driver's EEG signals per minute, and observe the changes in the brain waves of the driver from the awake state to the fatigued state within 30 min. The results show that as the driving time prolongs, the average power spectral density of the θ and α waves fluctuates significantly and shows a slow upward trend, while the β and γ waves have small fluctuations and show a slow downward trend. Finally, calculate the average power spectral density ratios of R(α/β), R(θ/β), and R(α + θ/β) of the driver. As can be seen from Fig. 7, R(α/β) changes little over time, while R(θ/β) and R(α + θ/β) change greatly over time and show an upward trend.

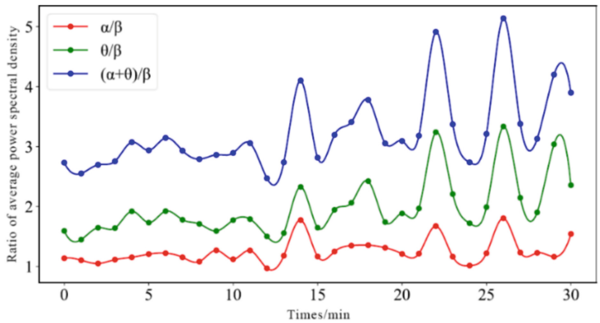


Fig. 7. Changes in the Average Power Spectral Density Ratios of $R(\alpha/\beta)$, $R(\theta/\beta)$, and $R(\alpha + \theta/\beta)$.

5 Model Construction

5.1 Model Construction and Evaluation

The fatigue driving results obtained from the recognition of facial features by the BP neural network, the recognition of heart rate features by the CNN network model, and the recognition of electroencephalogram features by the LSTM network model are used as inputs. The SVM is employed for decision-level fusion, and the true state of the driver is used as the output to train the model.

To evaluate the performance of the fatigue detection model, 300 sets of data are selected from the test set for testing, and common metrics such as accuracy, precision, and recall are examined. Each of these metrics can reflect the performance of the model in different dimensions, which helps to understand its advantages, disadvantages, and feasibility in practical applications. The confusion matrix of the test results is shown in Fig. 8, and the accuracy, precision, and recall of the model can be calculated based on the confusion matrix.

The calculation results are shown in Table 2. The precision of the model for identifying the awake state is 92.2% and the recall is 88.2%. The precision of the model

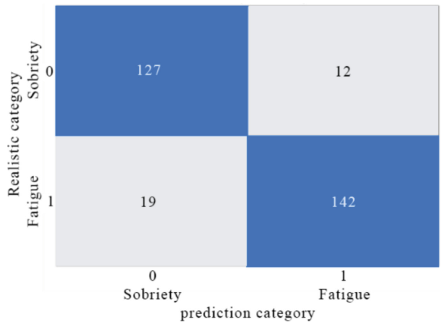


Fig. 8. Confusion Matrix

for identifying the fatigued state is 87.0% and the recall is 91.4%. The accuracy of the model is 89.7%, indicating that the model has certain generalization ability.

Table 2. Accuracy, Precision, and Recall of the Model

State recognition	Accuracy	Precision	Recall
Sobriety	89.7%	92.2%	88.2%
Fatigue		87.0%	91.4%

5.2 Potential Intervention Methods for Improving Drivers' Fatigue

This study proposes that multimodal fatigue stimuli include olfactory stimuli and auditory stimuli. When the system identifies that the driver is fatigued (the fatigued state is set as 1), the stimulation mechanism will be activated. The situation of multimodal fatigue regulation is shown in Fig. 9.

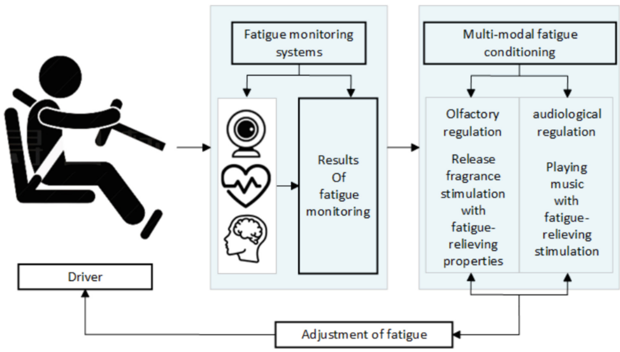


Fig. 9. Multimodal Fatigue Regulation Diagram.

First is the olfactory stimulation. When it is monitored that the driver is in a situation of fatigue driving or other conditions unfavorable to driving, we can release fragrance stimuli that can relieve and regulate the driver's fatigue through the olfactory stimulation strategy, so as to achieve the regulation of the driver's fatigue.

Similarly, for auditory stimulation, when the system monitors that the driver is in a fatigued state, music can be used to relieve the driver's fatigue. There are many problems in the possibility of using music to relieve the driver's fatigue state, such as the duration of music playback and what kind of music should be used to regulate fatigue. This requires our system to have a higher-level intelligence. In addition, characteristics such as individual differences and cognitive differences should also be taken into account, and this regulation strategy also needs more experience to verify its effectiveness and reliability.

6 Conclusion

In terms of the multimodal fatigue state recognition model, through in-depth analysis of facial, heart rate, and electroencephalogram features and the construction of the BCL-SVM model, the effective extraction and utilization of multimodal features have been achieved. The experimental results show that the accuracy of this model reaches 89.7%, the precision for the awake state is 92.2%, the precision for the fatigued state is 87.0%, and the recall rates are 88.2% and 91.4% respectively. This indicates that it has good generalization ability and provides a reliable technical approach for fatigue driving detection.

In terms of data collection and analysis, the use of a driving simulation platform and the recruitment of 24 different-type participants have enabled the collection of rich facial videos, heart rate signals, and electroencephalogram data. The detailed processing and analysis of these data have determined reasonable thresholds for facial features and revealed the changing trends of heart rate and electroencephalogram features from the awake state to the fatigued state, laying a solid foundation for model training and fatigue research.

The model of the adaptive adjustment method for intelligent car seats proposed based on the fatigue recognition results is innovative. When fatigue is detected, olfactory and auditory stimuli can effectively relieve the driver's fatigue and improve driving comfort and safety. In addition, although the multimodal fatigue-stimulation adjustment method still needs further research on individual differences in music adjustment, it provides new ideas for improving the driver's fatigue state.

This research has the potential to reduce the incidence of traffic accidents and enhance road traffic safety. Future research can focus on optimizing the multimodal fatigue-stimulation adjustment strategy, enhancing the adaptability of the model to complex environments and individual differences, and promoting the wide application of related technologies.

References

1. Sha, C., Li, R., Zhang, M.: Research on fatigue driving detection based on steering wheel grip force. Sci. Technol. Eng. **16**(30), 299–304 (2016)

2. Zhu, Y., Zhu, Y., Feng, Z.: Research on driving fatigue detection method based on EEG signal characteristics. J. Changchun Univ. Sci. Technol. (Nat. Sci. Ed.) **39**(05), 119–122 (2016)
3. Wang, F., Shichao, W., Liu, S.: Driving fatigue detection based on deep migration learning of EEG signals. J. Electron. Inf. **41**(09), 2264–2272 (2019)
4. Ma, S., Zhao, C., Sun, H.: A yawn detection algorithm based on convolutional neural network. Comput. Sci. **45**(S1), 227–229+241 (2018)
5. Zhu, F., Chen, J., Chen, J.: Driver fatigue detection based on improved Yolov3. Sci. Technol. Eng. **22**(08), 3358–3364 (2022)
6. Cao, J.: Fatigue Driving Monitoring Based on the Fusion Features of GELM Electroocular Signal and Visual Information. Northeastern Electric Power University, Jilin (2022)
7. Wang, Q.J., Mu, Z.D.: Heterogeneous signal fusion method in driving fatigue detection signals. J. Adv. Transp. **2021**, 4464890–4464901 (2021)
8. Xu, J.L., Min, J.L., Hu, J.F.: Real-time eye tracking for the assessment of driver fatigue. **5**(2), 54–58 (2018)
9. Dinges, D., Grace, R: PERCLOS: A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance. Federal Highway Administration, Office of Motor Carriers, Office of Motor Carrier Research and Standards, Washington (1998)
10. Miljevic, A., Bailey, N.W., Murphy, O.W., et al.: Alterations in EEG functional connectivity in individuals with depression: a systematic review. J. Affect. Disord. **328**, 287–302 (2023)
11. Jing, X.: Development of a Real-Vehicle Experimental Platform for Fatigue Driving Recognition and Intervention and Research on Odor Intervention Technology. Yangzhou University, Yangzhou (2020)
12. Pincus, S.M.: Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. U.S.A. **88**(6), 2297–2301 (1991)
13. Lecun, Y., Bottou, L., Bengio, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
14. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
15. Amidei, A., Poli, A., Iadarola, G., et al.: Driver drowsiness detection based on variation of skin conductance from wearable device. In: Proceedings of the 2022 IEEE International Workshop on Metrology for Automotive (Metro Automotive) (2022)
16. Li, Z.J., Chen, L.K., Nie, L., et al.: A novel learning model of driver fatigue features representation for steering wheel angle. IEEE Trans. Veh. Technol. **71**(1), 269–281 (2022)
17. National Highway Traffic Safety Administration. Drowsy driving [EB/OL]. 6 March 2022. https://www.nhtsa.gov/risky-driving/drowsy-driving

The Impact of Integration Between Visual and Haptic Texture Simulations on Comprehension of Counterfactual Artifacts in Mixed Reality

Ming-Chieh Chiang(✉) and Shih-Hung Cheng

National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Da'an District, Taipei City 106335, Taiwan (R.O.C.)
m11310103@gapps.ntust.edu.tw

Abstract. With the rapid advancement of virtual reality (VR) technology, haptic texture stimulation may serve as a crucial factor in achieving high-quality immersive experiences. Based on the principles of virtual sensory integration, this study investigates the cognitive impact of counterfactual design under haptic texture stimulation. Using a common object—a chair—as the virtual model in a mixed-reality environment, we examine the integration of three visual textures (original wood grain, ultra-smooth metal with high luminance, and an absolute light-absorbing material with low luminance) with two physical properties: surface roughness and temperature. In Experiment A, the haptic factor consisted of three levels of surface roughness (sandpaper #80, #600, and #1000), while Experiment B manipulated three seat temperatures (5 °C cold cushion, 15–20 °C room temperature cushion, and 50 °C heated cushion). Preliminary data analysis revealed that the absolute light-absorbing material provided a strong sense of compatibility under virtual visual stimuli, ultra-smooth metal evoked significant dissonance when paired with rough textures, and the 50 °C heated cushion induced notable perceptual conflict. Qualitative feedback suggested that participants tended to rely on haptic information to guide visual perception. The findings indicate that visuo-haptic consistency significantly affects immersion and trust, while roughness and temperature notably impact immersion. Additionally, the interaction between visual and thermal stimuli was significant, particularly with the absolute light-absorbing material. Individual differences were observed in the acceptance of counterfactual materials, likely depending on application contexts. In precision industrial applications, maintaining visuo-haptic consistency remains crucial.

Keywords: Mixed Reality · Counterfactual Artifacts · Cognitive Process · Haptic Texture

1 Research Background and Motivation

With the continuous advancement of Head-Mounted Display (HMD) technology, Virtual Reality (VR) has gradually evolved into Mixed Reality (MR), which integrates virtual objects into real-world environments [1]. MR allows virtual objects to be superimposed

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 306–324, 2025.
https://doi.org/10.1007/978-3-031-93715-6_19

onto the physical world, enabling users to interact with both virtual objects and environments, thereby enhancing interactivity and reinforcing the sense of immersion in virtual spaces [2].

Currently, VR technology primarily relies on visual stimuli as the main modality for experience delivery [3]. Although VR attempts to incorporate multiple sensory modalities (e.g., auditory stimuli), vision often dominates the user's perception and sense of immersion. Present implementations of haptic feedback are mainly achieved through controllers, such as vibration feedback and button interactions, providing only basic integration with visual information. However, fully replicating real-world sensory experiences remains a challenge, particularly in the fine-grained simulation of surface textures [4].

Haptic simulation plays a critical role in enhancing users' immersion and sense of realism. If VR and MR technologies can further improve haptic simulation, user interactions within virtual environments could become more natural and engaging [5]. Moreover, multisensory integration of haptic and visual stimuli can enhance the illusion of perceptual expectation [6].

Therefore, effectively integrating visual and haptic stimuli in MR environments to enhance user immersion and trust remains an essential research topic in the pursuit of more engaging and immersive virtual experiences.

The primary objective of this study is to experimentally investigate whether the inclusion of specific haptic stimuli in MR environments can enhance user immersion and trust in virtual objects. This research will examine the impact of visual-haptic consistency on multisensory integration and explore how different material properties influence users' cognitive and emotional responses. The findings are expected to provide valuable insights into future design research, device development, and user experience improvements, aiding designers in advancing immersive interaction technologies and expanding the applications of Mixed Reality.

2 Related Theoretical Framework

Virtual environments provide possibilities for representing scenes that cannot be replicated in the real world [7], such as fantasy worlds or past events that never occurred. This creates an expansive space for designers to explore speculative design objects and scenarios, inspiring reflections on the potential and implications of emerging technologies [8]. Within this context, this study focuses on the impact of haptic stimuli on users' cognitive experiences and emotional responses. It adopts the perspective of material speculation and employs counterfactual narratives as a research framework to investigate how interactive objects in virtual environments influence users' sense of trust and immersion.

From the perspective of material speculation theory, exploring material interactions and possibilities can prompt users to deeply reflect on future scenarios [9]. This approach allows the examination of how material experiences change under different multisensory integration conditions, providing valuable insights for forward-thinking design strategies. Material perception is not limited to a single sensory modality but involves the interaction between visual and haptic cues, which can aid users in recognizing material properties such as roughness in both virtual and physical environments [10].

Haptic feedback, as a key element in multisensory integration, can significantly enhance user interaction with virtual environments. In particular, under active touch conditions, users can engage more actively with virtual experiences, leading to faster sensory integration and an enhanced sense of immersion in virtual environments [11, 12].

Human perception is fundamentally a result of multisensory integration, with vision often playing a dominant role [13]. When visual and haptic perceptions align, the accuracy of perception is maximized [14]. Proper haptic simulation not only enhances the realism of materials but also facilitates cross-modal recognition of material properties, reducing sensory separation effects [10, 13] and even inducing sensory illusions across different modalities [15]. The surface characteristics of materials can generate highly consistent cross-modal perceptions through the interaction of visual and haptic stimuli, further improving the intuitiveness of design and system usability [16].

3 Experimental Design

3.1 Research Hypotheses

This study aims to explore how counterfactual virtual materials in an MR environment affect users' immersion and trust. Additionally, it examines whether the consistency between visual and haptic stimuli enhances multisensory integration. The study particularly focuses on the influence of roughness and temperature on sensory integration and user experience.

The study hypothesizes that when users enter an MR environment, their initial sense of immersion and trust will be influenced by the counterfactual nature of virtual materials. Furthermore, after incorporating haptic stimuli, if the visual and haptic feedback are consistent, immersion and trust will be enhanced; otherwise, they may decrease. Specifically, when counterfactual design elements appear in the virtual scene, users may experience perceptual incongruence (e.g., conflicts between visual and haptic sensations or unrealistically exaggerated material properties), which could trigger critical thinking. This cognitive response may temporarily weaken immersion. Conversely, when visual and haptic perceptions align, multisensory integration is strengthened, potentially mitigating the negative impact of counterfactual design.

This study also seeks to investigate the interaction effects of haptic stimuli on consistency and immersion. The experiment involves three types of visual material textures—wood, ultra-smooth metal, and absolute light-absorbing black—combined with two haptic factors (roughness and temperature) to simulate real-world material properties.

3.2 Participants and Sample Characteristics

This study recruited thirty participants aged 18 and above, with no restrictions on gender or background. All participants signed an informed consent form before the experiment and were informed that their data privacy would be fully protected throughout the study. Each participant underwent two sets of material tests and evaluations, encountering a total of eighteen stimulus conditions, and was required to provide subjective feedback on their sensory experience.

3.3 Experimental Design and Stimulus Description

This study selected the Ply Chair, a classic design by Jason Morrison, as the experimental stimulus. The chair adheres to the basic prototype of a standard chair, and its minimalist design helps control experimental variables. Its flat and spacious seat surface facilitates the replacement of experimental materials, while the simplicity of its backrest helps minimize potential biases (Fig. 1).

Fig. 1. Ply Chair by Jason Morrison.

Fig. 2. Texture mapping is integrated with a real chair.

In the VR environment, this study incorporates three types of virtual visual material textures, including:

1. Natural Wood Grain – closely resembling the original Ply Chair surface material.
2. (High-luminance) Ultra-Smooth Metal – set to white base color, smoothness 100, reflectivity 50.
3. (Low-luminance) Absolute Light-Absorbing Black Material – set to black base color, smoothness 0, reflectivity 0.

This study employs counterfactual material narratives (high-gloss metallic and absolute light-absorbing black) to define the visual presentation of the experiment. To ensure high contrast between material textures for easier participant perception, colors with high and low brightness levels were selected.

For the visual representation within the VR environment, the experimental setup was developed using Unity, incorporating all virtual materials and texture-switching functions tailored for a mixed reality setting. The virtual Ply Chair model was adjusted to match the dimensions of the physical chair, with a total height of 84 cm and a seat height fixed at 46 cm from the ground. The virtual model's position was further adjusted based on each participant's height to ensure proper alignment between the camera perspective and the real-world reference. The HMD's field of view (FOV) was set to 110°, simulating the natural human visual range (Fig. 3).

This study conducted two MR experiments on visual-haptic integration. Experiment A focused on roughness (as shown in Fig. 4), while Experiment B examined temperature (as shown in Fig. 5).

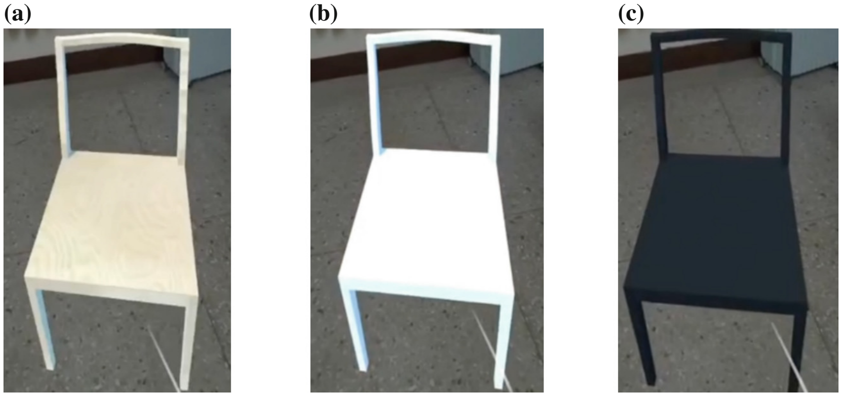


Fig. 3. (a). Natural Wood Grain. (b). Ultra-Smooth Metal. (c). Absolute Light-Absorbing Black Material.

A. Surface roughness: #80 cloth sandpaper(a), #600 cloth sandpaper(b), #1000 water sandpaper(c)

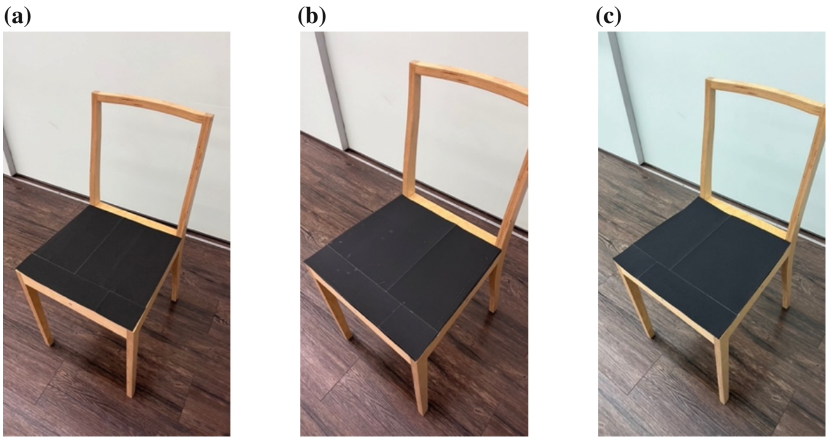


Fig. 4. (a). #80 cloth sandpaper. (b). #600 cloth sandpaper. (c). #1000 water sandpaper.

B. Temperature: room-temperature cushion (a), cooling cushion (b), heated cushion (c)

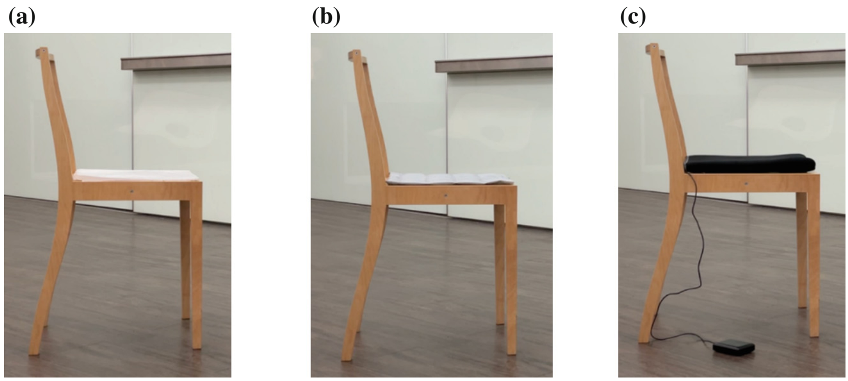


Fig. 5. (a). 15–20 °C room-temperature cushion. (b). 5 °C cooling cushion. (c). 50 °C heated cushion.

3.4 Experimental Procedure and Workflow

Both experiments in this study involve the combination of three visual material textures with three tactile characteristics. After comfortably wearing the Head-Mounted Display (HMD), participants will see a ply chair within the MR environment (as shown in Fig. 2). Once the virtual chair aligns with the physical chair, the experiment begins.

In experiment A (roughness), participants will first observe the chair with the original wood texture and rate its visual congruence on a 1–7 scale, where 1 represents extreme incongruence and 7 represents strong congruence. They will also provide a brief explanation for their rating. Next, three sandpapers with different roughness levels (#80/#600/#1000) will be sequentially placed on the chair's seat. Participants will touch and sit on each material, rating the tactile experience on the same 1–7 scale for each action.

Upon completing experiment, A (roughness), participants will proceed to experiment B (temperature), where they will sequentially experience three seat materials: a room-temperature cushion, a cooling cushion, and a heated cushion.

After completing both experiments, a final integrative question will be asked: "In the experiments on temperature and visual integration, did you perceive the experience as more visually guided or haptically dominant?" Participants will again rate their experience on a 1–7 scale, where 1 indicates strong visual dominance and 7 indicates strong haptic dominance, followed by a brief verbal explanation of their reasoning.

4 Experimental Data Analysis

This study collected quantitative data to examine the overall impact of visual-haptic integration under different conditions. The primary focus was on how visual stimuli (material textures), temperature, and roughness influenced perceived incongruence ratings and the interactions among these variables. The following sections present the analysis results:

4.1 Effect of Visual Stimuli on Incongruence Ratings

This study first examined whether there were significant differences in perceived incongruence ratings across different visual stimuli (natural wood grain, ultra-smooth metal, and absolute light-absorbing black) when participants relied solely on vision without physical interaction.

A one-way ANOVA test was conducted, revealing that visual material significantly influenced incongruence ratings ($F = 7.635$, $p = .001$). The homogeneity of variance assumption was met ($p = .644$), indicating that variances across conditions were comparable, allowing for direct comparison.

Post-hoc tests (LSD and SNK) showed that he absolute light-absorbing black material had the highest incongruence rating (Mean = 5.43), and the natural wood grain (Mean = 4.50) and ultra-smooth metal (Mean = 3.87) had significantly lower incongruence ratings (Fig. 6).

Pairwise comparisons showed that the absolute black material was significantly different from both the wood grain ($p = .023$) and metal ($p < .001$), while the difference between wood and metal was not significant ($p = .120$).

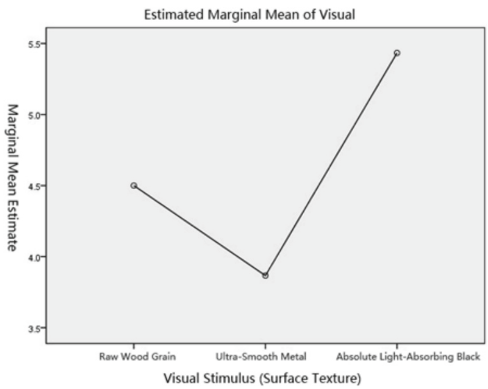


Fig. 6. Perceived Visual Consistency Across Different Surface Textures.

The results indicate that when visual information is limited, such as with the absolute light-absorbing black material, participants tend to assign higher scores for visual-haptic integration. This may be because the reduced visual cues make it more difficult for participants to form clear visual expectations, leading them to rely more on haptic perception.

4.2 Effect of Roughness on Incongruence Ratings

A one-way analysis of variance (ANOVA) was conducted to examine the effects of different roughness levels (#80, #600, and #1000 sandpaper) in both touch and sitting conditions. The results showed that roughness had a significant impact on incongruence scores ($F = 42.324$, $p < .001$), whereas in the sitting conditions, no significant difference was found ($p = .735$).

Post hoc LSD and SNK tests revealed that the high-gloss metallic surface had a significantly lower mean score (Mean = 4.28) compared to the original wood texture, indicating that participants experienced the strongest sense of incongruence with the metallic texture. The other two textures showed significantly higher congruence scores than the metallic texture.

These results suggest that haptic feedback plays a more significant role in incongruent perception, with hand touch being more effective than sitting in detecting roughness variations, as roughness had a smaller influence on the sitting condition.

4.3 Effect of Temperature on Incongruence Ratings

This study examined the effects of different temperature conditions on perceived incongruence scores in both touch and sitting scenarios. The results indicated that temperature had a significant effect in both conditions, with touch ($F = 24.892$, $p < .001$) and sitting ($F = 18.661$, $p < .001$) showing notable differences.

Post hoc LSD and SNK tests revealed that when surface temperature was high (50 °C heated cushion), the incongruence score for touch (Mean = 3.14) was significantly lower than that for the room-temperature cushion (Mean = 4.39) and cooling cushion (Mean = 4.60). Similarly, in the sitting condition, the 50 °C heated cushion (Mean = 3.29) received a significantly lower incongruence score compared to the room-temperature cushion (Mean = 4.40) and cooling cushion (Mean = 4.64).

These findings suggest that high-temperature environments may reduce the perceived incongruence between visual and haptic stimuli, as participants tend to focus more on temperature sensations, thereby influencing their perception and ratings.

4.4 Interaction Between Visual and Temperature Effects

This study further analyzed the effects of visual texture (surface material) and temperature (room-temperature cushion, cooling cushion, heated cushion) on perceived congruence scores and examined whether an interaction exists between these factors. Using multivariate tests, the results showed that visual texture (Pillai's Trace = .480, $F = 41.198$) and temperature (Pillai's Trace = .163, $F = 11.551$) significantly influenced congruence scores ($p < .001$). This indicates that both visual and temperature factors affect participants' evaluations of material congruence.

Moreover, a significant interaction effect between visual texture and temperature was found (Pillai's Trace = .312, $F = 12.065$, $p < .001$), suggesting that the impact of temperature varies depending on the material type (Figs. 7 and 8). Figure 7 illustrates the interaction effect on haptic perception, while Fig. 8 shows the effect in the seated condition.

In contrast, roughness showed only marginal significance ($p = .049$), indicating a relatively weaker effect compared to the stronger influence of visual and temperature factors.

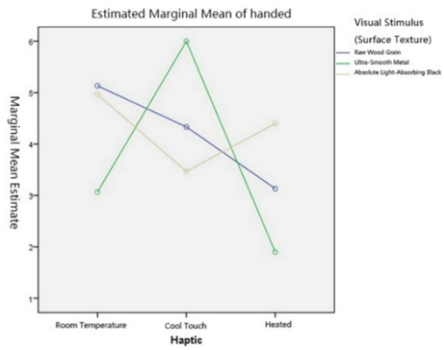


Fig. 7. Interaction Effects of Surface Texture and Temperature on **Haptic** Perception.

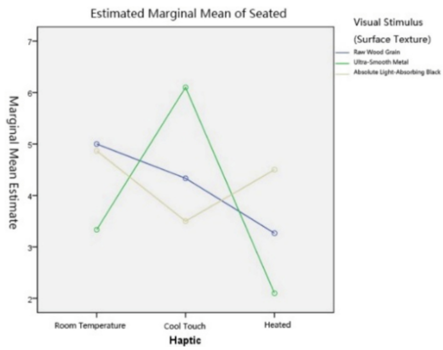


Fig. 8. Interaction Effects of Surface Texture and Temperature on **Seated** Perception.

4.5 Summary of Findings

This chapter presents statistical analyses of five key evaluations: visual-haptic integration perception (roughness and temperature), immersion changes, the effect of visual-haptic integration on immersion, and counterfactual design acceptance. The analysis includes descriptive statistics, paired t-tests, and correlation analysis to examine relationships and significance among variables. Additionally, frequency analysis was conducted to identify the most and least congruent material-condition combinations.

For visual-haptic integration ratings, 1 represents strong visual dominance, 7 represents strong haptic dominance, and 4 indicates balance. Other evaluation scores range from 1 (completely disagree) to 5 (completely agree), with 3 as neutral.

Descriptive Statistics

Roughness and temperature congruence ratings averaged above 4.5, indicating that participants generally relied more on haptic feedback than visual cues. Immersion ratings averaged 3.50 (on a 1–5 scale), suggesting a moderate-to-high level of immersion. The impact of visual-haptic integration on immersion was rated 3.93 (1–5 scale), showing that most participants believed sensory integration affected their immersive experience. Acceptance of counterfactual materials was rated relatively low ($M = 3.40$), indicating limited acceptance among participants (Table 1).

Table 1. Descriptive Statistics of Key Variables

Variable	Min	Max	Mean	SD
Visual vs. tactual (roughness)	2	7	4.93	1.574
Visual vs. tactual (temperature)	1	7	4.60	1.831
Immersion	2	5	3.50	1.042
Visual and tactile integration changes immersion	1	5	3.93	1.048
Counterfactual acceptance	1	5	3.40	1.248

Paired Sample t-Tests. To examine significant differences between conditions, paired sample t-tests were conducted (Table 2):

Table 2. Independent Samples t-test for Visual-Haptic Integration and Counterfactual Design (All tests were conducted with df = 29.)

Comparison	t-value	p-value	95% Confidence Interval
Visual-Haptic Integration Perception (Roughness) vs. (Temperature)	−0.952	.349	[−1.049,0.383]
Visual-Haptic Integration Perception (Roughness) vs. Immersion	4.285*	<.001	[0.749,2.118]
Visual-Haptic Integration Perception (Temperature) vs. Immersion	2.559*	.016	[0.221,1.979]
Impact of Visual-Haptic Integration vs. Immersion	1.783†	.085	[−0.064,0.930]
Acceptance of Counterfactual Design vs. Immersion	−0.367	.717	[−0.658,0.458]
Impact of Visual-Haptic Integration vs. Acceptance of Counterfactual Design	1.743†	.092	[−0.093,1.159]

Findings suggest that roughness significantly influences immersion ($p < .001$), whereas temperature has a moderate effect ($p = .016$). Visual-haptic integration shows borderline significance in affecting immersion ($p = .085$), indicating a potential but weak influence. Counterfactual material acceptance does not significantly impact immersion ($p = .717$), suggesting that user skepticism toward counterfactual materials is independent of their immersion level.

Correlation Analysis

A correlation analysis was conducted to examine the relationships between key variables. The results indicated that significant correlation between visual-haptic integration perception ($r = .373$, $p = .042$), suggesting that participants rated these two sensory dimensions with some consistency (Table 3).

No significant correlation between visual-haptic consistency and immersion ($p > .05$), implying that visual-haptic integration does not have a strong linear relationship with immersion. No significant correlation between counterfactual material acceptance and other variables, suggesting that participants' acceptance of counterfactual materials might be influenced by other unknown factors. Overall, the analysis showed that roughness and temperature perception were related, but their impact on immersion was not strongly linear. Counterfactual material acceptance appeared to be independent of other measured factors.

Table 3. Pearson Correlation Analysis Between Key Variables.

Comparison	Pearson's r	p-value
Visual-Haptic Integration Perception (Roughness) vs. (Temperature)	.373*	.042
Visual-Haptic Integration Perception (Roughness) vs Immersion	.063	.741
Visual-Haptic Integration Perception (Temperature) vs. Immersion	−.289	.121
Impact of Visual-Haptic Integration vs. Immersion	.189	.316
Acceptance of Counterfactual Design vs. Immersion	.159	.401
Impact of Visual-Haptic Integration vs. Acceptance of Counterfactual Design	−.058	.761

Most Congruent and Most Incongruent Combinations
Participants were asked to identify the most congruent and most incongruent material-variable combinations. As shown in Fig. 9, the most congruent combination was ultra-smooth metal with a cooling cushion. Participants' preferences were relatively dispersed, with natural wood grain being the most frequently chosen surface texture. During qualitative interviews, participants mentioned that wood was the most familiar material in daily life, making it easier to conceptualize and relate to. As shown in Fig. 10, the most incongruent combination was ultra-smooth metal with #80 sandpaper, followed by ultra-smooth metal with a heated cushion. This suggests that participants perceived metal-related combinations as the most conflicting, particularly when paired with rough textures or unexpected temperature variations.

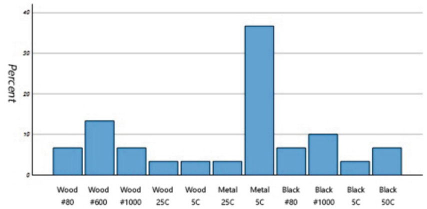


Fig. 9. Most congruent Combinations.

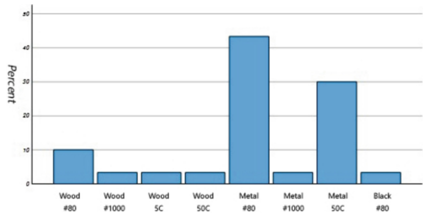


Fig. 10. Most Incongruent Combinations.

5 Participant Behavior and Qualitative Interview Analysis

5.1 Research Discussion

The Impact of Visual-Haptic Congruence. The experimental results show that in the visual-only phase (the first step of the experiment), participants perceived the absolute light-absorbing black surface as the most congruent. This could be due to its extremely low reflectivity, making it appear more naturally integrated into the virtual environment. Additionally, some participants noted that because the study did not explicitly specify whether the "black material" was plastic, fabric, or another type, they were more lenient in accepting its visual texture, leading to higher ratings. However, responses to the black material varied among participants, while 40% of participants believed that a highly light-absorbing black material should feel rough due to the way rough surfaces scatter light and increase absorption, 26.7% expected the material to feel soft, like velvet. Consequently, those who anticipated a soft texture perceived roughness as incongruent, leading to lower congruence ratings. This divergence in expectations resulted in a wider spread of ratings, reducing the statistical significance of the material's effect.

In contrast, natural wood grain had slightly lower congruence ratings, which may be attributed to virtual lighting conditions that do not perfectly replicate real-world lighting. This discrepancy caused the virtual wood color to appear lighter than expected. Participants had already seen the real chair before the experiment began, and if the virtual version did not match exactly, it increased perceived incongruence. Among the three textures, ultra-smooth metal had the lowest congruence ratings, possibly due to the absence of a perfectly mirrored surface in the virtual environment, which reduced its perceived realism. Additionally, inconsistencies in mixed Reality (MR) lighting, where real-world illumination interfered with virtual reflections, may have contributed to the material appearing unnatural.

Some participants provided more detailed observations:

"A visually smooth material should reflect light differently at different angles, but the chair's seat and backrest have inconsistent reflections, which affects its realism." (Participant 30).

"Compared to the previous material, this one feels less rough, but it still doesn't match my expectation of metal." (Participant 14).

About 30% of participants reported haptic illusions and increased tactile sensitivity when interacting with the ultra-smooth metal texture and #600 sandpaper combinations. Additionally, some participants noticed temperature changes, even though no actual temperature variable was introduced in this phase:

"It feels a little cold, just a bit." (Participant 10).

"Compared to what I imagined, this metal feels smoother and cooler." (Participant 18).

This suggests that participants subconsciously associated metal with coolness, leading them to perceive slight temperature variations even when none were present.

Interaction Between Visual and Temperature Stimuli. Data analysis revealed that when participants sat on a heated cushion, their incongruence rates significantly decreased, meaning they were more likely to perceive the experience as congruent. This effect may be attributed to increased sensory sensitivity at higher temperatures, which caused participants to focus more on temperature differences rather than visual-haptic mismatches. Additionally, the cognitive load effects suggest that strong temperature stimuli demanded greater attention, thereby reducing participants' ability to process minor visual-haptic conflicts. Conversely, in the room temperature and cooling cushion conditions, incongruence ratings remained stable. Some participants noted that room temperature and cooling cushions felt similar, possibly due to external environmental factors, such as the winter season during testing (15–20 °C ambient temperature), which may have affected temperature perception.

Further statistical tests confirmed a significant interaction between visual and temperature stimuli ($p < .001$), indicated that the effect of temperature on incongruence depended on the specific visual material being presented. For the absolute light-absorbing black material, temperature changes had a stronger effect on incongruence ratings, suggesting that participants relied more on haptic cues when visual information was minimal. In contrast, for the ultra-smooth metal material, temperature changes had a weaker effect, possibly because participants expected metals to be cold. When the temperature of the metal did not match their expectations, the inconsistency was harder to resolve, resulting in lower congruence ratings regardless of temperature.

Analysis of Incongruence Ratings in Touch vs. Sitting Conditions.
The data revealed that when participants evaluated materials through touch alone, they rated ultra-smooth metal as the most incongruent material, particularly when paired with rough textures (#80, #600 sandpaper). However, in the sitting conditions, incongruence ratings remained stable, with no significant changes. This could be attributed to two factors: weight distribution effects and seasonal effects. When sitting, pressure is distributed over a larger contact area, making coarse textures less noticeable than when touching with fingers. Additionally, since the study was conducted in winter, participants wore thicker clothing, which may have reduced sensitivity to tactile variations when sitting compared to using their hands.

Thus, haptic conflicts were more pronounced in the touch condition than in the sitting condition.

5.2 Participant Feedback and Perceptual Evaluation

Visual-Haptic Integration Perception (Roughness vs. Temperature). To understand participants' reliance on visual or haptic perception under different conditions, the questionnaire evaluated roughness and temperature on a seven-point scale (1 = fully vision-dominant, 7 = fully haptic-dominant).

Results showed that participants generally relied more on haptic perception, with mean scores for roughness (4.93) and temperature (4.60) both exceeding 4.5. Roughness provided clearer sensory information, making participants more dependent on touch to perceive surface properties. Some reported forming initial expectations based on visual cues but relied on touch to confirm or refine their perception, especially for unfamiliar counterfactual materials like absolute light-absorbing black.

Around 16% of participants exhibited different sensory reliance across conditions. Those who relied more on vision for roughness tended to depend on haptics for temperature, as they could not see the chair while sitting and found temperature perception more intuitive.

> *"Touch immediately conveys temperature, while vision helps imagine it first, but touch is more direct." (Participant 3).*

Conversely, some prioritized haptics for roughness but vision for temperature, as they used vision to form an initial material impression and touch to verify it.

"After knowing the material, I imagine its texture. Sitting confirms whether it matches my expectation, so I rely more on vision." (Participant 5).

Immersion Ratings. Participants rated their immersion on a five-point scale (1 = not immersive at all, 5 = fully immersive). The average immersion score was 3.50, indicating a generally high level of engagement in the experimental environment.

A paired t-test was conducted to compare four factors (visual-haptic integration for roughness and temperature, the impact of visual-haptic integration, and counterfactual design acceptance) against immersion. The results showed a significant difference between roughness perception and immersion ($p < .001$), suggesting that haptic feedback from surface texture influenced participants' sense of immersion in the virtual environment. Temperature had a weaker but still significant effect on immersion ($p = .016$), indicating that while temperature changes affected immersion, their impact was less pronounced than roughness. Additionally, roughness and temperature scores were significantly correlated ($p = .042$), suggesting consistency in participants' sensory evaluations, though their influence on immersion may differ.

The Impact of Visual-Haptic Integration on Immersion. This experiment hypothesized that the introduction of virtual objects along with haptic stimulation could influence immersion. To test this, participants were first asked to rate their overall immersion at the end of the experiment. Using this as a baseline, they then rated the impact of visual-haptic integration on immersion.

The results showed an average score of 3.93 (on a 1–5 scale), indicating that most participants perceived an increase in immersion. A paired t-test comparison between the two scores revealed marginal significance ($p = .085$), suggesting that visual-haptic integration had some effect on immersion, though not strongly significant.

Acceptance of Counterfactual Design Objects. The analysis of counterfactual design object acceptance revealed an average rating of 3.40 (on a 1–5 scale), indicating moderate acceptance among participants. Comparisons between this variable and other measures showed no significant correlation, suggesting that participants' acceptance of counterfactual materials was relatively independent of other factors such as immersion or visual-haptic congruence. Some participants mentioned that the counterfactual material descriptions lacked clarity, which may have influenced their ability to fully accept these materials. This aligns with prior research suggesting that a more structured world-building approach could enhance the believability of counterfactual design elements [9].

"I think the materials should be better integrated with real-world environments. Experiencing them in an actual physical setting might improve realism." (Participant 18).

Some participants also noted that extremely unrealistic material properties (e.g., ultra-smooth surfaces paired with rough tactile feedback) reduced their willingness to accept counterfactual designs.

Summary of Findings and Hypothesis Validation. This study aimed to investigate the effects of visual-haptic congruence on immersion and trust, as well as the role of counterfactual design in user experience. The results provide empirical support for the hypotheses, leading to the following key conclusions:

Hypothesis 1: The effect of counterfactual design on immersion.

1. Visual-haptic mismatch reduces immersion, but the impact varies by material.

Participants were most sensitive to mismatches in the high-gloss metallic surface combined with #600 sandpaper, which resulted in the lowest immersion scores. In contrast, the absolute light-absorbing black condition, despite being a counterfactual design, had a less pronounced effect. Some participants accepted the material's ambiguity and retained more imaginative flexibility, preventing a significant drop in immersion.

When visual-haptic mismatch occurs, participants invest additional cognitive effort to adjust expectations, affecting immersion.

In the high-gloss metallic surface condition, most participants expected metal to feel smooth, and any roughness significantly disrupted their perception. This type of mismatch required more cognitive adjustment, leading participants to focus more on tactile discrepancies. They actively described texture variations, suggesting that repeated comparison and analysis influenced their sense of immersion.

2. The impact of counterfactual design varies by individual.

Some participants found visual-haptic mismatches intriguing rather than immersion-breaking. In the absolute light-absorbing black condition, certain participants embraced the material's ambiguity, even reporting an enhanced immersive experience. However, in the high-gloss metallic surface condition, where visual details were more defined, participants were more sensitive to incongruences, leading to lower perceived congruence.

Overall, the findings partially support Hypothesis 1—visual-haptic mismatches do affect immersion, but the extent varies depending on the material's visual properties and level of detail. Counterfactual design can sometimes weaken immersion, but in cases where the material's properties remain ambiguous (such as black surfaces with undefined textures), it may have little effect or even enhance the immersive experience.

Hypothesis 2: The interactive effect of haptic stimulation on consistency and immersion.

3. Immersion increases when visual and haptic stimuli are consistent.

Results indicate that when participants' visual and haptic experiences were aligned, their immersion significantly increased. For example, in the absolute light-absorbing black condition combined with #80 sandpaper, many participants perceived high consistency and rated the congruence positively. Conversely, in the high-gloss metallic surface paired with #600 sandpaper, the mismatch led to decreased immersion.

4. Roughness has the strongest impact on immersion, followed by temperature.

Statistical analysis showed that roughness had a significant effect on immersion ($p < .001$), whereas temperature had a moderate but still significant effect ($p = .016$). This suggests that roughness variations were more perceptible and directly influenced immersion, while temperature perception was more subtle and indirect.

The findings support Hypothesis 2 and align with the Optimal Integration Theory of Multisensory Perception. When visual and haptic stimuli were congruent, immersion was significantly enhanced. Additionally, roughness had a greater impact than temperature, highlighting the critical role of haptic texture in material perception and immersion.

5.3 Practical Applications and Study Limitations

The findings of this study provide new insights into counterfactual material perception and visual-haptic integration in MR environments, with potential applications in various fields:

Applications

1. Counterfactual Design in Gaming and Immersive Experiences

   Immersion is a key factor in game design. This study shows that visual-haptic congruence, roughness, and temperature significantly impact immersion. Some participants found visual-haptic mismatches engaging, suggesting intentional sensory conflicts could be leveraged for unique gameplay mechanics in fantasy or sci-fi settings.
2. Industrial Design and Product Development

   Mismatched visual-haptic cues reduced trust in materials, which has implications for product design, material selection, and haptic feedback technology. In VR-based product experiences, ensuring visual and haptic alignment could improve user confidence in virtual prototyping and digital material simulation. High-temperature conditions reduced sensory conflicts, which may be applicable in smart furniture and adaptive tactile interfaces to enhance material realism through temperature modulation.

Study Limitations and Future Research Directions

1. Virtual Object Setup

   The virtual chair model had minor discrepancies in size compared to the real chair, which may have influenced participants' perception. The HMD field of view (FOV) was set to 110°, but minor perspective distortions may have affected spatial accuracy. Virtual and real-world lighting conditions differed, potentially affecting material appearance and realism.
2. Physical Stimuli Setup

   The roughness condition used sandpaper, but participants with different tactile sensitivity levels perceived the differences to varying degrees. Temperature conditions used different seat cushion materials, which may have introduced unintended texture variations. Future studies could apply identical surface materials with embedded temperature control for more precise comparisons.

3. Environmental Factors

   The experiment was conducted during winter, with ambient temperatures around 15 °C, which may have reduced the perceived differences between room temperature and cooling cushions. Participants were allowed to sit for extended durations to perceive temperature changes, but longer exposure sometimes reduced perceived realism (e.g., Participant 18 said that *"The chair shouldn't stay cold this long."*).

4. Evaluation Metrics

   Different rating scales were used for visual-haptic integration (1–7) and immersion/counterfactual acceptance (1–5), which complicated direct comparisons. The counterfactual design acceptance question was isolated, making it difficult to analyze its relationship with other factors. Future studies could integrate this metric within broader sensory evaluation frameworks.

6 Conclusion

6.1 Key Research Findings

This study explored the effects of visual-haptic congruence and incongruence on user immersion and trust while examining the role of counterfactual material design in MR environments. The key findings are summarized as follows:

1. Visual-haptic congruence significantly affects immersion and trust, while incongruence reduces them.

   Aligned visual and haptic stimuli led to higher immersion and trust, whereas incongruent stimuli resulted in a less natural experience and reduced credibility of the virtual environment.

2. Roughness and temperature significantly influence immersion.

   Roughness had a highly significant effect ($p < .001$), and temperature also showed a significant impact($p = .016$), including that both tactile dimensions play a crucial role in material perception.

3. Visual and temperature stimuli interact significantly.

   For absolute light-absorbing black materials, temperature changes had a stronger effect on perceived congruence, as participants relied more on haptic input when visual information was minimal.

For ultra-smooth metal materials, temperature had a weaker effect, as participants expected metals to feel cold, and deviations from this expectation were harder to resolve.

1. Acceptance of counterfactual materials depends on individuals and contexts.

   Some participants found visual-haptic incongruence interesting and engaging, while others found it disruptive to immersion. Counterfactual materials may be more suitable for entertainment and speculative design applications (e.g., games, virtual storytelling) rather than industrial applications requiring precision.

6.2 Research Contributions and Future Directions

This study quantified the impact of visual-haptic consistency on immersion and trust, addressing a gap in prior research on visual-haptic interaction. It also explored the feasibility of counterfactual design, offering new insights into whether mismatched sensory cues can create novel experiences beyond mere congruence.

Applications. The findings can inform haptic feedback technologies, enhancing immersive experiences by aligning material and temperature cues in virtual environments. In industrial design, the role of visual-haptic consistency in trust perception has implications for prosthetic design, smart materials, and virtual product displays.

Future Research. Further exploration is needed, including additional material properties like elasticity and stickiness, to provide a more comprehensive analysis of visual-haptic integration. Since the participant sample was relatively homogeneous, future studies should examine individual differences across diverse populations. Expanding research into other sensory modalities (e.g., auditory and olfactory interactions) could also deepen our understanding of multisensory integration in immersive experiences.

Acknowledgments. This research was funded by the National Science and Technology Council of Taiwan grant number NSTC 113-2410-H-011-012-.

Conflicts of Interest. The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Tenmoku, R., Ichikari, R., Shibata, F., Kimura, A., Tamura, H.: Design and prototype implementation of MR pre-visalization workflow. Research Organization of Science and Engineering, Ritsumeikan University & Graduate School of Science and Engineering, Ritsumeikan University (2006)
2. Nair, M., Fernandez, R.E.: Advancing mixed reality digital twins through 3D reconstruction of fresh produce. IEEE Access **12**, 4315–4326 (2024). https://doi.org/10.1109/ACCESS.2023.3348934
3. Cornelio, P., et al.: Virtually tasty: an investigation of the effect of ambient lighting and 3D-shaped taste stimuli on taste perception in virtual reality. Int. J. Gastronomy Food Sci. **30**, 100626 (2022). https://doi.org/10.1016/j.ijgfs.2022.100626
4. Sandor, C., Kuroki, T., Uchiyama, S., Yamamoto, H.: Exploring visuo-haptic mixed reality. Human Machine Perception Laboratory, Canon Inc., Tokyo (2006)
5. Kim, D., Kim, Y., Jo, D.: Exploring the effect of virtual environments on passive haptic perception. Appl. Sci. **13**(1), 299 (2023). https://doi.org/10.3390/app13010299
6. Kagimoto, M., Kimura, A., Shibata, F., Tamura, H.: Analysis of tactual impression by audio and visual stimulation for user interface design in mixed reality environment. In: Shumaker, R. (eds.) Virtual and Mixed Reality. VMR 2009. Lecture Notes in Computer Science, vol. 5622, pp. 326–335. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02771-0_37

7. Chamola, V., et al.: Beyond reality: the pivotal role of generative AI in the metaverse. IEEE Internet Things Mag. **7**(3), 126–135 (2024). https://doi.org/10.1109/IOTM.001.2300174
8. Dunne, A., Raby, F.: Speculative Everything: Design, Fiction, and Social Dreaming. MIT Press (2013)
9. Wakkary, R., Odom, W., Hauser, S., Hertz, G., Lin, H.: Material speculation: actual artifacts for critical inquiry. In: Proceedings of the 5th Decennial Aarhus Conference on Critical Alternatives, 17–21 August. Aarhus University and ACM, Aarhus, Denmark (2015). https://doi.org/10.7146/aahcc.v1i1.21299
10. Ndengue, J. D., Juganaru-Mathieu, M., Faucheu, J.: Material perception and material identification in product design. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vancouver, Canada, vol. 8, pp. 429–437 (2017)
11. Kang, N., Sah, Y.J., Lee, S.: Effects of visual and auditory cues on haptic illusions for active and passive touches in mixed reality. Int. J. Hum. Comput. Stud. **150**, 102613 (2021). https://doi.org/10.1016/j.ijhcs.2021.102613
12. Borst, C.W., Volz, R.A.: Evaluation of a haptic mixed reality system for interactions with a virtual control panel. Presence: Teleoperators Virtual Environ. **14**(6), 677–696 (2005). https://doi.org/10.1162/105474605775196594
13. Razza, B.M., Paschoarelli, L.C., Santos, H.M., Andrade, L.O.: The multisensory experience: a case study with five different products. Appl. Hum. Factors Ergon. Int. (2014). https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2108-1
14. Helbig, H.B., Ernst, M.O.: Optimal integration of shape information from vision and touch. Exp. Brain Res. **184**(3), 341–356 (2008). https://doi.org/10.1007/s00221-007-1111-3
15. Biocca, F., Kim, J., Choi, Y.: Visual touch in virtual environments: an exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. Presence Teleoperators Virtual Environ. **10**(3), 247–265 (2001). https://doi.org/10.1162/105474601300343595
16. Ott, R., Thalmann, D., Vexo, F.: Haptic feedback in mixed-reality environment. Vis. Comput. **23**(6), 843–849 (2007). https://doi.org/10.1007/s00371-007-0159-y

Blurring Self-touch Improves Sense of Body Ownership in Incongruence Between VR Avatar and Real User's Body Part

Kodai Hiramatsu(✉), Tomonori Kubota, Satoshi Sato, and Kohei Ogawa

Nagoya University, Nagoya, Aichi, Japan
hiramatsu.kodai.i0@s.mail.nagoya-u.ac.jp,
{kubota,ssato,k-ogawa}@nuee.nagoya-u.ac.jp

Abstract. As an application of Virtual Reality (VR) technology, VR programs have been proposed to simulate on-site experiences, such as practical training, in a virtual environment. To make users feel as if they are actually experiencing these activities in the virtual environment, it is important to design the program to increase the sense of body ownership and presence. However, in practical applications of such VR programs, little is known about the negative effects of using consumer-grade VR devices with limited tracking capabilities on body ownership and presence, and about methods to mitigate these effects. Limitations in tracking capabilities can cause discrepancies in the positional alignment of body parts between the user and their avatar, leading to incongruence between visual information and tactile feedback when users touch their own bodies. Developing methods to mitigate the effects of such inconsistent self-touch is essential for making VR programs accessible at the consumer level. The purpose of this study is to investigate the negative effects of incongruent self-touch on the sense of body ownership and presence and to develop a novel method to mitigate these effects by using visual effects to blur the areas where self-touch occurs. The results of the preliminary experiment suggest that VR devices with low tracking capabilities negatively effect the quality of the VR program experience, and the proposed method utilizing visual effects to mitigate this impact has shown potential effectiveness.

Keywords: virtual reality · avatars · sense of body ownership · presence · self-touch

1 Introduction

The use of Virtual Reality (VR) programs that allow users to engage in virtual environments is gaining attention in fields such as healthcare and education. By leveraging VR technology, it becomes possible to simulate experiences that

K. Hiramatsu and T. Kubota—These authors share first authorship.

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 325–335, 2025.
https://doi.org/10.1007/978-3-031-93715-6_20

are difficult to recreate in real life, such as responding to fires [1], within virtual environments. In the quality of the experience of these VR programs, it is essential to have the sense of body ownership and presence. A practical issue in the use of VR programs is the concern that when consumer-grade VR devices with limited tracking capabilities are employed, incongruences in the positions of body parts between the avatar and the user may arise, potentially leading to a decrease in the user's sense of body ownership and presence. Previous studies have investigated the conditions for generating body ownership and presence from various perspectives. For instance, it has been suggested that factors such as the shape [2] and appearance [3] of the avatar or tactile feedback [4] influence the sense of body ownership. Most of these previous studies assume that the user's body can be accurately tracked, but in practical use, the tracking ability varies depending on the VR device used. To effectively apply VR programs and the knowledge from related research findings at the user level, it is essential to develop methods that maintain the sense of body ownership and presence even when using consumer-grade VR devices.

One of the primary effects of positional mismatch is the incongruence between visual information and tactile feedback during self-touch. Self-touch refers to actions where one moves their own hands or feet to touch their body, including unconscious behaviors such as clasping hands or rubbing legs. Since self-touch naturally occurs during VR program experiences, addressing the effect of such incongruences is a key challenge for improving the quality of VR programs.

In the context of the Rubber Hand Illusion, it has been suggested that passive self-touch, which synchronizes tactile and proprioceptive sensations, is correlated with a sense of body ownership [5]. Additionally, it has been suggested that passive self-touch, which synchronizes tactile and proprioceptive sensations, is correlated with a sense of body ownership [6]. However, these studies assume the synchronization of visual and tactile stimuli, and the effects of self-touch under conditions of visual-tactile incongruence remain unknown.

Therefore, this study aims to achieve the following two objectives:

1. To demonstrate that self-touch under conditions of visual-tactile incongruence decreases the user's sense of body ownership over the avatar and their presence in the VR environment.
2. To develop a method to mitigate the negative effects of incongruent self-touch without enhancing tracking capabilities (i.e., without adding or modifying devices).

To address these objectives, we developed a VR program that allows users to control an avatar using VR devices. Positional incongruences between the user and the avatar were introduced to create conditions of incongruent self-touch. As the proposed method for Objective 2, a visual effect that blurs the area of self-touch on the avatar, as shown in Fig. 1, was developed. This method was applied to avatars performing incongruent self-touch, and its effectiveness was evaluated.

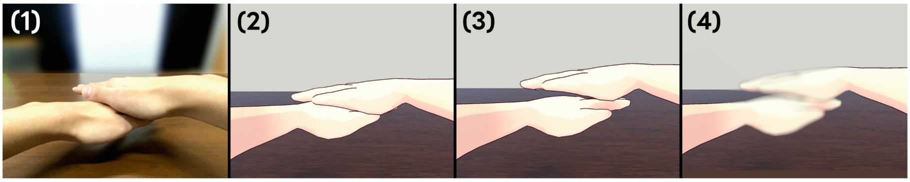


Fig. 1. Images of the proposed method: (1) An example of self-touch performed by the user in the VR program. (2) An avatar correctly reflecting the user's movements. (3) An avatar with tracking misalignment, resulting in incongruence with the user's movements. (4) An avatar with the proposed method applied under the condition of incongruence.

2 Methods

2.1 Objective

The following experiment aimed to investigate the effect of incongruent self-touch on body ownership and presence, and to evaluate the effectiveness of the proposed method.

The research objectives are the two outlined in the introduction. To address Objective 1, we developed a VR program designed to introduce positional incongruences between the user's physical body and their avatar. In the experiment, participants used VR devices to control the avatar's head, hands, and fingers. The experimental task required participants to perform self-touch by moving their right hand to touch their left hand. During the self-touch task, the program manipulated the tracking data obtained from the VR devices to simulate conditions of incongruent self-touch. After experiencing self-touch task, participants' sense of body ownership and presence were assessed using a questionnaire.

To address Objective 2, we developed a method using visual effects that blur the area of self-touch and implemented it in the VR program. For instance, as the right hand approaches the left hand, a blurring effect appears at the point of contact (Fig. 2). To ensure a seamless experience, the intensity of the visual effect dynamically increases as the touched and touching areas get closer, preventing the effect from abruptly appearing or disappearing. We considered that this method suppresses the perception of incongruence while maintaining the overall clarity of the VR environment. Additionally, since it does not rely on the tracking capabilities of the VR devices, it can be integrated into VR programs without requiring additional hardware.

The experiment is based on the following two hypotheses:

1. When self-touch is performed under the condition of visual-tactile incongruence, the sense of body ownership and presence decreases.
2. When self-touch is performed under the condition of visual-tactile incongruence, the proposed visual effect increases the sense of body ownership and presence compared to the condition without the effect.

2.2 Conditions

The experiment compared three conditions:

(a) Without positional incongruence, without visual effect
(b) With Positional incongruence, without visual effect
(c) With Positional incongruence, with visual effect

In the conditions involving positional incongruence ((b) and (c)), the avatar's position was deliberately offset during participants' self-touch actions. For the condition with the visual effect (c), a blurring effect was applied to the area of self-touch on the avatar. (a) corresponds to the situation where the VR program is accurately tracking the user's body, and (b) corresponds to the situation where tracking mismatches that can occur when using consumer-grade VR devices. (c) corresponds to the situation in (b) where the proposed method is applied.

By comparing (a) and (b), the effect of positional incongruence was investigated (Hypothesis 1). By comparing (b) and (c), the effectiveness of the proposed method was evaluated (Hypothesis 2). The experiment employed a within-participants design.

2.3 Experimental Situation

System Setup. A VR program was developed to enable participants to perform self-touch by controlling the avatar's head, hands, and fingers using a VR device (Meta Quest 3). The program applied positional incongruences and visual effects to validate the hypotheses. The program was created using Unity 2022.3.22f1[1].

Positional Incongruence. To replicate a situation where the position of the avatar's right hand is incongruent with the user's actual hand position, the avatar's right hand was offset upward relative to the tracked data. Specifically, the IK target for the right hand was shifted 0.05 m upward in Unity's coordinate space. To enable seamless toggling of positional incongruence, the offset was implemented with a gradual transition. For instance, when enabling positional incongruence, the offset value smoothly increased from 0 to 0.05 m over 0.5 s.

Blurring Effect. To validate the proposed method, a visual effect that blurs the area around the self-touch region was implemented using a shader program (Fig. 2). The effect applied a circular blur centered on the midpoint between the avatar's right and left hands within the participant's view. The intensity and radius of the visual effect were controlled based on the distance between the right and left hands during self-touch. When the right and left hands approached each other, the effect influence gradually increased from zero as the distance decreased. Upon reaching a certain proximity, the effect reached its maximum

[1] https://unity.com/.

influence, and as the hands moved apart, the influence gradually diminished back to zero. To calculate the distance between the hands, the sum of two distances was used:

1. The distance between the left wrist and the palm of the right hand.
2. The distance between the base of the left middle finger and the palm of the right hand.

This approach aimed to approximate the hand's volume for contact detection. Specifically, the control was implemented using an ellipsoid-shaped collision detection model.

Additionally, the blurring effect is always rendered in front of the avatar, regardless of their positioning.

Fig. 2. Self-touch performed with the proposed visual effect applied (first-person perspective).

Avatar. The avatar shown in Fig. 3, was operated by the participants in the experiment.

The avatar was created using VRoid[2] and configured in a format compatible with Unity's Humanoid Avatar system, enabling control of the hands, head, fingers, and other body parts. To ensure that the hands and forearms used for self-touch were not obscured, the avatar was dressed in short-sleeved clothing. The avatar's height was set at 170 cm, and participants, regardless of their own height, operated avatars of this uniform size.

[2] https://vroid.com/.

Fig. 3. The appearance of the avatar operated by participants during the experiment.

Virtual Environment. Since participants performed self-touch with their hands resting on a desk, a corresponding desk object was placed in the virtual environment at the same position as the real desk. This setup minimized discrepancies in tactile feedback caused by contact with the desk, reducing potential discomfort or inconsistencies between the real and virtual environments.

As the surrounding environment for the task, a virtual representation of a typical room interior, as shown in Fig. 4, was displayed in the virtual environment. This design choice aimed to mitigate the potential for participants to feel strangeness or discomfort in an overly simplistic virtual environment. To ensure that the presence of these interior elements did not influence the experimental results, objects unrelated to the task were positioned out of the participants' reach, preventing any unintended interaction.

To help participants recognize the relationship between their movements and the avatar's actions, a mirror was placed in front of them. The mirror allowed participants to observe the avatar's upper-body movements and self-touch actions. However, the lower body, including the untracked feet and waist, was intentionally excluded from the reflected view to maintain focus on the tracked areas.

Experimental instructions during the task were displayed as text messages on a message board placed in front of the participant in the virtual environment. This setup was used to signal the start and end of the self-touch task, accompanied by sound effects and text-based cues. Providing instructions through objects within the virtual environment avoided breaking participants' presence. Verbal instructions from the real world could draw participants' attention away from the virtual environment, potentially impacting their sense of presence. Using virtual objects to deliver instructions minimized external interference with the VR experience.

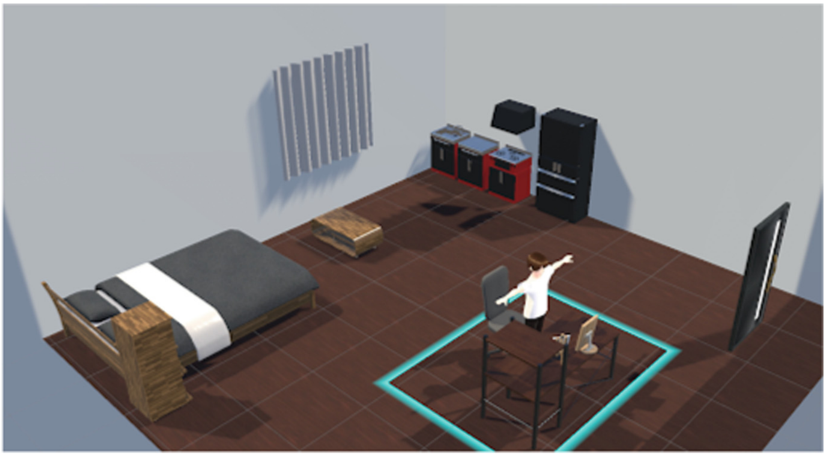


Fig. 4. Overview of the experimental environment in the virtual environment.

Devices and Software. The experiment used the Meta Quest 3[3] as the VR device, which handled virtual environment rendering and tracked the head, hands, and fingers.

The VR program was developed using Unity and executed on a computer equipped with standard graphics capabilities. The system used an Intel Core i7-13700 CPU and an NVIDIA GeForce RTX 4070 GPU. Additionally, VirtualDesktop was employed to establish a wireless connection between the Quest 3 and the PC, enabling the transmission of tracking data and the streaming of the virtual environment.

To align the Quest 3 tracking space with the Unity's coordinate space, a calibration process was performed at the start of the task. During calibration, participants sat in a real chair and rested both hands flat on the desk. In this position, the calibration process adjusted the Quest 3's rendered viewpoint to align with the avatar's eye level. Additionally, the desk object in the virtual environment was repositioned to match the avatar's hand placement. This ensured consistency between the layouts of the real-world and virtual experimental environments.

2.4 Tasks in the Experiment

Self-Touch Task. In this task, participants are instructed to move their right hand to make contact with their left hand. The participant (Avatar) performed the task while seated. The task involved using the right hand to touch the back of the left hand from above, repeating the motion at a pace of approximately 1.5 times per second. Instructions on the touching way and pace were provided beforehand.

[3] https://www.meta.com/jp/quest/quest-3/.

To prevent tracking issues caused by the right hand covering the left hand, the position and rotation of the avatar's left hand were fixed through processing within the VR program during the task. Participants were also instructed to keep their left hand resting on the desk without moving it. Even with the left hand fixed, the fingers remained movable. This prevented the complete loss of visual feedback for the left hand's movement.

Before and after the task, during the donning and removal of the headset, the participant's view gradually faded to black and brightened. This transition was designed to clearly delineate the shift between the VR environment and the real world. Additionally, avatar calibration (aligning the avatar's coordinates in the VR device's tracking space with those in the Unity coordinate space) was performed during the fade-to-black phase. This approach avoided sudden changes in the participant's viewpoint, aiming to reduce the likelihood of VR-induced motion sickness.

2.5 Measures

The evaluation was conducted using the questionnaire. The questionnaire was designed based on prior research [7–9]. Each item in the questionnaire was designed using a 7-point Likert Scale. Q1,2 are ownership-related questions, while Q3-6 pertain to presence. Q6 is a reverse-coded item.

The questions used are as follows:

- Q1 I felt as though the arms of the avatar visible in the virtual environment were my own.
- Q2 I felt as though the arms of the avatar reflected in the mirror within the virtual environment were my own.
- Q3 I felt that I existed within the virtual environment.
- Q4 During the experiment, there were moments when the virtual environment became my reality, and I forgot about the real world.
- Q5 I felt that the virtual environment I experienced during the experiment was not something I "saw" but rather a "place I visited".
- Q6 During the experiment, there were moments when I felt that I was not present in the virtual environment, but merely wearing a headset while sitting on a real chair.

(0: not at all - 6: very much)

The metrics for body ownership and presence were calculated as the average scores of their respective related questionnaire items. The participant's body ownership score in each condition is the value of the average of the Q1 and Q2 ratings. The presence score is the average of Q3-Q6. Scores ranged from 0 to 6.

2.6 Procedure

At the beginning of the experiment, participants were briefed on the overall procedure and important precautions, such as the risk of VR-induced motion sickness. During the explanation of the task, the experimenter demonstrated the self-touch motion to instruct participants on the proper hand movements and the pace of touching. To minimize the influence of factors other than self-touch on the experimental results, participants were instructed not to move around the virtual environment or interact with any objects apart from performing the self-touch task.

The task was performed across three blocks corresponding to the conditions described as (a), (b), and (c). In each block, participants completed the self-touch task, followed by answering a questionnaire about their experience under that condition.

The tasks and the instructions in each block are as follows:

1. The participant wears the VR headset, sits on a chair, and places both hands on the desk. At this point, the VR view remains blacked out.
2. The avatar and virtual environment are calibrated, and once calibration is complete, the blackout in the participant's view is lifted.
3. The instructions to start the task are delivered via the message board in the virtual space, and the participant performs self-touch for two minutes.
4. After two minutes, the closing instruction is displayed on the message board, and the participant's view fades to black.
5. The participant removes the VR headset and completes the questionnaire.
6. After answering the questionnaire, the participant's condition is checked to ensure there are no issues before proceeding to the next block. A minimum interval of two minutes is maintained before transitioning to the next block.

2.7 Participants

Three participants from inside the lab, all male in their 20s, took part in the experiment. In the experiment, two participants performed the tasks in the order of (a), (b), and (c), while one participant performed them in reverse order, (c), (b), and (a).

3 Results

The body ownership and presence scores for each participant under each condition are shown in Tables 1 and 2. In the positional incongruence condition (b), body ownership scores decreased by 43%, and presence scores decreased by 4% compared to the no positional incongruence condition(a). In the visual effect condition(c), body ownership scores increased by 38%, while presence scores decreased by 5% compared to the no visual effect condition(b).

Table 1. The sense of body ownership scores for each participant under each condition, and the participant-averaged scores.

Condition	Participant 1	Participant 2	Participant 3	Average Score
a	3	6	2.5	3.83
b	3.5	2	1	2.17
c	2	5	2	3

Table 2. The presence scores for each participant under each condition, and the participant-averaged scores.

Condition	Participant 1	Participant 2	Participant 3	Average Score
a	3.75	4.5	3.25	3.83
b	4	3.5	3.5	3.67
c	1.75	4.75	4	3.5

4 Discussion

The results of the preliminary experiment suggest the potential for future findings to support Hypotheses 1 and 2. For Objective 1, it was confirmed that self-touch under conditions of visual-tactile incongruence between the user and the avatar may decrease body ownership and presence. For Objective 2, it was confirmed that using the proposed method in this study may mitigate the sense of body ownership during self-touch under conditions of visual-tactile incongruence. However, regarding the sense of presence, the score decreased by 5% in the visual effect condition (c) compared to the no visual effect condition (b). Since this result does not align with Hypothesis 2, we aim to increase the number of participants in future studies to clarify the findings.

In this study, we developed a novel method utilizing visual effects to mitigate the effect of visual-tactile incongruence on the sense of body ownership and presence. A preliminary experiment was conducted, and the potential effectiveness of this method was confirmed. While the proposed method showed promise, the preliminary experiment involved a small number of participants, and some occasional issues with the experimental system, such as tracking malfunctions, were observed. Future research will conduct experiments with a larger sample size to further explore the effects of incongruent self-touch with VR avatars and evaluate the effectiveness of the proposed method.

Acknowledgements. This work was supported by JST Moonshot R&D Grant No. JPMJMS2011.

References

1. Congès, A., Evain, A., Benaben, F., Chabiron, O., Rebière, S.: Crisis management exercises in virtual reality. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 87–92, March 2020
2. Argelaguet, F., Hoyet, L., Trico, M., Lecuyer, A.: The role of interaction in virtual embodiment: effects of the virtual hand representation. In: 2016 IEEE Virtual Reality (VR), pp. 3–10, March 2016
3. Schwind, V., Lin, L., Di Luca, M., Jörg, S., Hillis, J.: Touch with foreign hands: the effect of virtual hand appearance on visual-haptic integration. In: Proceedings of the 15th ACM Symposium on Applied Perception, pp. 1–8, Vancouver British Columbia Canada, August 2018
4. Alimardani, M., Nishio, S., Ishiguro, H.: Humanlike robot hands controlled by brain activity arouse illusion of ownership in operators. Sci. Rep. **3**(1), 2396 (2013)
5. Henrik Ehrsson, H., Holmes, N.P., Passingham, R.E.: Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. J. Neurosci. Off. J. Soc. Neurosci. **25**(45), 10564–10573 (2005)
6. Hara, M., et al.: Voluntary self-touch increases body ownership. Front. Psychol. **6**, 1509 (2015)
7. Steed, A., et al.: An 'In the Wild' experiment on presence and embodiment using consumer virtual reality equipment. IEEE Trans. Visual Comput. Graphics **22**(4), 1406–1414 (2016)
8. Slater, M., Usoh, M., Steed, A.: Depth of presence in virtual environments. Presence **3**, 130–144 (1994)
9. Gonzalez-Franco, M., Peck, T.C.: Avatar embodiment. Towards a standardized questionnaire. Front. Robot. AI **5**, 74 (2018)

Embodying a Mixed-Reality Agent with a Wearable Snake-Shaped Robotic Appendage

Abdullah Iskandar[1,2](✉), Hala Khazer Shebli Aburajouh[3], Haya Al Abdullah[3], Roudha Al Yafei[3], Noor Al Wadaani[3], Osama Halabi[3], Mohammed Al-Sada[3,4], and Tatsuo Nakajima[1]

[1] Computer Science and Engineering, Waseda University, Tokyo, Japan
{a.iskandar,tatsuo}@dcl.cs.waseda.ac.jp,
abdullahiskandar@telkomuniversity.ac.id

[2] School of Computing, Telkom University, Bandung, Indonesia

[3] Department of Computer Science and Engineering, Qatar University, Doha, Qatar
{ha1906848,hayaalabdullah,roudhaalyafei,nooralwadaani,ohalabi,
mohammed.alsada}@qu.edu.qa

[4] KINDI Computing Research Center, Qatar University, Doha, Qatar

Abstract. Recent advances in MR and AI have paved the way for numerous applications of virtual agents. However, a pressing challenge lies in the interaction with such agents, as they lack physical embodiment, thereby being unable to interact with the real world. Accordingly, we explore the use of a wrist-worn snake-shaped robotic to embody a virtual MR agent. Our implementation comprises an MR agent implemented on an HMD, which is connected to the robotic appendage. We implemented two experiences using our system, a game and an object-finding task. To explore the effects of embodiment, we conducted a comparative study focusing on evaluating the effects of robot embodiment in the two scenarios. Results show that using full embodiment is effective in increasing immersion and realism of the MR agent when interacting directly with it, while an embodiment is not necessary in contexts when visual feedback is sufficient. We discuss the results and future work direction in light of our results.

Keywords: Mixed Reality · wearable robot · avatar robot

1 Introduction

The usage of robots in daily life has grown over the past few years, changing how we work, live, and interact with the world. Wearable robotic appendages present a vision of robots for daily use [2,11,36], which will play an important role in supporting our daily activities, such as recreational sports, household chores, and providing companionship. However, interacting and controlling such robots is an ongoing research challenge [2,12,27,36].

The complexity and wide variety of application domains make designing manual controls of wearable robots, such as by using a joystick, a complex, mentally and physically demanding task. Furthermore, such devices are expected to

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 336–355, 2025.
https://doi.org/10.1007/978-3-031-93715-6_21

support users with a high degree of autonomy and intelligence [2]. Therefore, previous research has concluded that a critical interaction paradigm for interacting with wearable robotic appendages is to interact with them as companions or agents; which provide anthropomorphic interaction modalities, and enable a highly intelligent and autonomous human-like interaction experience [5,8,26]. Such a finding is especially critical for daily use, where non-expert users are using such systems to carry-out tasks of daily living [4]. However, despite its criticality, interacting with wearable robotic appendages using the agent paradigm has been largely unexplored in previous research.

Although wearable robotic appendages can be versatile, most of such appendages lack interaction modalities and fidelity to establish an effective user experience. However, recent advancements in Mixed Reality (MR) and Head-Mounted Displays (HMDs) have shown that various types of robots can be augmented with visualizations to enhance the interaction experience, such as to utilize I/O modalities in the HMD to interact with the robot, or visualizations to construct a user experience [28]. Therefore, we explore using MR HMDs for developing and evaluating a user experience to interact with a wearable robot using an MR agent, exploring a highly anthropomorphic user experience that is intuitive for the daily user.

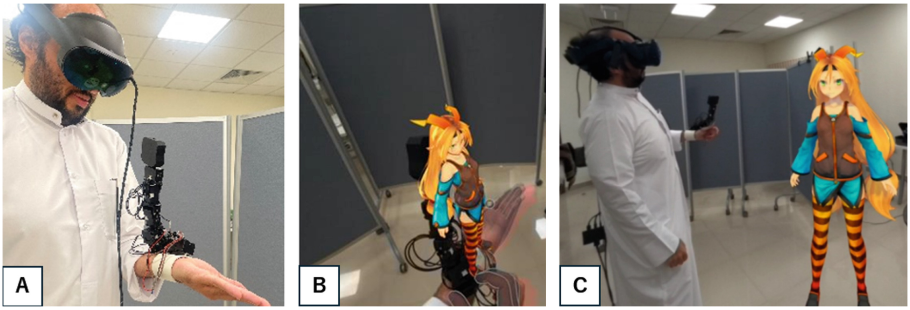


Fig. 1. (A) We present a system comprising a wrist-worn robotic appendage and an MR HMD. (B) Viewed from the HMD, The wearable robot can be used to embody a virtual character, offering novel multi modal interactive experience and enhancing the sense of presence of the character on the user's wrist. (C) The MR character can be presented in different sizes and locations around the user, where the character and the robot provide intriguing user experience within daily usage contexts.

The primary objective of this research is to develop a system comprising a wearable robot that is complemented by an MR agent, which can comprise a number of user experiences and agent embodiment methods using the wearable robot (as shown in Fig. 1). Furthermore, we use our system to design two main interaction scenarios Rock-Paper Scissor game (Sect. 5.1) and Object or Location Finding (Sect. 5.2). We proceed to evaluate our system through the two implemented scenarios, with emphasis on embodiment and usability factors related to daily use. The result shows that using full embodiment is effective in increasing immersion and realism of the MR agent when interacting directly with it, while an embodiment is not necessary in contexts when visual feedback is sufficient. Through these interaction scenarios and evaluations, it shows that participants preferred agent-only interactions over robot-augmented conditions due to greater simplicity, comfort, and ease of interaction, highlighting the need to address ergonomic and cognitive challenges when integrating physical robotics into MR environments, especially regarding weight and stability concerns.

We summarize the contributions of this paper as follows:

- Design and implementation of an anthropomorphic user experience with an MR-based agent capable of embodying and controlling a wrist-worn wearable robotic appendage.
- Contribute with evaluation based on a comparative study to measure effects of embodiment and interaction effectiveness of using a wearable robotic appendage to embody an MR agent.
- Contribute with insights and future work directions for realizing MR agents embodied by wearable robotic appendages.

2 Related Works

Our research expands two strands of related works; *wearable robotic appendages* and *virtual agents*. We explain each of these domains below as follows:

Wearable robotic appendages refer to robotic systems that can be worn and used in a variety of interactive contexts. For example, Orochi [2] is a serpentine-shaped robot integrated into a scarf hands-free communication interactions with both the real world and digital world. Similarly, LineFORM and ChainForm [20,21] is shape-shifting snake-shaped robot that can be used for a variety of interactions, such as a tangible user interface or for dynamic body constraints like an exoskeleton. A prominent challenge of wearable robotic appendages is interaction and control [33], therefore, weARable was a system that used MR to control a wrist-worn robotic appendage [33]. They explored various paradigms of interaction with a robotic appendage, such with direct commands, or through an MR character agent.

Virtual agents has long been explored in research works, which mainly refers to a systems that uses predefined rules or artificial intelligence (AI) to offer automated assistance or guidance to users through verbal or textual conversations [9]. For example, Teacher Avatar [35] is four virtual avatars used for educational field trips using a humanoid models akin to those in video games virtual guides in VR exhibitions [25] is a humanoid virtual agent to make VR exhibition visitors experience time-independent guided tours. Similarly, avatars, whether embodying users or systems, have been explored within robotics by using MR to overlay the avatar on the robot. For example, Jones et al. and Tajwani et al. [13,31] presented a teleoperated robotic system that allows remote users, wearing an MR HMD, to view the telepresence robot with an MR overlay of the operator on the robot. Similarly, Ihara et al. [10] explored using MR to overlay user's volumetric video feed in a remote environment and embodying the user's hands using miniature mobile robots. Therefore, remote users can visualize the local user's body and sense their physical interactions embodied by the robots.

Our work advances the state-of-the-art by exploring and evaluating embodiment and interaction with MR Agent embodied using multipurpose wearable robots. Our literature survey shows that these aspects have not been explored in previous research [28,34], especially aspects of cross-device interactions and effects of virtual-agent embodiment on wearable robotic appendages. These aspects are critical for designing suitable embodiment and visualization methods of MR agents with wearable robotic appendages.

3 System Design and Architecture

Our main design objective is to explore cross-device interactions and embodiment aspects of an MR agent using a snake-shaped robotic appendage. Accordingly, our system was designed and implemented to enable us to create a variety of user experiences.

Overall, we designed our system based on client-server connections between an HMD and a robot system, which enable us to experiment with different MR HMDs, interaction methods, and robot configurations without the need of major system modifications. The high-level architecture of our system is shown in Fig. 2, which comprises a number of systems, namely the *HMD system* and *Robot System*. Each of these systems are explained in the below subsections.

3.1 HMD System

The HMD system is mainly responsible for displaying and interacting with the virtual environment, and controlling the robot in response to the agent or events in the MR system. The HMD system is implemented on the Meta Quest Pro due to its versatile MR capabilities and built-in interaction methods [17]. This system comprises the following subsystems:
Agent subsystem: which oversees the behavior and decision-making of MR agent, and includes the animator, which controls animations and visual effects of the agent, and the Motion Synchronizer system, which controls the robot physical motion based on the agent's actions.
Input/Output subsystem: which mainly controls interactivity with the MR agent, such as using voice commands, auditory feedback, or hand-gestures.
Application system: which comprises application-level logic and utilizes the other subsystems to enable creating various interactive experiences merging the MR agent and robot, which we utilize to implement our interaction scenarios.

3.2 Robot System:

This system is used mainly to control the robot's motion. We implemented this system using a client-server architecture based on websockets between the HMD system and the robot system [2,15].

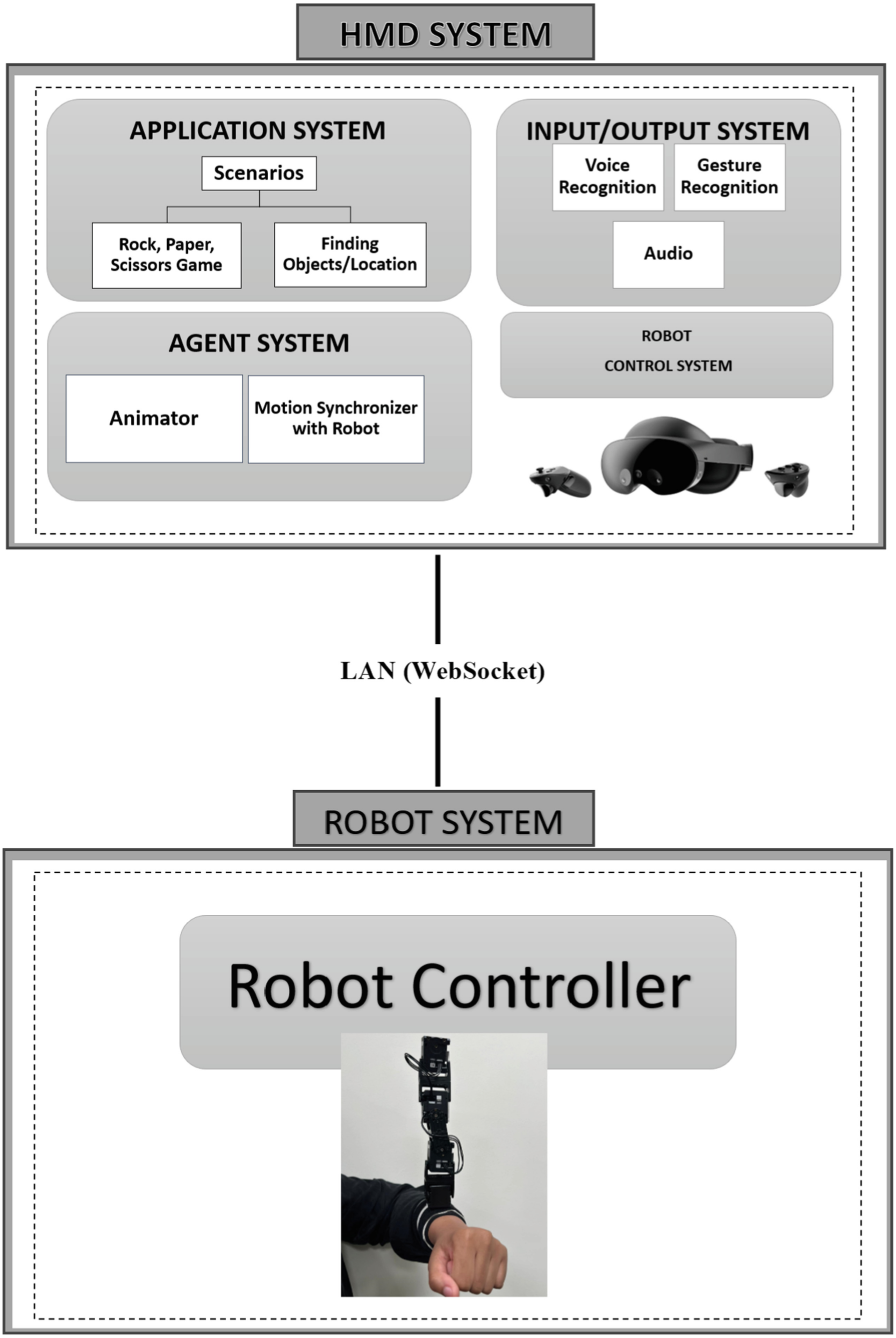


Fig. 2. High-level architecture of the design.

4 System Implementation

4.1 HMD System

The HMD System is implemented using the Meta Quest Pro HMD [17], which was chosen due to its versatility and variety of interaction methods. The software system is implemented using Unity game engine due to its compatibility with Meta Quest Pro [30]. Accordingly, all the subsystems has been implemented using C#, as components, and game-objects within Unity game engine environment.

Agent System: We used Unity-Chan character which was developed by Unity [29]. Unity-Chan has various auditory feedback and animations, such as walk, scissor, hands up, and others. Therefore, we used various existing animations and auditory feedback to create the virtual agent experience, which were developed within the Unity game engine using its animator. Moreover, within our agent system, we have implemented a motion synchronizer system that facilitates communication between the agent and the robot. This enables the robot to receive movement commands corresponding to the current actions of the agent.
Input/Output System: this system comprises a number of subsystems that are used to provide interaction modalities with the virtual agent, namely using voice commands and gestures.
Voice Commands system: was implemented using the Meta Quest Text to Speech (TTS) SDK powered by Wit.ai [18]. This system allows us to convert spoken voice into text, after which they are sent over the network to wit.ai web-service that is used to infer meaning from various void commands, which then can be converted to agent/robot actions. We used the microphone within the Meta Quest Pro to capture users' voice commands.

Gesture Recognition System: A Meta Quest has a hand tracking system feature, which we used as basis for detecting both robot hands and various hand gestures using their Pose Detection SDK [16]. We use the Hand tracking feature to place the MR agent on the user's wrist exactly above the wearable robot.

4.2 Robot Design and Control System

The *robot interfacing and Control System* was designed with a client-server architecture for controlling the robot, based on websockets, enabling real-time, bidirectional communication with the robot. Our client-server system is based on a publisher-subscriber model [2], with two websocket services implemented for setting up the robot (setting servomotor speed, acceleration, angular limits, etc.), controlling the robot using position control, and reading various feedback from the robot (e.g., current servomotor angles, applied torque, temperature). Our system is implemented using C# and integrates with the Robotis [24] SDK, to directly control the robot over serial-USB. Overall, this architecture provides a robust and flexible method of controlling the robot using a variety of input methods.

We designed a snake-shaped robotic appendage that is wrist-worn, which is used to create a variety of interaction and embodiment potentials with the MR Agent. The robot weighs 328g and 34cm in length (including the base). The robot comprises 7 DoFs, and was implemented using five XC330-T288-T servomotors and one 2XL430-W250 servo motor, which were configured to achieve high redundancy in a snake-shaped form factor (as shown in Fig. 3). The servomotors are connected together using plastic frames (PLA), and are daisy-chained and controlled over TTL. The robot was fitted on a base that can be worn on the user's wrist, which was molded from thermoplastic material and included a 3D printed frame to fit the robot. The base is padded with a thin layer of foam to ensure comfort when worn, and is secured using Velcro around the user's wrist.

We utilized robot control software based on previous research [2,12]. Robot control software is based on the network model and a publisher-subscriber model using the Web Socket [33]. The data between robot and computer was sent and received using a JSON packet. Robot movements are pre-recorded and play back with the predetermined motions of the character.

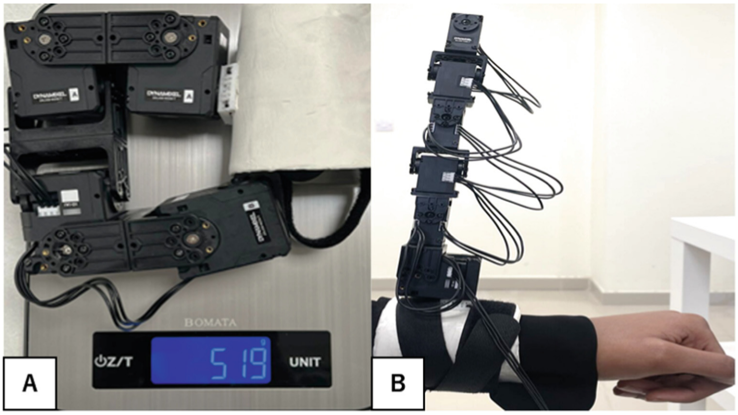


Fig. 3. A) Robot structure; B) Robot is wrist-worn by the user and fully extended upwards.

5 Interactive Scenarios

We implemented two main scenarios that demonstrated how cross-device experiences can be created using an MR agent and a wearable robot. Based on our described implementation in Sect. 4, we developed two interactive scenarios explained as follows.

5.1 Scenario 1: Rock-Paper-Scissors Game

Scenario 1 comprised the casual game of Rock, Paper, Scissors. We chose this scenario as it combines various interaction modalities and agent animations.

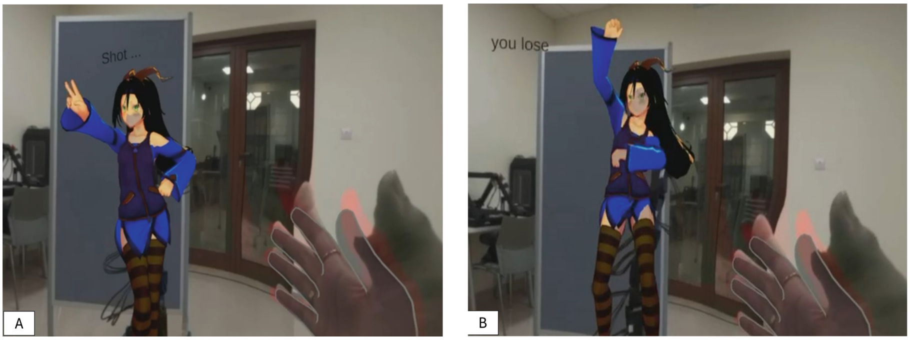


Fig. 4. Scenario 1, which is rock-paper-scissors game. A) Shows the user playing the game with the AR Agent, B) shows the agent reacting to beating the user.

The scenarios start with the user's saying "start" verbal command to initiate the game shown in Fig. 4. Then, the MR Agent would start saying "rock, paper, scissors", and users have to make one of the three hand gesture of the game. Based on the results, the MR agent would either say "I won" or "I lost", with matching celebration/losing animations and movements by the robot according to the character's movements (as shown in 4). If the user ties with the MR agent, the agent performs a neutral animation. The results are also shown as text to the user for confirming the result of the game. Upon finishing, the game repeats again. Users can end the game by saying "End".

The robot motions were generated, played back and synchronized with the MR agent's motion. We recorded various robot motions that included leaning forward/backward and rotations at various speeds and accelerations so that the motions and generated inertia portray a sense of the physical existence of the agent on the user's wrist. When the MR agent moves, each animation by the character triggers a matching motion on the robot matching its body motions. Therefore, users could feel the MR agent's movements through the wrist-worn robot, which was synchronized with the MR agent's motions.

5.2 Scenario 2: Object/Location Finding Scenario

The object/location finding scenario is an experience in which the MR agent assists users in finding objects or locations in the surrounding environment. The experience starts with the user saying "Help". Then, the MR agent asks, "What are you looking for?", then the user can respond by asking for a specific object or location. Upon answering, the agent would walk to the desired object/location and say "Here it is", and the robot would point at the same direction as the object/location as shown Fig. 5. Due to restrictions to access the HMD's camera-feed for conducting real-time object recognition (Further discussed in future work), and for evaluation purposes, we implemented two pre-programmed objects (tools, snacks) and one location (exit door). Unlike scenario

1, scenario 2 uses a human-sized MR agent that navigates the surrounding environment using the localization system of Quest pro. Scenario 2 is implemented in this manner to evaluate effectiveness of the robot at pointing to the desired object/location in conjunction with the MR agent.

Based on the directions of the objects or locations, we implemented a robot animation at the HMD system for pointing the robot to a specific direction similar to a compass. The robot motion is triggered when the MR agent approaches the desired object/location (Fig. 5.B), and was configured based on the directions of each of the objects/locations (as shown in Fig. 5).

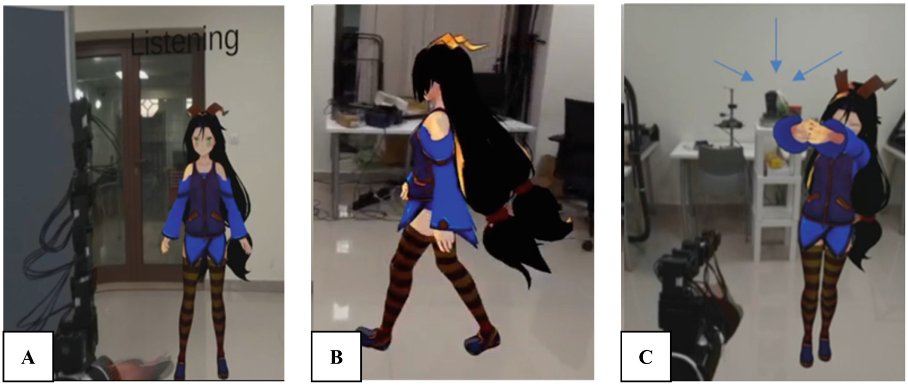


Fig. 5. Scenario 2, object/location finding. A) The MR Agent waiting for user to ask about an object or location to find, upon the user asking about a location/object, B) The MR agent leads the user towards the object/location by first walking to it, then C) pointing and saying where it is, the robot also points to the location of the objective.

6 Evaluation

This section covers our evaluation procedure with subsections covering the user study design, participants scenarios and conditions, and flow as follows:

6.1 User Study Design

This user study aims to evaluate the interaction between the users and an MR agent with a wearable robotic appendage. The objectives of our study are:

A. Investigate the effectiveness of the robot in embodying the MR agent and enhancing the interaction experience.
B. Measurement of impressions and opinions of the system.

Accordingly, our study includes four conditions with two independent variables:

1. MR Agent Embodiment: whether the robot is used to embody the MR agent or provide visual guidance with the agent or not.

2. Task Type: which included two tasks based on the developed experiences, which are rock-paper-scissors and locating an object in the surrounding environment.

Our study includes two dependent variables:

1- Factors of physical embodiment, enjoyability and helpfullness of the system under tasks and embodiments.

2- Users impressions and preferences.

Overall, the outcomes of this study are essential for understanding the interaction dynamics between users and a MR agent and wearable robotic appendages, especially from the perspective of embodiment and it's perceived usefullness in various interactive scenarios.

6.2 Participants:

We hired 12 participants, aged 20 to 23, who come from diverse backgrounds and majors and universities. All participants were women. 7 participants reported having briefly used VR before, while the remaining 5 have never used VR.

6.3 Scenarios and Conditions:

Our user study comprises two scenarios (Explained in Sects. 5.1 and 5.2) which are as follows:

1) Rock Paper Scissors game (S1): We compare playing the game in two conditions. The first condition is playing the game with the MR agent embodied by the robot (*S1AR*), and the second condition is playing the game with the MR agent only (*S1A*). To balance potential learning effects, half of the participants started with (S1AR) then (S1A), the other half in reverse manner.

2) Finding object/location (S2): This task requires participants to compare two conditions. The first condition is to locate objects with the MR Agent embodied by the robot*S2AR*, and the second condition is using the MR agent only *S2A*. To avoid potential learning effects, half of the participants started with S2A followed by the S2AR, while the other were in reverse order.

6.4 Flow

First, participants were briefed about the experiment, and had a 5-minute familiarization to try the system. Then, participants started S1 in a random condition, followed by a short questionnaire about their enjoyment and impressions. Then, participants completed the other condition, followed by the same short questionnaire. Lastly, they took a questionnaire comparing the conditions of S1, followed by a 10-minute break. S2 started with a random condition, and users were instructed to try the scenario to find the tools, snacks and exit door (as explained in Sect. 4.2). Then, participants took a short questionnaire that measured the perceived helpfulness of the system. Next, participants tried the other conditions, followed by the same questionnaire. Lastly, each participant had a debriefing session including a semi-structured interview questions and a questionnaire about the overall impressions of the system, which included the system usability questionnaire (SUS) [1,6]. Each study lasted for approx. 60 min per participant (Fig. 6).

7 Results

In this section, we discuss the results of the two scenarios performed by the participants and the overall experiment.

7.1 Results of Scenario 1 (S1)

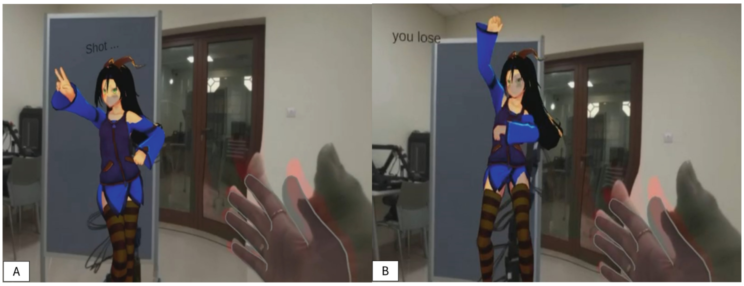


Fig. 6. S1 Enjoyment questionnaire results.

Scenario 1 was designed to evaluate the embodiment of the MR agent by the robot. The results of the questions are shown in Fig. 7. The first question (**Q1**) measured how enjoyable the game experience was in *S1A*, where participants thought it was enjoyable ($M = 5.67, SD = 0.65$). Users reported that they appreciated the agent's stability on the arm, describing it as weightless and effortless to interact with. The second question (**Q2**) measured how enjoyable the game experience was in *S1AR*. The results showed a lower enjoyment rating ($M = 4.83, SD = 1.19$) compared to *S1A*. The inclusion of the robot in *S1AR* was noted to require more effort, which may have reduced user enjoyment. A paired t-test was conducted to compare the enjoyment ratings between the two conditions, yielding significant results ($t(11) = -2.59, p = 0.025$), confirming that participants significantly preferred *S1A* over *S1AR*.

The results of questions three to seven for S1 are shown in Fig. 7. According to Q3, 'I feel that the robot made the experience more enjoyable', the robot made the experience more enjoyable (M = 4.25, SD = 1.14). Q4, 'To what extent

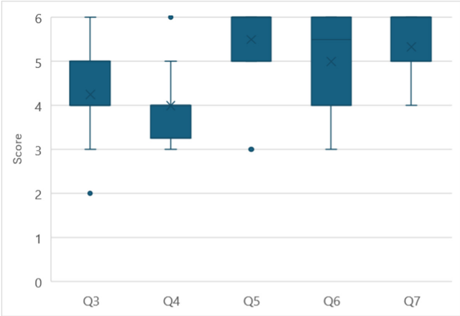


Fig. 7. S1 Questionnaire results of Q3-Q7.

did the presence of the wearable robot enhance your overall gaming experience?', scored (M = 4.00, SD = 0.85), indicating that people thought that the wearable robot improved their gaming experience. In Q5, 'How satisfied were you with the movement-synchronization between the MR-agent and robot during game-play?', scored (M = 5.50, SD = 0.90), indicating high satisfaction with the synchronization of the wearable robot and the AR agent during the game, facilitating the game experience. In Q6, 'I felt the character is physically moving on my arm', participants rated the physical movement of the character in their arm with (M = 5, SD = 1.2). This result indicates that most participants felt a strong sense of the character physically moving on their arms during the scenario. In addition, Q7, 'How well did the wearable robot and the AR agent work together to provide feedback during the game?', has (M = 5.30, SD = 0.98), where participants rated how well the wearable robot and the AR agent worked together to provide feedback during the game.

When ranking the conditions according to the enjoyment of the S1 experience from most to second-most liked, *S1A* resulted in (M = 1.58, SD = 0.51), followed by *S1AR* with (M = 1.42, SD = 0.51). This result is justified by participants indicating that the presence of the robot slightly affected the experience of the game due to its weight (Further discussed in Sect. 6.3, Q3), which made it feel unstable on the arm during game play. However, we conducted statistical analysis and it did not yield significant results.

7.2 Results of Scenario 2 (S2)

In the first two questions, participants rated how helpful the conditions were in the object/location finding scenario. The results of the questions are shown in Fig. 8.

For the first question (**Q3**), which evaluated the helpfulness of the *S2A* condition, the results showed that participants rated it as moderately helpful ($M = 5.33, SD = 1.23$). For the second question (**Q4**), which evaluated the helpfulness of the *S2AR* condition, the results showed a slightly lower rating

$(M = 5.17, SD = 1.34)$. These results are illustrated in Fig. 8. A paired-sample t-test was conducted to compare the two conditions. The results indicated no significant difference in the helpfulness ratings between *S2A* and *S2AR* $(t(11) = 0.352, p = 0.732)$.

For the third question (**Q5**), "Which one will you choose in real life?" participants expressed a clear preference for the *S2A* condition, with 8 participants selecting Agent only for the object/location finding scenario.

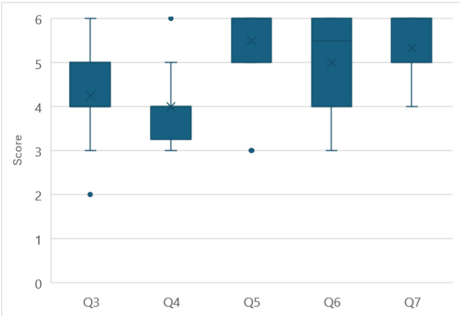


Fig. 8. S2 Helpfulness Questionnaire results.

According to the last two questions, when ranking the experience from most helpful to least helpful, where 1 indicates the favorite, the participants ranked the conditions as follows: *S2A* $(M = 1.5, SD = 0.90)$ and *S2AR* $(M = 2.25, SD = 0.75)$. Statistical analysis did not yield significant effects between the two conditions.

7.3 Overall Experience Results

According to the results in Fig. 9, We asked questions about Q1 'How satisfied were you with the overall interaction experience? in Rock, Paper, Scissor (S1AR - Agent with robot)', and Q2 'How satisfied were you with the overall interaction experience? in Object/Location finder (S2AR - Agent with robot)'. The result showed that the participants found the interaction experience with the Agent and Robot conditions satisfying in both S1 and S2. The average rating was (m = 4.92, SD = 1.08) in both scenarios. We performed a paired samples t-test between the conditions, yet results showed no significant effects (t(11) = 0.00, p = 1.00).

Q3 'How would you rate the weight of the wearable robot? [6 is very light]' shows that the wearable robot was slightly heavy (m = 3.83, SD = 0.83). The results of Q4 'I think the wearable robot was safe [6 is very safe]' indicates that participants thought the robot was safe (m = 5.42, SD = 1.24). Q5 'I felt that the MR agent was real [6 is strongly agree]' scored (m = 4.83, SD = 0.94) indicating indicates moderate embodiment effect.

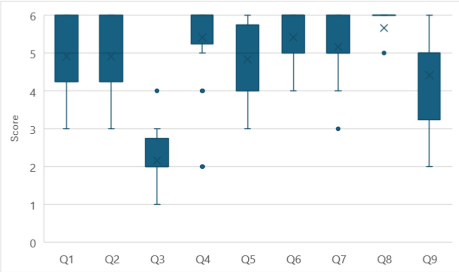


Fig. 9. Post-Study Questionnaire results.

Q6 'I think the system of the wearable robot and the MR agent was friendly [6 is friendly]', was rated with (m = 5.2, SD = 0.79). Q7 'I think the wearable robot and MR agent system was responsive [6 is responsive]' showed that the system was mostly responsive with a score (m = 5.17, SD = 0.94). Q8 'I think the wearable robot and MR agent system was exciting [6 is exciting]' resulted in (m = 5.67, SD = 0.89). Finally, in Q9 'I think the system of the wearable robot and the MR agent was life-like [6 is life-like]', the results were (m = 4.42, SD = 1.24).

According to the System Usability Scale (SUS) [14], the system received a score of 76.8, corresponding to a 'C' grade. The scale indicates a moderate level of usability for the system, which is expected for an early prototype using novel technologies that have not yet been adopted by end users.

8 Discussion

In this section, we discuss the results of each scenario and discuss the overall results as follows:

8.1 Scenario 1 (S1)

In S1, the preference for the agent-only condition (S1A) over the robot-augmented condition (S1AR) highlights a significant finding regarding user enjoyment. Participants consistently rated the agent-only interaction as more enjoyable, citing reasons such as the slow response of the wearable robot, heaviness, and instability of the wearable robot, which leads to users being distracted by the double directing from the agent and robot. This preference aligns with previous research that emphasizes the importance of user comfort and seamless interaction in immersive environments [3,11,12]. However, despite the significance difference between two conditions in enjoyment, the overall score for the S1AR condition was reported to be nearly 5 points, indicating a high level of participant satisfaction with the enjoyment derived from using this condition.

Further results also indicate that the participants expressed high satisfaction with the synchronization between the wearable robot and the MR agent during the game-play, and that the robot conveyed the sense of physical presence of the MR-agent on their arms. This positive feedback indicates that while the robot's weight may have slightly diminished enjoyment, it effectively facilitated the sense of embodiment for the MR agent on users wrists. This dual perspective underscores the complexity of integrating robotics into MR scenarios, balancing technological augmentation with user-centric design considerations to optimize both interaction quality and user satisfaction.

Although there were minor differences in the preference scores between S1A and S1AR, the difference was not statistically significant. Therefore, we believe that this is due to variability in individual user experiences and preferences between conditions [3]. Therefore, we conclude that the robot and comfort was a significant factor in shaping the users' opinons.

8.2 Scenario 2 (S2)

Despite participants rating both of S2's conditions similarly in terms of perceived helpfulness, their overwhelming preference for the agent-only condition (S2A) suggests a clear inclination towards simplicity and direct interaction. This preference aligns with user-centric design principles that prioritize intuitive user interfaces and minimal cognitive load [23,32], and also demonstrates the user's preference of the MR-agent over the MR Agent with the robot.

We believe that in S2AR, users could have possibly been visually distracted by the simultaneous movements of the MR-agent and the robot, and thereby eventually just looked at the MR-agent and rendered the robot as irrelevant to the context. As our statistical results did indicate a significant effect between conditions of S2, such findings further indicate that while the robot did not hinder task performance, its presence did not confer significant advantages over the agent-only condition. The post-study questionnaires about user's preferences within real-life applications (Fig. 9 Q3) and their rankings of the experiences highlight the importance of user comfort and preference in MR applications involving practical tasks.

8.3 Overall Analysis

In S1, participants found the robot less favorable compared to interacting solely with the MR agent. We believe that this preference was mainly due to perceived stability and ease of interaction without the added physical components. S2 results did indicated that the robot did not play a significant role in enhancing the efficiency or interaction experience. However, both of S2 conditions (with and without the robot) had similar scores in terms of perceived helpfulness.

Accordingly, the mentioned user ratings in S1 and S2 shows a divergence in their opinions about entertainment-focused applications (S1) and practical/task-oriented applications (S2). While embodiment through the robot enhanced entertainment value and interaction dynamics in S1, its added complexity did not translate into significant advantages for practical tasks in S2. The results showed that users' thought that direct interaction with the virtual agent sufficed the interaction needs.

Overall, participants were very satisfied with the interaction experience across the two scenarios, indicating effective engagement and perceived safety of our system. These results underscore the potential of such systems to provide enjoyable and effective applications, and pave the way for exploring further deployment contexts.

However, participants identified a number of shortcomings. In S2, simultaneous feedback using the robot and the MR agent distracted users and impacted the system's efficiency. Furthermore, ergonomic factors like the stability of the wearable robot and comfort were significant concerns among participants, potentially affecting prolonged use and adoption of the system. Most importantly, participants thought the wearable robot should be lighter, as results show that the system's weight negatively affected user's judgment and opinions. Such aspect is well established and in line with findings from previous work [33], and should accordingly be addressed when designing future robotic systems.

8.4 Future Work

In light of our design, implementation and evaluations, we discuss a number of future research directions as follows:

Agent Embodiment and Visualizations: Virtual agents are versatile, and can be embodied with different sizes, locations, and visual appearances. Such factors can be embodied in various methods using different types of wearable and mobile robotic platforms. For example, human-sized agents can be embodied partially, such as their hands, fingers or arms during physical interactions [7,10]. Therefore, exploring agent visualization and embodiment's potentials is critical for enabling robust and seamless user experiences on wearable robots.

AI-Based Conversation Systems and Non-Verbal Communication: Generative AI and large-language models have paved the way for various applications of AI for conversational systems and robot control [19,22]. Using such systems can create a seamless user experience with the MR agent beyond the limitations of voice commands, which could enable more engaging and meaningful interactive experiences similar to character-based humanoid robots [37]. Such research direction is critical for the advancement of virtual agents and wearable robotic appendages. Furthermore, nonverbal cues, such as hand-gestures or facial expressions, should be investigated to explore how they can shape conversations and interactions with the MR Agent and robot.

Physical Manipulation Tasks: Although we explored aspects of MR agent embodiment on two basic tasks that demonstrate the importance of embodiment when interacting with the agent, an important tasks to explore is physical manipulation tasks, which is a critical task domain of wearable robotic appendages [2]. Future work should investigate conducting physical manipulation tasks delegated to the robot by the user, or in collaboration between the user and the agent/robot. Therefore, designing such interactive experiences presents an intriguing research direction for further utilization of MR Agents and wearable robotics appendages.

Robot Design: Although our robot was lighter in weight than previous designs, we believe that the design of the robot was heavy enough to induce negative results, especially during prolonged use. Therefore, future systems should look into reducing the weight of the robot, such as by using cable-driven mechanism, smaller servomotors, or simplified robot structure with lower DoFs. Such highly flexible structures could be used to better embody the MR-agent's body, similar to humanoid [37].

Object Recognition and Localization: A critical shortcoming of our system stems from the HMD's restrictions to provide access to its raw camera feeds, limiting our capability to use such information for recognizing and localizing objects in the surrounding environments. As such capability is essential for precise physical manipulations using the robot, it is essential to either equip the robot with sufficient cameras and sensors to enable precise manipulations, or to use an HMD that provide access to such capabilities.

Extended Evaluations: Our current study is limited in terms of sample size, diversity and evaluated tasks. Future work should extend the evaluations with larger and more diverse sample sizes. Furthermore, the evaluations should focus on additional tasks (e.g. physical manipulation, haptic feedback, etc.).

9 Conclusion

This project investigates using wearable robotics with MR agents, exploring a novel approach for embodying MR agents in various interaction domains. This integration enriches various tasks and activities by offering practical assistance, entertainment, and companionship. Accordingly, Our study contributes with insights for effectively using MR agents with wearable robots in leisure and serious applications. The users' preference for direct and sole interaction with MR agents, in all evaluated tasks, highlights the importance of constant embodiment of the MR agent using the robot. Future research should focus on enhancing the design of the wearable robot, especially reducing its weight. Further critical application domains should be investigated, including physical manipulation and haptic interaction with the embodied MR agent. These research directions are crucial for optimizing human-MR interactions and ensuring seamless integration into daily usage contexts.

Acknowledgments. This project is supported by Qatar Japan Research Collaboration fund (QJRC) M-QJRC-2023-325. This project is also supported by JST SPRING, Grant Number JPMJSP2128.

References

1. Aaron Bangor, P.T.K., Miller, J.T.: An empirical evaluation of the system usability scale. Int. J. Human Comput. Interact. **24**(6), 574–594 (2008). https://doi.org/10.1080/10447310802205776
2. Al-Sada, M., Höglund, T., Khamis, M., Urbani, J., Nakajima, T.: Orochi: investigating requirements and expectations for multipurpose daily used supernumerary robotic limbs. ACM (2019). https://doi.org/10.1145/3311823.3311850
3. Al-Sada, M., Jiang, K., Ranade, S., Kalkattawi, M., Nakajima, T.: Hapticsnakes: multi-haptic feedback wearable robots for immersive virtual reality. Virtual Real. **24**(2), 191–209 (2020). https://doi.org/10.1007/s10055-019-00404-x
4. Al-Sada, M., Khamis, M., Kato, A., Sugano, S., Nakajima, T., Alt, F.: Challenges and opportunities of supernumerary robotic limbs. Front. Robot. AI. **5**, 74 (2017)
5. Beaudouin-Lafon, M.: Designing interaction, not interfaces. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 15–22. AVI 2004, Association for Computing Machinery, New York, NY, USA (2004). https://doi.org/10.1145/989863.989865
6. Brooke, J.: SUS: a quick and dirty usability scale. Usability Eval. Ind. **189**, 4–7 (1995)
7. Fabre, E., Rekimoto, J., Itoh, Y.: Exploring the Kuroko paradigm: the effect of enhancing virtual humans with reality actuators in augmented reality. In: Proceedings of the Augmented Humans International Conference 2024, pp. 79–90. AHs 2024, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3652920.3652945
8. Fink, J.: Anthropomorphism and human likeness in the design of robots and human-robot interaction. In: Ge, S.S., Khatib, O., Cabibihan, J.-J., Simmons, R., Williams, M.-A. (eds.) ICSR 2012. LNCS (LNAI), vol. 7621, pp. 199–208. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34103-8_20
9. Guimarães, M., Prada, R., Santos, P.A., Dias, J.a., Jhala, A., Mascarenhas, S.: The impact of virtual reality in the social presence of a virtual agent. In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. IVA 2020, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3383652.3423879
10. Ihara, K., Faridan, M., Ichikawa, A., Kawaguchi, I., Suzuki, R.: Holobots: augmenting holographic telepresence with mobile robots for tangible remote collaboration in mixed reality. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. UIST 2023, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3586183.3606727
11. Iskandar, A., Al-Sada, M., Halabi, O., Nakajima, T.: Investigating requirements and expectations of wearable telexistence robotic systems. In: 2023 IEEE 29th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), pp. 265–266 (2023). https://doi.org/10.1109/RTCSA58653.2023.00040
12. Iskandar, A., Al-Sada, M., Miyake, T., Saraiji, Y., Halabi, O., Nakajima, T.: Piton: investigating the controllability of a wearable telexistence robot. Sensors **22**(21), 8574 (2022). https://doi.org/10.3390/s22218574
13. Jones, B., Zhang, Y., Wong, P.N.Y., Rintel, S.: Vroom: virtual robot overlay for online meetings. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–10. CHI EA 2020, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3334480.3382820

14. Klug, B.: An overview of the system usability scale in library website and system usability testing. J. Libr. User Exp. **1**(6), 602 (2017). https://doi.org/10.3998/WEAVE.12535642.0001.602
15. Liu, Q., Sun, X.: Research of web real-time communication based on web socket. Netw. Syst. Sci. **5**, 797–801 (2012)
16. Meta: Hand Pose Detection | Oculus Developers. https://developer.oculus.com/documentation/unity/unity-isdk-hand-pose-detection/ (2024). Accessed 08 July 2024
17. Meta: Meta Quest Pro: Premium Mixed Reality. https://www.meta.com/jp/en/quest/quest-pro/ (2024). Accessed 08 July 2024
18. Meta: Voice SDK Overview | Oculus Developers. https://developer.oculus.com/documentation/unity/voice-sdk-overview/ (2024). Accessed 08 July 2024
19. Miyake, T., Wang, Y., Yang, P.c., Sugano, S.: Feasibility study on parameter adjustment for a humanoid using LLM tailoring physical care. In: Ali, A.A., et al. (eds.) Social Robotics, pp. 230–243. Springer Nature Singapore, Singapore (2024). https://doi.org/10.1007/978-981-99-8715-3_20
20. Nakagaki, K., Dementyev, A., Follmer, S., Paradiso, J.A., Ishii, H.: Chainform: a linear integrated modular hardware system for shape changing interfaces. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology, pp. 87–96. UIST 2016, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2984511.2984587
21. Nakagaki, K., Follmer, S., Ishii, H.: Lineform: actuated curve interfaces for display, interaction, and constraint. ACM (2015). https://doi.org/10.1145/2807442.2807452
22. OpenAI: ChatGPT. https://chatgpt.com/ (2015). Accessed 17 July 2024
23. Oviatt, S.: Human-centered design meets cognitive load theory: designing interfaces that help people think. In: Proceedings of the 14th ACM International Conference on Multimedia, pp. 871–880. MM 2006, Association for Computing Machinery, New York, NY, USA (2006). https://doi.org/10.1145/1180639.1180831
24. ROBOTIS Co., Ltd.: Dynamixel: All-in-one smart actuator. https://en.robotis.com/shop_en/ (2024). Accessed 08 July 2024
25. Rzayev, R., Karaman, G., Wolf, K., Henze, N., Schwind, V.: The effect of presence and appearance of guides in virtual reality exhibitions. ACM (2019). https://doi.org/10.1145/3340764.3340802
26. Schmidt, A.: Implicit human computer interaction through context. Personal Technol. **4**, 191–199 (1999). https://doi.org/10.1007/BF01324126
27. Shi, Y., Dong, W., Lin, W., Gao, Y.: Soft wearable robots: development status and technical challenges. Sensors **22**(19), 7584 (2022). https://doi.org/10.3390/s22197584
28. Suzuki, R., Karim, A., Xia, T., Hedayati, H., Marquardt, N.: Augmented reality and robotics: a survey and taxonomy for AR-enhanced human-robot interaction and robotic interfaces. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI 2022, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3491102.3517719
29. Technologies, U.: Unity-Chan! Model | 3D Characters | Unity Asset Store (2020). https://assetstore.unity.com/packages/3d/characters/unity-chan-model-18705. Accessed 08 July 2024
30. Technologies, U.: Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine. https://unity.com/ (2024). Accessed 08 July 2024

31. Tejwani, R., Ma, C., Bonato, P., Asada, H.H.: An avatar robot overlaid with the 3d human model of a remote operator. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7061–7068 (2023). https://doi.org/10.1109/IROS55552.2023.10341890
32. Thees, M., Kapp, S., Strzys, M.P., Beil, F., Lukowicz, P., Kuhn, J.: Effects of augmented reality on learning and cognitive load in university physics laboratory courses. Comput. Human Behav. **108**, 106316 (2020). https://doi.org/10.1016/j.chb.2020.106316, https://www.sciencedirect.com/science/article/pii/S0747563220300704
33. Urbani, J., Al-Sada, M., Nakajima, T., Höglund, T.: Exploring augmented reality interaction for everyday multipurpose wearable robots. Institute of Electrical and Electronics Engineers Inc. August 2018. https://doi.org/10.1109/RTCSA.2018.00033
34. Walker, M., Phung, T., Chakraborti, T., Williams, T., Szafir, D.: Virtual, augmented, and mixed reality for human-robot interaction: a survey and virtual design element taxonomy. J. Hum. Robot Interact. **12**(4), 1–39 (2023). https://doi.org/10.1145/3597623
35. Woodworth, J., Lipari, N., Borst, C.: Evaluating teacher avatar appearances in educational VR. IEEE (2019). https://doi.org/10.1109/VR.2019.8798318
36. Woodworth, J., Lipari, N., Borst, C.: Evaluating teacher avatar appearances in educational VR. IEEE (2019). https://doi.org/10.1109/VR.2019.8798318
37. Yang, P.C., et al.: Hatsuki : an anime character like robot figure platform with anime-style expressions and imitation learning based action generation. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 384–391 (2020). https://doi.org/10.1109/RO-MAN47096.2020.9223558

Exploring Mixed Reality Design Considerations for Adaptable User Interfaces to Improve Interaction on Physical Textured Surfaces

Shwetha Subramanian[1], Renee Bogdany[1], Michael Crabb[2], Roshan L. Peiris[1], and Garreth W. Tigwell[1](✉)

[1] Rochester Institute of Technology, Rochester, NY 14623, USA
garreth.w.tigwell@rit.edu
[2] University of Dundee, Dundee, Scotland, UK

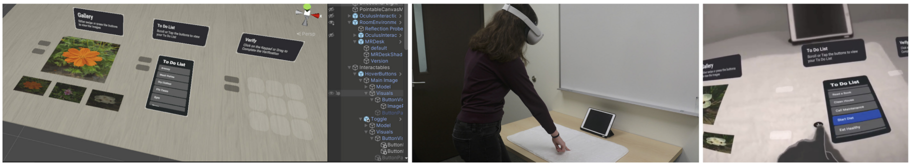


Fig. 1. (1) The three apps we created, (2) study setup, and (3) user's MR view.

Abstract. Mixed Reality (MR) offers new ways to interact with physical environments, but user preferences for gesture interactions on textured surfaces (e.g., carpet, wood) vary. To address this, MR systems should provide alternative interactions–allowing users to choose gestures, such as tapping instead of dragging. We explored the impact of user-controlled alternative interactions on MR textured surface experiences with a 40-participant study. Results show alternative interactions enhance user experience by accommodating individual needs and preferences. However, effective MR design requires consideration of the interplay between user, task, and surface. We provide design insights for future MR systems.

Keywords: Adaptable UIs · Textured Surfaces · User Experience

1 Introduction

Mixed Reality (MR) creates opportunities to explore the world in new ways [38] and has been tested in domains such as education, entertainment, healthcare, and work [11,18,28,31]. Yet, the MR interaction design space is in its infancy .

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 356–376, 2025.
https://doi.org/10.1007/978-3-031-93715-6_22

One type of MR experience that shows potential is turning physical surfaces into interactive surfaces. Prior work showed direct interaction with textured surfaces (e.g., a wooden surface) with an augmented digital UI can enhance the user experience [40]. However, user preferences for textured surface interaction varied [40], suggesting that a customized interaction experience is needed to address individual user experience needs.

Adaptable user interfaces (UIs), which support alternative interaction, could enhance user experiences by aligning an individual's interaction preferences with the surface type they are using (e.g., the UI could offer the user to tap on digital elements instead of dragging digital elements when interacting on a rough textured surface). Therefore, while other MR research has concentrated on adaption in contactless and mid-air interactions [9,29], we are interested in understanding how to enhance the user experience when contacting with physical textured surfaces. Contactless MR experiences also lack haptic or tactile feedback, which can limit immersiveness [6], whereas leveraging textured surfaces in our environments provides a low-cost method for enhancing immersion.

With this in mind, we focus on exploring the user experience of MR textured surface interaction through user-controlled adaptable UIs to understand how different interaction gestures affect usability and what design considerations need to be made to optimize the user experience. We ran a 60-minute user study with 40 participants who completed interactions on eight different surfaces while exploring the benefits and challenges of user-controlled alternative interactions (Fig. 1). We created three MR applications (Image Gallery, To-Do List, and Security Verification). Each MR application supported two interaction styles from a standard set with varying surface contact time duration (tap, press, flick, drag).

We found that alternative interactions can improve the user experience by supporting individual user needs during MR textured surface interaction. However, it is important for future technology design to consider the relationship between the person, task, and surface being used. Surface-based alternative interactions should consider surface physical characteristics, available surface real estate, and environmental conditions. Furthermore, alternative interactions should cue users to when alterations take place and give information regarding how this alteration impacts control. We share design insights to help guide future development that will help to maintain a good user experience.

2 Related Work

2.1 Interacting with Various Surfaces

MR has created opportunities to explore new forms of HCI through overlaying digital imagery on top of real-world surroundings. These range from common environments such as interactive surfaces in kitchens and offices [23,26] to unconventional surfaces like ice walls and the human body [17,44]. Recent work has looked at identifying gestural interaction from the acoustic wave detected during surface interactions [19]. Researchers have also explored various ways to bring physicality to MR with gloves and other wearable devices [2,35], to shape-changing interfaces [30], and electrovibration to add texture to objects and surface [3]. Although researchers have explored using additional hardware to expand the immersive experience, one under-researched area is leveraging the textured surfaces already found in our surroundings. Avoiding additional hardware keeps things simple and, most importantly, low-cost and obtainable for all. Rather than purchasing various additional electronic components, *what if we sought to improve immersion, and consequently UX, by taking advantage of the multitude of surface textures already in our environment*?

Tigwell and Crabb [40] previously investigated how textured surfaces can be used as a low-cost way to enhance MR interaction experiences so that people feel more connected to digital interfaces. In this work, they introduced Household Surface Interactions (HSIs) and the opportunity that MR brings to make all surfaces interactive. HSI research is growing (e.g., investigating the feasibility of tables and pillows as smart home controls [7]).

2.2 The Physiology of Surface Interaction

Physiological and psychological studies have investigated how we experience textured surfaces (e.g., [16]), and object materials can offer various interaction affordances [13,21] while associated textures affect interaction [10], which can be taken advantage of for the benefit of users. Furthermore, when designing with a particular material in mind, the material itself can impact what function a specific object can perform and what input/output methods are possible [36]. TexelBlocks was a dynamic surface of multiple textured squares that mix and match through a rotating mechanism [8]. This allowed for a creative surface that could either fully display as one texture or include a mixture of textures at one time. The authors describe several use cases, such as immersive storytelling and games, however, there was no user study.

The goal of providing direct manipulation where there is both predictability and control for the user is important [37]. Yet, prior work has highlighted that challenges exist when using non-typical materials for interaction, such as ice [44], and surfaces may inherit different properties due to the interaction itself (e.g., when interacting with liquids, the fluid properties and degree of contact will affect interaction [15]). MR can be used to support surface interaction on a wide variety of surfaces, but it does raise the question of *how do we support designers in creating optimal interfaces and interactive experiences on textured surfaces*?

2.3 Adaptable UIs and Alternative Interactions

One strategy to improve the user experience when interacting with textured surfaces during MR is to explore adjustments to the UI and provide alternative interactions. MR experiences can be improved by leveraging context awareness to optimize position, size, and the level of information of floating windows within a real-world view [9,29]. However, the work was not conducted in the context of physical surfaces. In our work, we want to focus on a more tangible situation where people are directly interacting with augmented physical surfaces. The issue is that in most cases, we cannot alter a person's physical environment, but we can influence digital design, so *how do we adapt MR interactions and interfaces for different textured surface materials*?

When creating alternative interactions, user-led control can be used to allow users themselves to take a proactive role in determining how adaptions should be implemented [12]. This is in contrast to system-led adaptions, where user models are created to facilitate adaption [32]. Although Tigwell and Crabb's [40] work did not investigate user-controlled alternative interactions, they did recommend that such solutions could potentially address the challenges their users experienced. We use their recommendation to guide our work.

2.4 Research Questions

Considering what has been achieved in prior work, our work focuses on understanding the user experience of MR textured surface interaction with user-controlled alternative interactions. Our research questions are:

(1) *What design priorities will maintain usability when creating alternative interactions for textured surface interaction?*
(2) *What are the challenges and opportunities in using alternative interactions during textured surface interaction?*

3 Method

3.1 Pilot Study

Before we provide the specific details of our main study, we want to highlight some details from our pilot study, which were helpful in determining user study details such as duration, surface materials to use, and clarity of questions.

Our *pilot study* involved five participants (Male = 2, Female = 3) with the following age ranges: 18–24 = 4, 25–34 = 1. Our participants had varied experience with AR/VR/MR technologies (on a scale of 0–9 where 0 is 'No experience' and 9 is 'Very Experienced'); Mean for AR = 3.4, VR = 3.4, and MR = 1.4. We realized for the main study that we should find out about prior experience with the Quest 2, and so we added it as a question in our main study.

We initially sourced many different textured materials to use in our research, including rugs, carpets, wooden materials, tiles, etc., but we knew it would be challenging to include them all, and therefore we had to balance selecting a large enough number with evaluation time. During the pilot studies, we went from participants using 18 surfaces to using eight—we found that participants could sufficiently try out eight surfaces within 60 min and discuss their experiences. Similar to prior work [40] we selected eight materials to cover a range of different real-life textured surfaces (e.g., flooring, countertops).

3.2 Apparatus and Materials

Experimental Equipment. We used the Meta (Oculus) Quest 2 running Oculus Build v38. Our apps ran on Unity version 2020.3.27f1. We chose the Quest 2 because its Passthrough API facilitated the necessary immersive MR experience. The Quest 2 possesses a large diagonal Field of View (113°) compared to contemporary MR headsets (e.g., Magic Leap = 50°, HoloLens 2 = 52°).[1] We required our users to physically touch the surfaces, which meant they would be standing close, and other headsets could not show the fully interactive environment. Due to the restricted field of view of the Hololens 1 and 2, we decided that the Quest 2 faired significantly better.

Surface Materials. We sourced materials from home and hardware stores. Our final surface selection (informed by the pilot study) included two carpets, one rubber mat, one rug, three tiles, and one wooden surface (see Table 1 and Fig. 2). We classified our surfaces based on material, texture, and color. Although the hue of the surface materials was not so important since Passthrough shows the external worldview in grayscale, we wanted to have variability since participants would still be able to perceive color variations of light.

Table 1. Descriptions for the eight textured surfaces used in our study.

Code	Material	Texture Description	Color & Design
C1	Carpet	Short pile, coarse	Dark, no pattern
C2	Carpet	Short pile, coarse	Light & Dark, pattern
M	Rubber Mat	Grooves, smooth, hard	Dark, no pattern
R	Fur Rug	Fine fibers, medium pile, very soft	Light, no pattern
T1	Glossy Tile	Flat, smooth, hard	Light, no pattern
T2	Marble Tile	Flat, matte, rough, hard	Light, patterns
T3	Ceramic Tile	Wavy surface, smooth, hard	Light, pattern
W	Wood Countertop	Flat, smooth substance, hard	Medium, pattern

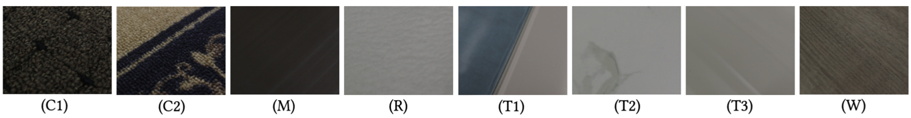


Fig. 2. Close-up of the 8 textured surface materials.

Interaction System. Our chosen interactions are based on a gesture guide [45] and fall on a continuum of increasing contact time (fastest to slowest):

[1] All FoV information is taken from https://vr-compare.com/.

1. **Tap**: User briefly touches a surface with their fingertip.
2. **Flick**: User quickly brushes a surface with their fingertip.
3. **Press**: User touches a surface for an extended period of time.
4. **Drag**: User presses with one finger and moves over without losing contact.

We created three MR apps, inspired by existing mobile device features, to investigate user-controlled alternative interactions: *Image Gallery*, *To-Do list*, and *Security Verification* (Fig. 3). User interaction with each app varied in both effort and time required. Tapping (*Image Gallery* and *To-Do List*) involves lightly touching the thumbnail to highlight an image. Pressing (*Security Verification*) requires significantly more pressure and visually urges users to depress the buttons. This gives the illusion of holding and pressing a button. Interactions such as flicking horizontally and vertically (*Image Gallery* and *To-Do List*) require significant motion and contact of the fingers over the surface compared to tapping. Whereas dragging (*Security Verification*) is accompanied by an altered key shape to depict a pattern that requires a combination of adding pressure while performing a pattern motion.

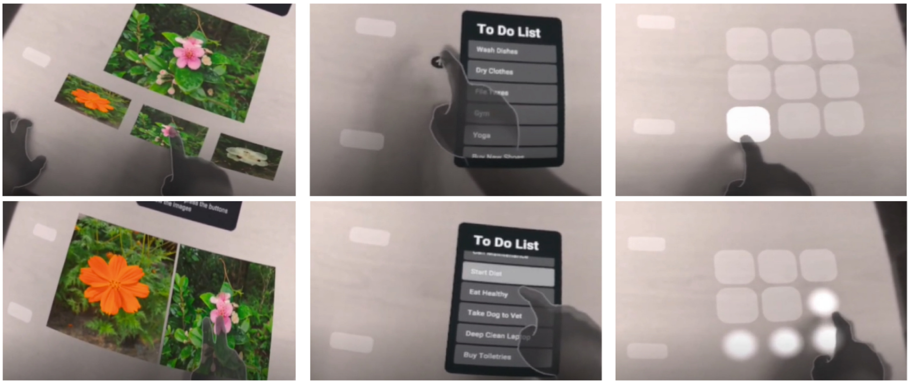


Fig. 3. Top Left: Image Gallery (Tap), Top Middle: To-Do List (Tap), Top Right: Security Verification (Press), Bottom Left: Image Gallery (Horizontal Flick), Bottom Middle: To-Do List (Vertical Flick), Bottom Right: Security Verification (Drag).

The **Image Gallery** UI (Fig. 3, left top and bottom) displayed three images that could either be viewed by flicking horizontally from side to side or by tapping on their thumbnails. The interactions featured here were *tapping* and *horizontal flicking*. Changing interaction styles for Image Gallery resulted in an immediate change in application layout to inform our participants that a change in interaction method had occurred.

The **To-Do List** UI (Fig. 3, middle top and bottom) displayed a list of various items that participants could either highlight and flick through vertically or by touching arrow buttons on the immediate left of the body of the list. The interactions featured here are *tapping* and *vertical flicking*. Switching between interaction styles meant the adaptive UI made arrow buttons visible/invisible.

The **Security Verification** UI (Fig. 3, right top and bottom) functions as a common PIN/pattern password input. Participants either press to enter a pin or employ a drag gesture similar to pattern input on smartphones. The interactions here are *pressing* and *dragging*. Changing the interaction style for this UI made no visible change to the UI layout.

Our apps were placed side by side so participants could easily reflect on similarities/differences during the study exploration phase. We placed two buttons on the right-hand side of each app so that our participants could toggle between interaction styles (shown in Fig. 3).

Interview Guide. During the study, we asked questions to understand our participants' in-the-moment thoughts toward alternative interactions. During the post-exploration interview, we inquired about participant preferences, their likes and dislikes about the MR experience, their thoughts on design, interaction styles, interaction-texture conflicts, usability, future implementation of MR in household surface interactions, as well as thoughts on alternative interactions (including attitudes toward manual vs. automatic control).

3.3 Participant Information

Participant Recruitment. We advertised our study on campus and through social media. Interested participants completed a screening questionnaire, which contained questions regarding demographic information and prior experience with AR, MR, and VR technologies. We took precautions to avoid recruiting participants who might experience issues with MR and textured surface interaction (e.g., motion sickness, photo-sensitivity, skin allergies).

Demographic Details. Our main study included 40 participants (Male = 19, Female = 15, Non-Binary = 3, Female/Non-Binary = 1, Prefer Not to Say = 2). Our main study participants could be categorized into the following age ranges: 18–24 = 28, 25–34 = 11, 65+ = 1.

Our participants had varied experience with AR/VR/MR technologies and the Quest 2 (On a scale of 0–9 where 0 is No experience and 9 is Very Experienced; Mean of AR = 3.28, VR = 3.7, MR = 1.63 and Quest 2 = 1.55). Thirty-six participants were acquainted with AR (e.g., playing mobile AR games like Pokemon Go). Five participants had tried AR experiences for visualization and productivity (e.g., Target Inc.'s 'See It In Your Space', Measuring Distance in AR). Thirty-five participants were aware of VR, and 24 mentioned that they had tried HMDs such as the Oculus Quest (now Meta), as well as having played video games (e.g., Beat Saber) or witnessed immersive experiences. Twenty-two had at least some MR experience, whereas 18 had no experience using MR. Seventeen participants were aware of or had used the Quest 2 prior to the study.

3.4 Procedure and Analysis

Our Institutional Review Board (IRB) approved study was scheduled for 60 min, and we reimbursed participants with $25 cash (USD). All sessions took place in a dedicated experiment room on campus.

We presented participants with the scenario of *interactions in the home* to provide a relatable context to our study, especially because digital technology is commonplace within homes [43], and emergent smart homes [20,27,48]. However, our findings generalize more broadly since the textured surfaces in our study are found in many settings.

We recorded each participant's view through the MR headset to support later analysis. We also recorded audio in two locations: next to the researcher (Samsung Galaxy M51 phone) and in front of the participant (iPad Pro 10.5").

After a short briefing and introduction to the MR Headset, we asked our participants to explore our adaptive UI apps on various textured surfaces. We used a *William Latin Square* counterbalancing method [46] for our within-subject design to reduce any surface texture order effects. We asked our participants to rate each surface based on the comfort of interacting with the surface materials (1–7, where 1 = uncomfortable and 7 = comfortable) and ease of use of the interactions (1–7, where 1 = unusable and 7 = usable), as well as asking for their interaction style preferences for each app. We encouraged our participants to use the *Think Aloud* method [25] so we could capture in-the-moment thoughts, and we also asked questions to prompt feedback. Qualitative approaches in surface interaction research are found to be extremely beneficial in understanding insights such as user preferences (e.g., [47]). Wobbrock et al. [47] have shown the value of participatory design in eliciting possible gestures to complete surface interaction tasks, although their work was not focused on how surface material texture affects user-preferred gestures or the potential of alternative interactions.

Our *qualitative* procedure also aims to extend prior textured surface interaction research [40], which sought to *quantitatively* evaluate interaction types on each surface systematically and in isolation—the tasks in Tigwell and Crabb [40] required participants to cycle through different interactions without the option to switch back and forth between interaction types on an individual surface. Our procedure intentionally avoids isolated interactions since people are typically exposed in real-world applications to UIs that encourage using a mixture of interaction types—in some cases, users can even switch interaction types on the same input elements (e.g., modern smartphone keyboards allow users to seamlessly switch between tapping and swiping on keys to input letters).

In the post-exploration phase, we interviewed our participants. We followed Braun and Clarke's [5] steps to perform an inductive thematic analysis.

4 Findings

4.1 General Perspectives on the MR Experiences

Our participants reported a desire to use textured surfaces themselves in the future if alternative interactions and surface MR experiences were commercially

Table 2. Participant preferences (count) for interaction styles and ratings (scale of 1–7) for comfort and ease of use for the 8 textured surfaces. (Note: T = Tapping, HF = Horizontal Flicking, VF = Vertical Flicking, P = Pressing, D = Dragging, App 1 = Image Gallery, App 2 = To-Do List, App 3 = Security Verification. We also report multiple modes in a few places).

	App 1		App 2		App 3		Comfort				Ease of Use			
	HF	T	VF	T	D	P	Min	Max	Mean	SD	Min	Max	Mean	SD
C1	2	38	5	35	0	40	1	7	3.36	1.66	1	7	3.76	1.56
C2	7	33	7	33	2	38	1	7	3.73	1.62	1	7	4.03	1.58
M	27	13	23	17	24	16	1	7	4.51	1.46	2	7	4.93	1.31
R	24	16	24	16	28	12	3	7	5.95	1.15	3	7	5.53	1.11
T1	32	8	29	11	26	14	3	7	5.91	1.14	3	7	6.06	0.92
T2	9	31	10	30	6	34	1	7	3.73	1.62	2	7	4.58	1.43
T3	17	23	8	32	6	34	1	6	3.78	1.45	1	7	3.78	1.53
W	15	24	16	22	16	23	1	7	4.94	1.57	2	7	5.21	1.40

available. Our participant's comfort and ease of use preferences for the combinations of interaction styles and surfaces varied (see Table 2).

When reviewing the participant preference ratings for interaction style across apps and surfaces, we found that a large majority of participants had a preference for shorter contact interactions (tapping in App 1, tapping in App 2, pressing in App 3) when using the surfaces that had a coarseness or were not flat (i.e., C1, C2, T2, T3) compared to the smoother and softer surfaces, which generally had less skewness in participant interaction preference. While this observation may seem predictable, it is important to note that apart from App 3 used with C1, no other condition had unanimous agreement, demonstrating that some people's preferences can be opposite to others when it comes to surface texture and interaction and supporting findings from the literature. Furthermore, the Fur Rug (R), Glossy Tile (T1), and Wooden Countertop (W) were the top three rated surfaces for both comfort and ease of use. Notably, six surfaces for comfort and three surfaces for ease of use received at least one score on both the minimum and maximum ends of the rating scale, again demonstrating how widely preferences on textured surfaces vary, thus reinforcing the importance of exploring alternative interactions to improve user experience.

Interaction-Texture Conflicts. We found that interaction-texture conflicts can occur during either surface changes or, in general, when participants interact with a surface using an interaction style they do not prefer. For example, an undesired interaction style can be used on a surface that the user may find uncomfortable or not easy to use. Since we were interested in user experience, we asked what participants would do when facing a conflict of this nature.

Participants described that they would not want to continue performing certain tasks that were inconvenient on certain surface textures, which assists in motivating our reason to understand the benefits of alternative interactions for addressing contextual interaction issues in MR environments. For example, P22 said, "*sometimes you like scrolling, sometimes you'll like tapping*" and their interaction preferences can also be affected by how they are sitting.

It is clear that different scenarios require interaction options that best match the context for the individual. There was also feedback highlighting the need to ensure the textured surfaces are 'compatible' with the interaction type (i.e., MR systems for textured surface interaction will maintain a good user experience when input methods are optimized to reduce input errors). However, there were participants who mentioned that they would tolerate ineffective interaction-texture pairs if the task was of high priority.

Surfaces for Interaction Purposes. We asked participants what sort of surfaces they would interact with if an MR system was installed in their homes. Our participants provided many insightful comments about the use cases they imagined with surfaces. Overall, we found most of our participants suggested kitchen countertops, desks, and walls as being ideal surfaces for interaction, while some participants mentioned their bed or bed sheets, and other participants considered glass-based furniture, and even outside.

Potential Issues of Using Mixed Reality at Home. Our participants expressed some concerns about MR within home life and real-world applications. For example, P26 had concerns about using MR in an evolving environment with "*other people in the room*", as well as hardware limitations for tracking.

In general, our participants discussed realistic issues (e.g., P14 highlighted environmental issues pertaining to lighting which can be a major issue considering the multiple light sources at homes), while P18 wondered what the lasting impact would be when interacting with surfaces at home that do not typically have a lot of contact (e.g., wearing away materials). Other concerns included messy surfaces, issues pertaining to temperature, and sharp objects such as knives or spills not being visible through HMD cameras.

Thoughts on Teaching the System for Automatic Adaption. Recognizing that people have different preferences for interaction-texture pairs, we asked participants how they might want to teach the system of their preferences, which would make for a more personalized experience. Our participants identified many ways in which alternative interaction systems can be taught and function automatically. The two general approaches proposed were (1) the system knows what to do because of surface type, and (2) the user informs the system.

One example of the first category was offered by P10 involving a sensor to "*detect when either scrolling or tapping*" paired with "*a machine learning algorithm that's gonna predict what you're going to like more when it sees a new surface.*" While there are definitely benefits to the system freeing up the user's time by making decisions, those decisions still need to be informed in some way. Altering interaction based on the frequency of usage and user trends was brought up significantly by participants, but we know people can differ in their preference for the same surface types. Therefore, user input is likely to be the most useful method. P14 suggested this could be achieved by teaching the "*system based on our ratings*", which are submitted during use.

Participants recognized the importance of onboarding and its potential use to train both the user and the system simultaneously. P23 mentioned how people are used to doing initial setups when enabling biometric authentication on mobile devices by recording a set of fingerprints, and P23 used this example to indicate a potential simple calibration setup on first use.

General Design. Our participants provided valuable feedback and suggestions for design and usability. Although this is better covered in the main themes from our analysis (see 4.2, 4.3), we would like to highlight suggestions that were provided out of the purview of alternative interactions regarding user experience.

Our participants provided us with insights and suggestions for the future design of the MR experiences, especially taking inspiration from familiar technology (e.g., P10 mentioned how the design of phones has got to a point where people can pick one up and use it even if it is not one they own). A common design language will certainly aid in maintaining a good user experience.

Since interaction in MR textured surfaces is a visual task, our participants highlighted the importance of good visual design, especially when it could affect accessibility. For example, P13 discussed the importance of color contrast when the color of surfaces changes. In addition to aesthetic visual design, our participants discussed the information side of visual design. Mainly, visual cues need to be actively incorporated into the design of alternative interactions, which will help improve the user experience along with an intuitive and easy to use design.

Next, we share insights into the two main themes we identified during our analysis. Theme 1: *Alternative interactions are beneficial, but there is ambivalence toward automatic control* (Sect. 4.2), and Theme 2: *Users want the ability to customize and control the UI* (Sect. 4.3).

4.2 Alternative Interactions Are Beneficial, but There Is Ambivalence Toward Automatic Control

We asked our participants about their thoughts on automatic adaption in order to understand whether applying it in the future could, in theory, make for a more seamless interactive experience for enhancing the user experience of MR textured surface interaction.

Most participants found user-controlled alternative interactions to be quite beneficial due to the ability to switch to a more suitable interaction style when using a particular surface texture. We acknowledge that different users have different sets of priorities and preferences, and providing alternatives can be of great help, as summarized by P37: "*[Alternative interactions] definitely makes it a lot more versatile because I feel like everyone would have a different opinion on what they would like more [...] I had a different opinion for all the surfaces, so it makes it easier for like us anywhere.*"

While P37 mentioned the versatility of the UIs with alternative interaction, some other participants acknowledged the changeability of preferences based on moods and certain situations. For example, not changing the type of interaction due to laziness or changing an interaction due to sheer boredom. Other examples include situational moods where participants expressed that they would keep or change an interaction just because they 'felt like it'.

Although there were participants in favor of automated adaption of the UI and resulting interaction input, other participants expressed their displeasure and hesitancy toward the idea of automatic adaption. For example, P21 said: "*I mean, the biggest thing is just how quickly automated it becomes. Like I think for me, personally, it's better to be a lot more conscientious of those things and like a lot less of allowing a machine to run my life.*"

Other similar comments also drew attention to the point that sometimes people have a particular way they want to do something, even if it is not what might be considered optimal. The compromise comes down to user choice. Our participants mentioned systems that, when providing automatic adaption, there needs to be a manual option as well. In particular, P16 discussed the idea of consent and it is an important part of the system seeking permission to adapt so that users remain in control.

Alternative interaction is extremely beneficial and eliminates issues pertaining to a lack of user preferences, but the notion of automatic adaption gives some users hesitancy due to the uncertainty behind it. The discomfort brought by a sudden automatic change without prior notification to users is something that cannot be discarded by the vast majority of users.

4.3 Users Want the Ability to Customize and Control the UI

Users want to control and customize the UI, whether it is with respect to aesthetics or options for alternative interaction.

Aesthetics and the UI Should be Customizable. During the post-task interview, we received feedback that the aesthetics, color, layout, and design of the UI are something that should be customizable, owing to the unique preferences of users. The ability to manually resize, reorder, and refurbish the UI in MR could potentially incentivize people to use the system. For example, P26 discussed how phones provide users with support to change the layout and other visual elements (e.g., text size), and therefore is "*definitely a good move to let people customize [MR textured surface interaction]*" because locking users into one way of using gestures/interactions within programs would "*be an issue*".

Design flexibility is a crucial component in meeting the needs of a large user base. Some participants also brought up customization of the aesthetics and color of the UI, which may aid with readability on certain surfaces. Not only are adjustments to visual elements useful for people to meet their aesthetic design preferences, but accessibility too. A few participants spoke about the visual clutter in the virtual space, which could be altered if needed to improve the user experience. The ability to adjust UI layouts would make the interaction more accessible for people with dexterity and motor impairments or even to match handedness (P23, who is left-handed, commented on interaction challenges and the desire to flip the UI). However, it is likely right-handed people would benefit from customization as well due to how users performed interactions.

Control for Accessing UI and Adaption Should lie with the User. Whether it be for controlling the adaption of the UI or the general UI in MR, participants urged that the main control should be manual and they can always be in control. For example, P35 said: "*I want to be able to control where things are at all times [...] So, for me, the best usability would be making sure that you're able to control where everything goes. I think drag and drop functions will be really good for that. And then also just something where you could see like a full menu and move things from there would work really well there too.*"

Leaving the system to make decisions was not preferable. Participants discussed wanting control over the items displayed and the UI. Providing users with control over what happens virtually is imperative and will help to improve the user experience. P39 brought up the idea of comfort level, which is something that will vary not only between people but also within an individual. MR textured surface interaction is not commonplace, and people will need time to adjust to interacting with surfaces in this context.

The user should be in control of not only the UI and adaption but also the nature of control. Alternative input controls were suggested, such as voice input, but that is out of the scope of our study focus. More related were comments on hand gestures as a way to control the UI and the adaption of the interface.

5 Discussion

Creating alternative designs that adequately meet user needs in changing contexts can be challenging [14,34,41]. Alternative interactions must follow a cycle of an individual using a given system, deciding that an adaption needs to take place, selecting what the adaption should be, and evaluating if an adaption is acceptable [14]. These alternative interactions should consider the person requiring the adaption, the application being adapted, and the environment where the adaption is taking place. It is, therefore, difficult to create guidance that can be followed on a granular level, and a more holistic approach must be taken.

5.1 UX of Adaptable Textured Surface Interaction

What design priorities will maintain usability when creating alternative interactions for textured surface interaction?

Our participants found adaptable UI interaction beneficial, commenting that they appreciated having the option to choose between interactions they could perform. Participants discussed that when interacting with surfaces with unusual properties, alternative interaction methods alleviated the inconvenience associated with performing task gestures. Our participants commented that key considerations surrounding maintaining the usability of textured surface interaction should include a complete understanding of the task ecosystem and the inclusion of system characteristics to inform users of when adaptations are taking place.

Prior work highlights a current challenge in the AR/VR space is the lack of guidelines [1], thus we reflect on our findings to provide design insights.

Design Insight 1: Interaction Relationships over Task Performance. Textured surface alternative interactions should consider the relationship between the person, the task, and the surface being used. The effect of overall environmental constraints on an individual's ability to carry out a task must be considered when developing textured surface interactions. When creating digital interfaces that interact with physical objects, it is not easily possible to alter these objects to improve task performance. We prioritize creating digital alternative interactions that consider an individual's needs, the surface's characteristics, and task-based requirements. Our participants described potential accessibility challenges that may be faced, discussing that physical and visual accessibility areas may further influence how alternative interactions can be made. Further consideration of the types of alternative interactions is required, and additional work is needed to understand accessibility challenges within MR systems.

Design Insight 2: Surface Characteristics over Interface Norms. Surface-based alternative interactions should consider surface physical characteristics, available surface real estate, and environmental conditions. Our participants commented that home environments are spaces that change daily and that the requirements of home spaces alter depending on the needs of individual people, as reported in previous work [40]. Understanding how surface-based UI fits into home areas is essential, paying attention to what individual surfaces are used for, and the overall environmental conditions for a given time of day. For example, a breakfast bar in a kitchen can be used as a space for eating and preparing food, a space to carry out work, and a place to facilitate social interaction. Activities within an environment alter the physical space available that can be used for mixed-reality devices. In addition, lighting conditions in a space may also alter a surface's overall contrast levels, creating additional constraints on how a surface can be used (for both physical and virtual objects). We believe that further consideration of how surface-based applications and interactions are embedded within home environments is vital, and we recognize that in situ work such as this cannot occur until previous challenges are solved.

Design Insight 3: Continuous UI Onboarding over Expected Behavior Patterns. Alternative interactions should cue users to when adaptions take place and give information regarding how this alteration impacts UI control. While our participants were receptive to the idea of alternative interactions, they highlighted challenges relating to understanding what effect alternative interactions have had on individual interface interaction methods. In our work, we examined three interfaces that used different levels of UI alteration to highlight how interaction methods changed. Our applications demonstrated different levels of layout change connected with an interaction change (e.g., Image Gallery demonstrated the most visible change). Our participants discussed that adding more visual or auditory cues would be beneficial in assisting users in understanding the expected user behavior for a given task, thus highlighting the importance of this design consideration to improve user experience. We believe that further consideration is needed to inform users of interaction methods for surface interactions. Additional UI information could include traditional methods, such as including additional UI elements, but may also focus on using element micro-animations to subtly demonstrate the default interaction in a given situation.

5.2 Designing Alternative Interactions for Textured Surfaces

What are the challenges and opportunities in using adaptable UI during textured surface interaction?

The adaption of UIs, when done automatically, should take into consideration the type of surface being interacted with. Automatic adaption must be paired with some form of manual feature to provide control to the user. The design and usability of a surface-based interactive experience can be improved by providing customizability to users.

Challenge 1: The Contrast Between the Real World and UI Elements. MR interfaces can introduce users to situational impairments. Prior work using projection-based AR reported on the distortion of UI elements on different surface textures [40]. We planned to avoid those projection issues by using an MR HMD and we uncovered additional situational impairments that must now be considered. One main area that participants focused on was how the color of the surface affected their perception of the user interface. Participants suggested that the UI colors should be customizable depending on the surface they are interacting with.

Example Implementation: Specific situational visual impairment guidelines are lacking [42], but one method to overcome contrast challenges would be to use pre-existing color guidance such as those present within the Web Content Accessibility Guidelines [22]. Future work could also examine enabling the MR system to understand the color properties of the surface to then alter the colors of UI elements to increase contrast to a level that makes digital content easily viewable by the user. This would also allow for personalized color profiles [39] and would increase the visual accessibility of MR systems. Some prior work has explored compensating for imperfections that appear in images when projected on patterned surfaces [33], but this was not using HMDs.

Challenge 2: MR UIs Obscure Surface Properties that Users may Find Unpleasant. Our participants reported that there could be situations where surfaces being interacted with may have unfavorable properties for certain types of interaction. Prior work reported on similar concerns [40], but it was with projection not HMDs. It seems that this participant concern is significant enough to apply across different immersive experience technologies. Our participants discussed how surfaces may have dirt or liquid on them that they would be unaware of due to MR interfaces obscuring surface properties. This obfuscation could be potentially problematic for users if they used MR interfaces on, for example, a kitchen countertop when reading a food recipe [4]. Our participants also pointed out other surface objects may pose additional risks (e.g., knives in a kitchen) if obscured by MR display elements.

Example Implementation: Future surface-based MR systems must highlight potentially unfavorable interaction areas for users or position elements so those areas are not a focus for interaction. For example, Slider UI elements may be positioned on a tiled surface to avoid crossing over tile edges (that are likely to be sharp). Future systems may also point out other potentially hazardous surface properties that are naked to the human eye (e.g., temperature extremes). We see this area as being a challenge that has the potential to be addressed using a combination of computer vision, machine learning, and sensor-based techniques.

Opportunity 1: Tasks Deemed as a High Priority May Not Require Automatic Adaption Due to User-Perceived Speed Requirements. We asked participants if there would ever be a situation where they would continue using an interaction style that may otherwise be inconvenient to use on a particular surface. In this case, interaction-texture conflicts could be considered problematic or ambivalent. Participants commented that if the tasks they are performing are of a high priority, or the seemingly 'preferred' interaction style happened to be uncomfortable for a specific task within an application, they would maintain their desired interaction style. Some participants described that they were heavily biased towards short gesture interactions and would use only those interaction gestures regardless of the type of surface they were interacting with. However, participants also described that the type of interaction they would be comfortable using might vary depending on aspects such as their emotional state or the position of their body in relation to an interactive surface.

Example Implementation: Considering how participants had such varied opinions on interaction styles and when particular gestures should be used, we believe there are opportunities to create UI elements to support multiple interaction styles simultaneously. One way to implement this would be to use techniques that are similar to those currently used within mobile phone text entry where swipe [24] and traditional tap gestures can be used simultaneously.

Opportunity 2: Alternative Interactions for Textured Surface Interaction Can be Automatically Implemented But Should be User-Led in the First Instance. Prior work on surface interaction acknowledges that customization is good but can be at the expense of the user's time [17]. Our participants expressed that they see the benefit of manual adaption of UI interaction. The continual adaption of interfaces can take away from the overall experience of using a given system [14]. Therefore, care must be given to offer control of alternative interactions to users in a way that does not detract from the task being carried out but also to seek input from users when additional information is necessary. Our participants commented they would like manual control for UI elements and would like the ability to create interfaces that work for their specific needs. An additional challenge when creating alternative interactions for textured surface interactions is that while the overall task outcome must remain constant, the UI interactions will change based on the interaction surface.

Example Implementation: We believe there are many opportunities to be explored in this area, particularly in onboarding users into textured surface interactions. It may be possible to implement 'default' interaction gestures for a given task but to provide users with the option to allow these to be overwritten for automatic alternative interactions. This approach would allow users to prioritize consistency in gesture or surface gesture usability. We believe there are opportunities surrounding how to understand user preferences for individual textured surfaces and how this could be used to create personalized interfaces.

5.3 Limitations and Future Work

First, the software we created in the study involved gestures such as tapping, pressing, flicking, and dragging. Due to certain parallax issues, we could not include pinching and spreading gestures in this work. We understand that hand-tracking through head-mounted cameras works best when a user's palm faces the camera. Our use case was the opposite of this, where the palm faced the surface being interacted with, causing tracking difficulties that could not be overcome. While we see this as a limitation in our work, the dual-finger dragging action within pinching and zooming is similar to a single-finger dragging action of a swipe gesture in terms of contact time with a surface.

Second, the surfaces used within the study only covered part of the area of the desk and were compact. We used this approach to test multiple surfaces within one experimental session quickly, as the setup allowed us to swap surfaces in situ. Future work should build on our findings by conducting in situ research within a real-world setting that would allow for the inclusion of surfaces with much more varied properties and dimensions.

Finally, we decided to give users control over when adaption took place to UI elements and not to examine when automated alternative interactions should occur. Both aspects are important but require significantly different study designs. We believe that future work should take a longitudinal approach situated within users' own environments to examine when automatic alternative interactions should occur. This would require prior knowledge of users' preferences for different surfaces and application features—our current study is a necessary first step toward this goal.

6 Conclusion

As technology becomes embedded in our homes, it is crucial to question the relationship between the physical surfaces we use daily and the potential of new digital UI platforms that may cohabit with these spaces. In this work, we focused on understanding how user experience during MR textured surface interaction is affected when using alternative interactions to address interaction challenges. We designed several applications with user-controlled alternative interactions and ran a qualitative study with 40 participants. We found that alternative interactions have the potential to improve user experience by supporting individual user needs during MR textured surface interaction. However, the design of future technology must consider the relationship that exists between the person, the task, and the surface being used. We provide design insights to help guide future development that will help to maintain a good user experience.

Other insights related to challenges and opportunities were: (1) the contrast between the real world and UI elements must be considered; (2) MR UIs obscure surface properties that users may find unpleasant; (3) tasks deemed as a high priority may not require automatic adaption due to user-perceived speed requirements, and (4) alternative interactions for textured surface interaction can be automatically implemented but should be user-led in the first instance.

Our recommendations can be used to improve the overall usability and user experience of future applications involving textured surface interaction.

Acknowledgments. We thank the participants of our study.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Ashtari, N., Bunt, A., McGrenere, J., Nebeling, M., Chilana, P.K.: Creating augmented and virtual reality applications: current practices, challenges, and opportunities. In: Proceedings of CHI, pp. 1–13. ACM, New York, NY, USA (2020). https://doi.org/10.1145/3313831.3376722
2. Bau, O., Poupyrev, I.: REVEL: tactile feedback technology for augmented reality. ACM Trans. Graph. **31**(4), 1–11 (2012). https://doi.org/10.1145/2185520.2185585
3. Bau, O., Poupyrev, I.: Revel: Tactile feedback technology for augmented reality. ACM Trans. Graph. **31**(4), 89:1–89:11 (2012). https://doi.org/10.1145/2185520.2185585
4. Bonanni, L., Lee, C.-H., Selker, T.: Cooking with the elements: intuitive immersive interfaces for augmented reality environments. In: Costabile, M.F., Paternò, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1022–1025. Springer, Heidelberg (2005). https://doi.org/10.1007/11555261_95
5. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006). https://doi.org/10.1191/1478088706qp063oa
6. Buchmann, V., Violich, S., Billinghurst, M., Cockburn, A.: Fingartips: gesture based direct manipulation in augmented reality. In: Proceedings of GRAPHITE, pp. 212–221. ACM, New York, NY, USA (2004). https://doi.org/10.1145/988834.988871
7. Chamunorwa, M., Wozniak, M.P., Vöge, S., Müller, H., Boll, S.C.: Interacting with rigid and soft surfaces for smart-home control. Proc. ACM Hum. Comput. Interact. **6**(MHCI), 3546746 (2022). https://doi.org/10.1145/3546746
8. Chan, V.H., Chan, Y.C., Peng, P.H., Cheng, L.P.: Texelblocks: dynamic surfaces for physical interactions. In: Proceedings of CHI EA. ACM, New York, NY, USA (2021)

9. Cheng, Y., Yan, Y., Yi, X., Shi, Y., Lindlbauer, D.: Semanticadapt: optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In: Proceedings of UIST, pp. 282–297. ACM, New York, NY, USA (2021). https://doi.org/10.1145/3472749.3474750
10. Döring, T.: The interaction material profile: understanding and inspiring how physical materials shape interaction. In: Proc.eedings of CHI EA, pp. 2446–2453. ACM, New York, NY, USA (2016). https://doi.org/10.1145/2851581.2892516
11. Fu, Y., Hu, Y., Sundstedt, V.: A systematic literature review of virtual, augmented, and mixed reality game applications in healthcare. ACM Trans. Comput. Healthcare **3**(2), 1–27 (2022). https://doi.org/10.1145/3472303
12. Garrido, A., Firmenich, S., Rossi, G., Grigera, J., Medina-Medina, N., Harari, I.: Personalized web accessibility using client-side refactoring. IEEE Internet Comput. **17**(4), 58–66 (2012). https://doi.org/10.1109/MIC.2012.143
13. Giaccardi, E., Karana, E.: Foundations of materials experience: an approach for HCI. In: Proceedings of CHI, pp. 2447–2456. ACM, New York, NY, USA (2015)
14. Gorman, B.M., Crabb, M., Armstrong, M.: Adaptive subtitles: preferences and trade-offs in real-time media adaption. In: Proceedings of CHI. CHI 2021, ACM, New York, NY, USA (2021). https://doi.org/10.1145/3411764.3445509
15. Häkkila, J., Colley, A.: Towards a design space for liquid user interfaces. In: Proceedings of NordiCHI, pp. 34:1–34:4. ACM, New York (2016). https://doi.org/10.1145/2971485.2971537
16. Häkkila, J., He, Y., Colley, A.: Experiencing the elements – user study with natural material probes. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9296, pp. 324–331. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22701-6_24
17. Harrison, C., Benko, H., Wilson, A.D.: Omnitouch: wearable multitouch interaction everywhere. In: Proceedings of UIST, p. 441–450. ACM, New York, NY, USA (2011)
18. Hughes, C.E., Stapleton, C.B., Hughes, D.E., Smith, E.M.: Mixed reality in education, entertainment, and training. IEEE Comput. Graphics Appl. **25**(6), 24–30 (2005). https://doi.org/10.1109/MCG.2005.139
19. Iravantchi, Y., Zhao, Y., Kin, K., Sample, A.P.: Sawsense: using surface acoustic waves for surface-bound event recognition. In: Proceedings of CHI. ACM, New York, NY, USA (2023). https://doi.org/10.1145/3544548.3580991
20. Jimenez-Mixco, V., de las Heras, R., Villalar, J.-L., Arredondo, M.T.: A new approach for accessible interaction within smart homes through virtual reality. In: Stephanidis, C. (ed.) UAHCI 2009. LNCS, vol. 5615, pp. 75–81. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02710-9_9
21. Jung, H., Stolterman, E.: Material probe: Exploring materiality of digital artifacts. In: Proceedings of TEI, pp. 153–156. ACM, New York, NY, USA (2010). https://doi.org/10.1145/1935701.1935731
22. Kirkpatrick, A., O'Connor, J., Campbell, A., Cooper, M.: Web content accessibility guidelines (WCAG) 2.1 (2018). www.w3.org/TR/WCAG21/. Accessed 08 Dec 2022
23. Koike, H., Sato, Y., Kobayashi, Y., Tobita, H., Kobayashi, M.: Interactive textbook and interactive Venn diagram: natural and intuitive interfaces on augmented desk system. In: Proceedings of CHI, pp. 121–128. ACM, New York, NY, USA (2000)
24. Kristensson, P.O.: Discrete and continuous shape writing for text entry and control. Ph.D. thesis, Institutionen för datavetenskap (2007)
25. Lazar, J., Feng, J.H., Hochheiser, H.: Research Methods in Human-Computer Interaction. Morgan Kaufmann (2017)

26. Lee, C.H., Bonnani, L., Selker, T.: Augmented reality kitchen: enhancing human sensibility in domestic life. In: ACM SIGGRAPH 2005 Posters. ACM, New York, NY, USA (2005). https://doi.org/10.1145/1186954.1187022
27. Leonidis, A., et al.: Ambient intelligence in the living room. Sensors **19**(22), 5011 (2019). https://doi.org/10.3390/s19225011
28. Li, X., Yi, W., Chi, H.L., Wang, X., Chan, A.P.: A critical review of virtual and augmented reality (VR/AR) applications in construction safety. Autom. Constr. **86**, 150–162 (2018)
29. Lindlbauer, D., Feit, A.M., Hilliges, O.: Context-aware online adaptation of mixed reality interfaces. In: Proceedings UIST, pp. 147–160. ACM, New York, NY, USA (2019). https://doi.org/10.1145/3332165.3347945
30. Lindlbauer, D., Grønbæk, J.E., Birk, M., Halskov, K., Alexa, M., Müller, J.: Combining shape-changing interfaces and spatial augmented reality enables extended object appearance. In: Proceedings of CHI, pp. 791–802. ACM, New York, NY, USA (2016). https://doi.org/10.1145/2858036.2858457
31. Molloy, W., Huang, E., Wünsche, B.C.: Mixed reality piano tutor: a gamified piano practice environment. In: 2019 International Conference on Electronics, Information, and Communication (ICEIC), pp. 1–7. IEEE (2019). https://doi.org/10.23919/ELINFOCOM.2019.8706474
32. Montague, K., Hanson, V.L., Cobley, A.: Designing for individuals: usable touchscreen interaction through shared user models. In: Proceedings ASSETS, pp. 151–158. ACM, New York, NY, USA (2012). https://doi.org/10.1145/2384916.2384943
33. Nayar, S.K., Peri, H., Grossberg, M.D., Belhumeur, P.N.: A projection system with radiometric compensation for screen imperfections. In: ICCV Workshop on Projector-camera Systems (PROCAMS), vol. 3 (2003)
34. Peissner, M., Edlin-White, R.: User control in adaptive user interfaces for accessibility. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013. LNCS, vol. 8117, pp. 623–640. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40483-2_44
35. Pezent, E., et al.: Explorations of wrist haptic feedback for AR/VR interactions with Tasbi. In: Proceedings of CHI EA, pp. 1–4. ACM, New York, NY, USA (2020). https://doi.org/10.1145/3334480.3383151
36. Schmid, M., Rümelin, S., Richter, H.: Empowering materiality: Inspiring the design of tangible interactions. In: Proceeding of TEI, pp. 91–98. ACM, New York, NY, USA (2013). https://doi.org/10.1145/2460625.2460639
37. Shneiderman, B., Maes, P.: Direct manipulation vs. interface agents. Interactions **4**(6), 42–61 (1997). https://doi.org/10.1145/267505.267514
38. Speicher, M., Hall, B.D., Nebeling, M.: What is mixed reality? In: Proceedings of CHI, pp. 1–15. ACM, New York, NY, USA (2019). https://doi.org/10.1145/3290605.3300767
39. Sridharan, S.K., Hincapié-Ramos, J.D., Flatla, D.R., Irani, P.: Color correction for optical see-through displays using display color profiles. In: Proceedings of VRST, pp. 231–240. ACM, New York, NY, USA (2013). https://doi.org/10.1145/2503713.2503716
40. Tigwell, G.W., Crabb, M.: Household surface interactions: understanding user input preferences and perceived home experiences. In: Proceedings of CHI, pp. 1–14. ACM, New York, NY, USA (2020). https://doi.org/10.1145/3313831.3376856
41. Tigwell, G.W., Flatla, D.R., Menzies, R.: It's not just the light: Understanding the factors causing situational visual impairments during mobile interaction. In: Proceedings of NordiCHI, pp. 338–351. ACM, New York, NY, USA (2018). https://doi.org/10.1145/3240167.3240207

42. Tigwell, G.W., Menzies, R., Flatla, D.R.: Designing for situational visual impairments: Supporting early-career designers of mobile content. In: Proceedings of DIS, pp. 387–399. ACM, New York, NY, USA (2018). https://doi.org/10.1145/3196709.3196760
43. Venkatesh, A.: Digital home technologies and transformation of households. Inf. Syst. Front. **10**(4), 391–395 (2008)
44. Ventä-Olkkonen, L., AAkerman, P., Puikkonen, A., Colley, A., Häkkilä, J.: Touching the ice: in-the-wild study of an interactive icewall. In: Proceedings of OzCHI, pp. 129–132. ACM, New York, NY, USA (2014). https://doi.org/10.1145/2686612.2686630
45. Villamor, C., Willis, D., Wroblewski, L.: Touch gesture reference guide. Touch Gesture Reference Guide (2010). http://www.lukew.com/touch/
46. Williams, E.: Experimental designs balanced for the estimation of residual effects of treatments. Aust. J. Chem. **2**, 149–168 (1949). https://doi.org/10.1071/CH9490149
47. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceeding of CHI, pp. 1083–1092. ACM, New York, NY, USA (2009). https://doi.org/10.1145/1518701.1518866
48. Zhong, Y., et al.: Smart home on smart phone. In: Proceedings of UbiComp, pp. 467–468. ACM, New York, NY, USA (2011). https://doi.org/10.1145/2030112.2030174

Author Index

J. Y. C. Chen and G. Fragomeni (Eds.): HCII 2025, LNCS 15790, pp. 377–379, 2025.
https://doi.org/10.1007/978-3-031-93715-6