Mixed Reality Multi-user Asymmetric Telecollaboration

by

Faisal Zaman

A thesis submitted to the
Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Faculty of Computer Science and Engineering.

Victoria University of Wellington
2024

Supervisory Committee

Primary Supervisor
Associate Professor Taehyun Rhee

Secondary Supervisor
Dr Craig Anslow

School of Engineering and Computer Science,
Victoria University of Wellington

To my parents

Acknowledgments

I would like to express my gratitude to my supervisors, Taehyun Rhee and Craig Anslow, for their unwavering support, guidance, and encouragement throughout my PhD journey. Their valuable insights and feedback have been instrumental in shaping my research work and keeping me focused on my goals.

I'd like to thank Jacob Young (aka *sensei*) for his valuable insights and support. Thanks for indulging all my eccentric questions and enduring the agony of reading my poorly written drafts. Thanks to Andrew Chalmers for all his help. We've been on more trips and conferences than any rock bands, and somehow, you've managed to keep us on the path amidst all the jet lag and distractions.

I'm deeply indebted to the following people, listed in no particular order: Nadia Pantidi, Rafael Kuffner dos Anjos, Daniel Medeiros, Ian Loh, Nicolas Vergnaud, Warren Butcher, Lohit Petikam, Christian Suppan, Simon Finnie, Rongsen Chen, Richard Roberts, Stephen Thompson, Siyun Thompson, Benjamin Powley, Holly Downer, Diana Siwiak, and Junhong Zhao.

Thank you, Dad, for the morning calls. Thank you, Mom, for supporting me unconditionally in all my crazy pursuits. Finally, thank you, Fahad, for always being on my side (even when we disagree).

Lastly, thanks to the support, assistance, and funding provided by the Computational Media Innovation Centre (CMIC) throughout this PhD. The work in this thesis was supported by the Entrepreneurial Universities Programme, funded by the Tertiary Education Commission (TEC), New Zealand.

Publications

The following work was published as a result of the investigations performed in the course of this thesis:

- *(Submitted to ISMAR 2024)*
  F. Zaman, C. Anslow, A. Chalmers, T. Rhee. "CollabXR: A Systematic Review of Research, Applications, and Opportunities in XR Collaboration."
- **F. Zaman,** C. Anslow, A. Chalmers, T. Rhee. "MRMAC: Mixed Reality Multi-user Asymmetric Collaboration." In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Sydney, Australia, 2023.
- **F. Zaman,** C. Anslow, T. Rhee. "Vicarious: Context-aware Viewpoints Selection for Mixed Reality Collaboration." In Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST), Christchurch, New Zealand, 2023.
- **F. Zaman.** "[DC] Improving Multi-User Interaction for Mixed Reality Telecollaboration", in IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 940-941), 2022. IEEE.

Abstract

Creating a seamless mixed reality (MR) collaborative environment to facilitate natural collaboration between multiple remote and local users within a shared physical space is challenging. Existing collaboration technologies, such as 2D and 3D videoconferencing, as well as virtual reality (VR) solutions, fall short of delivering a fully immersive, real-time, and cohesive collaborative experience for multiple users. Remote users often feel disconnected from the shared physical space. This thesis aims to overcome these limitations by developing a multi-user immersive MR system. The primary objective is to enable remote users to perceive the physical environment and collaborate effectively with local users in real-time. First, a framework is described that enables local users to live-stream their physical environment to multiple remote users while seamlessly blending 3D virtual assets. This provides a more immersive and interactive collaborative space with minimal latency. The system seamlessly integrates these elements, allowing remote and local users to collaborate within physical spaces as if they were present together. Second, the effectiveness of the system is evaluated through performance assessments and user studies. The results of these evaluations demonstrate the system's ability to induce various forms of presence among participants in mixed-reality collaboration. These findings enable a second application that employs a context-aware method for selecting and dynamically switching or highlighting optimal viewpoints based on user actions and the current context. This allows participants to explore the collaboration space from different perspectives. The application is also evaluated, and the results of the evaluation provide insight into enabling effective mixed-perspective collaboration within the multi-user MR system. These findings also show reduced cognitive load and improved task understanding among participants. Thus, it demonstrates its effectiveness in enhancing the collaborative experience.

Contents

Acknowledgments **iii**

Publications **v**

Abstract **vii**

1 Introduction **1**

1.1 Telecollaboration in Physical Space 5
1.2 Collaborative Mixed Reality Environment 6
1.3 Problem Statement 7
1.4 Contributions 11
1.5 Thesis Structure 12

2 Background and Related Work **15**

2.1 Immersive Telepresence 15
2.2 Computer-Supported Cooperative Work 16
2.3 AR Telecollaboration 17
2.4 VR Telecollaboration 18
2.5 Asymmetric AR-VR Telecollaboration 19
2.6 360-degree Panorama 19
2.7 Review of Existing Literature 22
2.7.1 Scope of Survey 25
2.7.2 Methodology 27
2.7.3 Taxonomy 28
2.8 Results 42
2.8.1 Discussion 45

3 MRMAC: Mixed Reality Multi-user Asymmetric Collaboration System 53
3.1 Introduction 53
3.2 Related Work 55
3.2.1 Multi-user Collaboration in VR 55
3.2.2 Multi-user Collaboration in AR 56
3.2.3 Multi-user Asymmetric Collaboration in MR 57
3.3 MRMAC Design 58
3.3.1 Asymmetric Telecollaboration System 58
3.3.2 Multi-user Asymmetric Collaboration (MAC) 59
3.3.3 Key Design Features 61
3.4 MRMAC Implementation 62
3.4.1 Architecture Overview 62
3.4.2 Communication 65
3.4.3 Awareness 66
3.4.4 Streaming and Synchronization 68
3.5 Conclusion 72

4 Evaluation of MRMAC 73
4.1 MRMAC Evaluation: System 73
4.1.1 System Setup 73
4.1.2 Results 75
4.1.3 Discussion 76
4.2 MRMAC Evaluation: User Role 78
4.2.1 Design and Hypotheses 78
4.2.2 Study Tasks 80
4.2.3 System Setup 82
4.2.4 Participants 82
4.2.5 Procedure 82
4.2.6 Measures 84
4.2.7 Results 85
4.2.8 Interview Insights 91
4.2.9 Discussion 96
4.3 MRMAC Evaluation: Baseline Comparison 99
4.3.1 Design and Methodology 99
4.3.2 Experimental tasks 100
4.3.3 Experimental Conditions and Setup 100

4.3.4 Procedure 102
4.3.5 Measures and Hypotheses 102
4.3.6 Participants 103
4.3.7 Results 103
4.3.8 Discussion 112
4.4 Conclusion 116

5 Vicarious: Context-aware Viewpoints Selection for Mixed Reality Collaboration 119
5.1 Introduction 120
5.2 Related Work 122
5.2.1 Viewpoint Sharing in Telepresence 122
5.2.2 Mixed-Perspectives Sharing 123
5.2.3 Viewpoint Sharing and Transition 124
5.3 Vicarious 125
5.3.1 Design Overview 125
5.3.2 Context-aware Viewpoint Selection 127
5.3.3 Visualization and Interaction 131
5.3.4 Networking 132
5.4 User Study 133
5.4.1 Hypotheses 133
5.4.2 Participants 134
5.4.3 Experimental Conditions 134
5.4.4 Experimental Tasks and Setup 135
5.4.5 Procedure 136
5.4.6 Measures 136
5.5 Results 137
5.5.1 Social Presence 137
5.5.2 Spatial Presence 138
5.5.3 System Usability 139
5.5.4 Completion Time and Task Load 140
5.5.5 Simulator Sickness 141
5.5.6 Preferences 142
5.6 Discussion 143
5.6.1 Open-ended Feedback 145
5.6.2 Communication Patterns 147
5.6.3 Design Implications 147

5.6.4 Limitation and Future Work . . . 148
5.7 Conclusion . . . 149

6 Conclusion 151
6.1 Summary of the Contributions . . . 152
6.2 Limitations . . . 154
6.3 Future Work . . . 155
6.4 Conclusion . . . 156

A Ethics 159

B User Evaluation Questionnaire 165

Bibliography 173

List of Figures

1.1 An immersive telecollaboration scenario. 1
1.2 Reality-Virtuality Continuum. 2
1.3 Extended Reality Market Size, 2021 to 2030 (USD Billion). 3
1.4 The research topic for this thesis lies at the intersection of XR and CSCW, as well as XR, Visual Analytics, and CSCW. 9

2.1 Illustration of global and local coordinates for a 360° camera with θ and ϕ to indicate the direction of a point on a sphere. 20
2.2 A total of 557 papers that employed XR collaboration research between 2000 and 2023 were classified by (left-to-right): system input type, device type, interaction modalities, viewpoint sharing, collaboration dynamic and topology, and evaluation method in user studies. 23
2.3 The workflow for the systematic literature review. 29
2.4 (a) Stitching a Panorama using multiple images, (b) Point cloud sampled from the real surface, (c) Holograms, also called volumetric displays, are composed of voxels (volume + pixel). 31
2.5 Interaction techniques in XR collaboration: Drawing annotation (a), viewpoint conversion (b), pointing (c), hand gestures (d), ray to objects (e), overlay icons and label text on top shared objects (f), gaze (g). Diagram adapted from [99]. 33
2.6 Visualization of Ego, Exo, and World-in-Hand viewpoints. 35
2.7 Clustering results from the text mining. 39
2.8 Distribution of 395 papers that were further assessed to determine the current XR collaboration trend are organized by year of publication, publication venues, citation impact, type of publication, and area of application. 43
2.9 Distribution of publications across the top 10 venues. 44
2.10 Distribution of publications by year. 45

2.11 Device types (a), User interaction types (b), Evaluation criteria (c) 45
2.12 Author Contribution Visualization: Bubble radius proportional to number of publications. 46
2.13 Distribution of User Studies Based on Participant Group Size 47

3.1 The MRMAC system, including the local AR-Hosts and remote VR-Travelers, 360° camera, controls, and display setup. 54
3.2 Different types of collaboration setup in a multi-user asymmetric MR collaboration. Remote space to Local space (A) One-to-One ($1 \rightleftarrows 1$), (B) One-to-Many ($1 \rightleftarrows n$), (C) Many-to-One ($n \rightleftarrows 1$), (D) Many-to-Many ($n \rightleftarrows n$). 60
3.3 MRMAC: overview of the system architecture and the network data flow, demonstrating how the primary processing pieces are mapped to the VR-Travelers and AR-Hosts. 64
3.4 Communication and awareness cues in MRMAC: Annotation, 3D pointer, and avatar embodiment in collaborative environments. 66
3.5 Multiple avatar position control includes (a) Applying offset compensation to shift the avatar from its original position, and (b) adjusting the drawing of the 3D Pointer based on the offset compensation. 67
3.6 Personalized avatar generation process. 69
3.7 View sharing in MRMAC: AR-Hosts can see the viewpoint of VR-Travellers through a picture-in-picture (PiP) window. 70

4.1 System performance metrics in milliseconds (ms) or frames per second (FPS) for each performance metric. 75
4.2 Scalability of MRMAC's network latency as the number of users increases. . . 76
4.3 MRMAC Scenario: At a physical work site, local AR user (a) Adam (host) initiates a collaboration session using MRMAC in which remote VR users (David and Jane) (b, c) collaborate with Adam. MRMAC broadcasts the captured real-world environment to multiple remote users. Teleporters' identities (represented as avatars) share audio-visual cues to augment their remote collaboration. 79
4.4 User interface and MRC space in the *Training task* (A and B); user interface and MRC space in the *Main task* (C and D). 80
4.5 Lego and domino blocks structures: (a) virtual replicas, (b) corresponding structures made using physical blocks. 81
4.6 Room layout and placement of the VR-Travelers and AR-Host. 83

4.7 User study setup for MRMAC prototype. Left: multiple remote users are exploring the local user environment. Right: the local user in front of the table, identifying the specific LEGO pieces that the remote user instructs to select. The avatar of the remote user positioned within the shared space, guiding the local user using a 3D pointer. 84
4.8 Participants' spatial presence scores comparison between AR and VR users. . 86
4.9 Social presence between AR and VR users. 87
4.10 Participants' SUS scores comparison between AR and VR users. 88
4.11 Additional system usability questionnaire (Q1 and Q2 are positive aspects while Q3 and Q4 are negative aspects). 88
4.12 Participants' responses on the NASA-TLX questionnaire ($0 : verylow$ to $100 : veryhigh$, lower is better). 90
4.13 Increase in the simulator sickness score (pre- and post-exposure). 91
4.14 Statistical results of user preference questionnaire. 92
4.15 Comparison of baseline setups in MRMAC evaluation, including the local AR-Hosts and remote VR-Travelers, 360° camera, controls, and display setup. 99
4.16 Overview of the User Study. 101
4.17 Spatial Presence scores across different conditions. 104
4.18 Social Presence scores across different conditions. 105
4.19 The task completion time for each condition. 106
4.20 NASA-TLX score from 0 (low workload) to 100 (high workload) (lower is better) 107
4.21 SUS results across different conditions. 108
4.22 Users ranked the three conditions based on preference for presence, partner presence, fun, task support, and overall preference (Q1-Q5). 110

5.1 Vicarious teaser: Two remote users direct a local user in a physical task. 121
5.2 Schematic view of egocentric (left) and exocentric (right) viewpoint. 125
5.3 Illustration of user FoV. As shown on the sphere. 126
5.4 Pipeline of the pre-processing step. The pre-processing step first computes optical flow from an input 2D video, estimates saliency based on the optical flow, and finally, the binary threshold of the saliency map 128
5.5 Visual cues: (a) annotate (b) gestures (c) 3D pointer. 130
5.6 Visualization and Interaction options include: (a) Smooth follow (b) Pinned to a fixed location (c) Zoom in/out (d) Resize. 131
5.7 Indication of spatial context: Arrow indicator 132

5.8 Viewpoint perspectives of local and remote users, highlighting remote user C's viewpoint in the GS condition. 133
5.9 Experimental conditions: (a) Manual Selection (b) Guided Selection (c) Automatic Selection. 135
5.10 Lego and domino blocks structures: (a) virtual replicas, (b) corresponding structures made using physical blocks. 136
5.11 Social Presence results in each subscale. 138
5.12 Spatial Presence results in each subscale. 139
5.13 System usability results (80.3 or higher is considered good, 68 and above classified as average, and below 51 considered poor. 140
5.14 Average task completion time for each condition. 141
5.15 NASA-TLX score (0: very low to 100: very high), overall score is shown to the right. 142
5.16 Increase in the simulator sickness score (pre- and post-exposure) 143
5.17 User preference among the four conditions. 144

List of Tables

2.1 The collaboration scenarios targeted by this research lie in the *same time, same place* and *same time, different place* quadrant of the Time/Space Matrix. 16
2.2 Summary of existing review papers involving collaboration in AR/VR/MR. . 27
2.3 Empirical derivation of the topics for the publication type and application areas to improve field encapsulation. 40

4.1 System Performance Measurements 75
4.2 Task completion times (in seconds) and error frequency, each with standard deviation in parentheses. 89
4.3 Error frequency in the user study (mean with standard deviation in parentheses). 106
4.4 Group Statistics and Post-Hoc Test Results of the SUS. 108
4.5 Average ranking value of user preferences for each question, role, and condition. 111

5.1 A high-level comparison of previous work outlined based on viewpoint sharing. 124
5.2 Decision Matrix for Optimal Viewpoint Selection 130

Chapter 1

Introduction

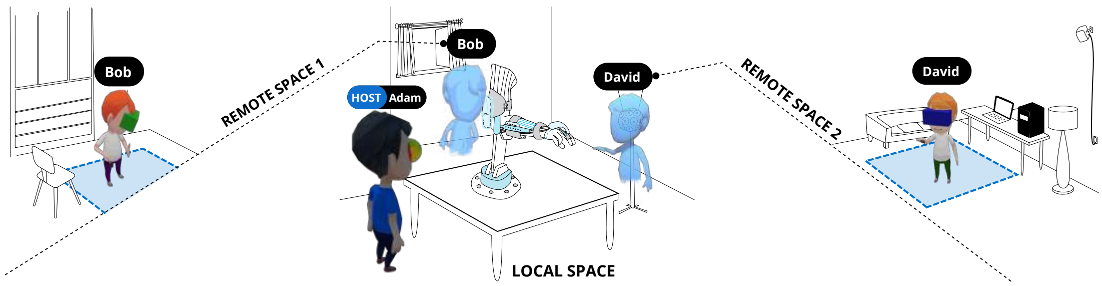


Figure 1.1: An immersive telecollaboration scenario.

> *"Imagine you are part of a team involved in assembling a complex piece of machinery that requires input from different specialized experts located in different cities or countries. During an on-site inspection, you realize significant structural modifications are necessary, but it can take hours or even days for on-site experts to arrive and diagnose the problem. To stay on time and within budget, seamless collaboration between the remote and on-site teams becomes imperative, despite the geographical distances. Conventional video conferencing may not suffice, as physical presence in the same space as the equipment is necessary to achieve situational awareness, identify the problem, and troubleshoot it effectively. A collaborative meeting space where remote experts can join, access designs and analytic tools, and share data is essential. How can we create such a space? How can we bring remote experts to that space? What would this space look like?"*

While this is a specific example, it illustrates the fundamentals of the immersive telecollaboration concept, which plays a role in many different applications. The term *'ideal space'* in this context refers to an immersive telecollaboration environment (see Figure 1.1). Immersive telecollaboration combines immersive technologies, such as Virtual Reality (VR) and Augmented Reality (AR), with remote collaboration tools to enable multiple users to work together as if they were in the same physical location. This allows geographically distant users to come together within the shared virtual space, providing remote users with a strong sense of telepresence. The term *'telepresence'* is defined as the ability for individuals to feel fully present in a remote location through the utilization of audio, video, and other communication mediums [203].

The concept of immersive telecollaboration is in line with the Mixed Reality (MR) continuum [147], which describes a gradual shift in the user experience from the physical environment to a virtual computer-generated world (see Figure 1.2). In the context, the concept of telepresence assumes a central role in creating a seamless transition from physical to virtual spaces in MR environments. On one end of the continuum, AR presents a view of the physical world where visual and auditory senses are augmented with virtual information that enhances the user experience [11]. On the other end, VR completely replaces the user's view of the physical world with a computer-generated representation, giving unique experiences that are unattainable within the limits of their physical form [27]. By blending different realities, MR technologies enable new opportunities for collaborative human experiences, with remote mediation being a key focus [142, 230, 237, 251]. MR, with its unique feature of supporting viewpoint independence, facilitates mutual grounding in communication, where collaborators can build a shared understanding of each other's actions and expectations by working together [48]. This capability highlights MR's potential to become the next big computing platform, a notion supported by Precedence Research, as market is projected to hit around USD 345.9 billion by 2030, growing at a CAGR of 33.09% during the forecast period 2022 to 2030 (see Figure 1.3) [190].

Researchers have explored ways to shared telepresence and telecollaboration among remote users for decades. Early examples include Office of the Future [184], the TELEPORT system [33], Viewport [271] which envisioned transforming conventional office environ-

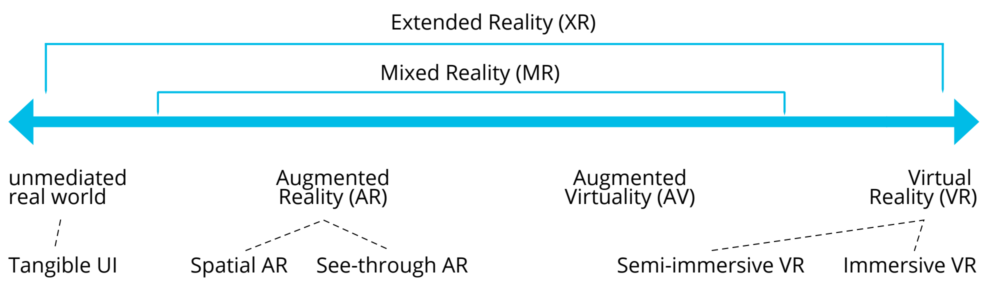


Figure 1.2: Reality-Virtuality Continuum adapted from Milgram & Kishino [147].

ments and redefining the way people work and collaborate. These projects aimed to combine 3D models of physical spaces with real-time computer-generated imagery, creating realistic and interactive virtual workspaces. These concepts have undergone refinement over many years of collaborative efforts and have had a significant impact on the development of VR and AR technologies. Several commercial products have emerged as state-of-the-art enterprise solutions [214, 146]. These products often go beyond software and incorporate custom-designed furniture and specially designed configurations of hardware components like displays, cameras, microphones, and speakers to create immersive and seamless communication experiences in remote spaces [47, 90, 213]. However, while these solutions may be acceptable for certain tasks, they restrict users from fully exploring their partner's physical space and lock their view in the direction of the camera which significantly reduces the users' sense of co-presence [93, 115], potentially leading to slow tasks completion [68, 70, 274].

It's also important to recognize that in many instances, the remote individual possesses knowledge about tasks or physical artifacts. This expertise creates an inherent 'asymmetry' due to the physical nature of the task and the distributed working arrangement because their specialized knowledge or skills are not readily available to others in the distributed team, leading to discrepancies in understanding and decision-making [179]. This challenge becomes particularly relevant in telecollaboration scenarios where current technologies often struggle to adequately support the transfer of specialized knowledge or skills between

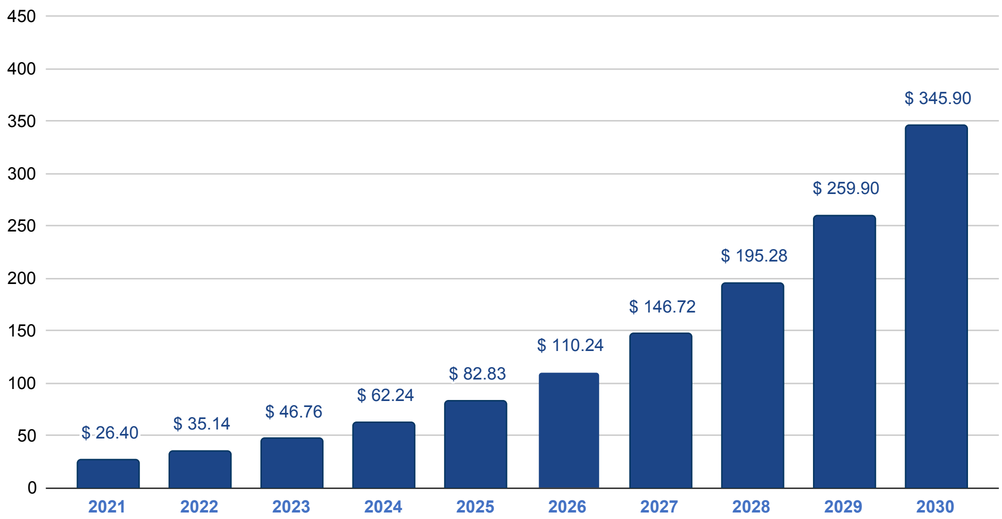


Figure 1.3: Extended Reality Market Size, 2021 to 2030 (USD Billion).

collaborators, leading to inefficiencies in effective collaboration. Remote users, not physically present with the artifact, thus struggle with spatial awareness and presence [233]. This can lead to difficulties in accurately positioning or interacting with the artifact from a distance, hindering effective collaboration and coordination. Additionally, issues such as latency, connectivity problems, delays in communication, limited visibility, and restricted access to non-verbal cues can further disrupt the collaboration, coordination, and interaction experience, ultimately decreasing usability [119]. Although video-mediated communication provides visual access, but lacks the nuanced non-verbal cues essential for facilitating effective interactions, both in co-located and remote contexts, especially for complex physical tasks (e.g., inspection, repair, or maintenance) [45]. The paradigm of MR technology offers solutions to address these limitations due to its natural interface and the potential for seamless integration with the real world [227, 94, 7]. MR extends beyond traditional audio-video interactions by integrating virtual elements into the physical world or augmenting reality with digital content. Asymmetric setups leveraging AR and VR technologies not only support addressing different roles and interactions but also enable multimodality to provide remote guidance or assistance in various contexts [46, 238]. For instance, an AR user may use one type of Head-Mounted Display (HMD) in a physical environment, while a VR user might wear a different HMD and interact with a virtual representation of that same physical environment.

While asymmetric telepresence systems have significantly evolved over recent years, they are mostly designed to support only two users simultaneously [161, 166, 237, 193]. Human experiences of social interactions typically involve more than one other person. A rich exploitation of telepresence should allow interactions with multiple people. Therefore, an investigation into the user experience of a telepresence system involving more than two users simultaneously engaging in various modes (physical, augmented, and virtual) becomes necessary, particularly in contexts that prioritize human activity and the fidelity of the virtual environment. Creating an immersive collaborative space to enable such seamless collaboration between multiple remote and on-site teams requires a combination of hardware, software, and network infrastructure. We want to see and interact with remote collaborators, as naturally as we do when we are in the same physical room: gesturing, pointing, waving, and using all of the subtle nuances of both verbal and nonverbal communication.

In recent years, several commercial solutions have emerged that address some of the issues described so far, such as Microsoft Remote Assist [1], which integrates the physical and digital worlds to facilitate collaboration in multi-user settings. However, in this sys-

[1] https://dynamics.microsoft.com/en-us/mixed-reality/remote-assist/

tem, remote users have limited physical presence and interaction capabilities. Additionally, solutions like Microsoft Mesh [2] allow collaborators to convene in a shared virtual space for collaboration with multiple users. However, this platform does not enable users to see their collaborators' physical spaces. As a result, telepresence systems, especially those tailored for collaboration in physical spaces where users can view collaborators' spaces in real-time, remain limited and typically facilitate one-to-one communication [193].

Therefore, this thesis aims to explore how to transform that shared virtual space into a collaborative physical space, allowing remote participants to see the local user's surroundings and join in to guide local users. To achieve this objective, the thesis investigates the development of technology that facilitates the described asymmetric local and remote interactions in physical space. It examines the design and implementation of technologies that enable remote users to experience telepresence, support both non-verbal and verbal communication cues, and provide viewing access between spaces. In the following sections Section 1.1 and Section 1.2, the goals of telecollaboration in physical space and how they can be addressed by a collaborative MR environment are outlined. Then, in Section 1.3, the specific problems addressed in this thesis are defined.

1.1 Telecollaboration in Physical Space

Telecollaboration in physical space involves facilitating collaboration and communication between individuals who are geographically separated but working in the same physical environment [55]. This often occurs in settings where teams and individuals need to work together across geographical boundaries. As demonstrated in previous studies [51], video- and audio-mediated communication cues can improve coordination, maintain situational awareness (which refers to understanding elements in the environment, their context, and anticipating future actions or developments [219]), and facilitate conversational grounding in remote collaborative physical tasks [68, 120, 180]. However, despite these advancements, traditional methods still lack important non-verbal cues present in collocated work, such as gaze direction, gesture cues, and depth perception of the physical task environment [7, 121, 254, 275].

MR systems, which merge the physical and virtual worlds through the use of AR or VR technologies, allow participants to engage in a shared environment that combines digital elements with the physical space. Thus, MR has the potential to improve remote collaborative performance by enabling efficient deictic references and acknowledgments through the sharing of AR annotations or a cursor pointer on the shared task space view [7, 69, 255],

[2] https://www.microsoft.com/en-us/microsoft-teams/microsoft-mesh

as well as gaze and gesture cues [166, 169]. This technology can reduce a user's task response time and mental workload of procedural tasks in numerous applications (e.g., medicine [16], manufacturing [253] and assembly [82]). Moreover, MR naturally support collaboration in 3D virtual-real fusion environments, enabling the seamless integration of augmented communication cues with real-world working scenarios [250, 252, 246]. These technologies have also created opportunities for novel, natural, and intuitive interaction interfaces, such as gaze, gestures, and Tangible User Interfaces (TUI), thereby enhancing the user experience [17, 7, 169, 162].

Telecollaboration in MR often involves asymmetric interactions, where participants may have varying levels of access to information or control within the collaborative environment [259]. This asymmetry can be intentional, such as in educational settings where a teacher guides students through an MR experience [50, 236], or it can arise naturally in professional settings where experts collaborate with novices [193]. To facilitate this in asymmetric remote collaboration, the task environment is captured by spatial mapping and imaging technologies, providing remote collaborators with access to the collaboration space [175]. With the widespread availability of high-speed data connections and various sensors, the potential for sharing users' physical space for remote collaboration has greatly increased. Therefore, the MR-based remote collaborative platform has great potential to enable users to share communication cues and improve user experience and task performance [84, 175, 173, 111, 45], as well as share empathy with each other and maintain awareness of each other's activities [63, 169, 117, 162]. The advantages of MR have attracted academic researchers and industrial developers to explore remote collaborative applications for physical tasks. Numerous remote collaborative architectures, models, methods, and systems have been proposed and developed to enhance the user experience and task performance. These include sharing live panoramas, annotations, 3D pointers, avatars, gestures, and gaze cues, as demonstrated in platforms such as JackIn [104, 117], Showme [5], RemoteFusion [2], Mini-Me [173], BeThere [212], 3D helping hands [227, 239], Vishnu [123], Virtual replica [274], AlphaRead [42], and others.

1.2 Collaborative Mixed Reality Environment

The Mixed Reality Collaborative (MRC) space represents a convergence of physical and virtual worlds, where VR and AR technologies seamlessly blend to create a shared, immersive environment for remote collaboration. This dynamic environment leverages the strengths of both VR and AR to offer a holistic telecollaboration experience. Within the MRC space, VR and AR users assume distinct yet complementary roles, each contributing to the collaborative experience:

- *VR Users*: VR users immerse themselves in a fully digital environment, often characterized by realistic 3D simulations. They typically employ hand controllers or even full-body tracking suits to interact with objects and navigate the virtual world. VR users often take on roles that require spatial awareness, such as manipulating 3D models or simulating physical tasks. Their primary advantage lies in the complete immersion and isolation from the physical world, enabling deep focus and presence.
- *AR Users*: AR users, on the other hand, remain anchored in the physical world while overlaying digital information and objects. They view the real environment through AR glasses or devices, enriching their surroundings with digital content. AR users are well-suited for tasks that require context awareness, like annotations or real-time information retrieval. AR's strength lies in its ability to blend digital and physical elements seamlessly, making it an ideal tool for enhancing situational awareness.

In MRC space where knowledge of the local space is crucial for task performance (e.g., remotely conducting architectural assessments or overseeing construction progress), collaboration systems that integrate VR and AR offer significant benefits. Combining AR and VR into one system allows for a space where local AR users interact with remote VR users who can 'teleport' into the local space.

1.3 Problem Statement

Collaborative MR with asymmetric interfaces poses novel challenges to creating Mixed Reality Collaborative (MRC) space where remote and local users can effortlessly work together. In the MRC space, both a high sense of presence and co-presence are crucial for effective collaboration. A high sense of presence immerses users fully in the virtual environment, enhancing their engagement and interaction, while co-presence fosters a shared sense of being present with other collaborators, facilitating communication, coordination, and collaboration. In addition to users' sense of presence in the MRC space, spatial presence is also crucial. Spatial presence refers to the subjective sensation of being physically present in a mediated environment, such as a VR or AR setting, despite the actual physical absence from that environment [204]. Examples of design challenges include enabling users to make spatial references and to have a sense of awareness of other user's interaction with the MRC space (i.e., workspace awareness [85]), which are commonly associated with increased task performance [85, 173, 174]. Moreover, it is essential to explore the effects of asymmetry in user experience and collaborative processes [20, 236]. To ensure effective collaboration, it is important that every user can collaborate effectively regardless of the devices they are able to access. In this context, effective collaboration requires overcoming barriers posed by diverse devices and interfaces to ensure seamless communication, coordination, and cooperation among users. Creating an inclusive and accessible collaborative environment is crucial, enabling all participants to actively engage and contribute regardless of their technological setup. Additionally, understanding the complexity of interaction among asymmetric interfaces could enhance or enable inherently asymmetric collaborative scenarios, such as remote assistance [212, 251], training scenarios [234, 202] for co-located setups facilitating insightful comparisons [77].

Telepresence systems strive to create high-fidelity and coherent representations, aiming to make users feel as if they are truly present in remote environments [193]. This involves rendering both real-world and virtual elements with high fidelity to ensure visual consistency. For example, virtual objects integrated into real-world scenes should have realistic lighting and cast shadows accurately onto physical surfaces. Additionally, to induce a sense of presence for remote users, allowing them to feel as though they are physically present in the remote location, even though they are not physically there is crucial. This is also an important aspect of video communication technology. A crucial aspect of video communication technology is the ability to sense the presence of others at a remote location [56]. In these scenarios, it is not just about conveying information; the aim is for the remote user to feel that they are 'really there' in the environment and to feel that they share the space together [100, 198]. Therefore, remote users should feel they are present in that remote location without physically being there.

This thesis begins by addressing multi-user asymmetric collaboration in MRC space, serving as a strong foundation for enabling remote users to seamlessly collaborate with local users in their physical space. To bridge this research gap, a novel server-client structure for real-time 4K video streaming is proposed, allowing multiple remote users to experience the spatial presence of the local space. The motivation behind this approach is to enable remote users to immerse themselves more fully in the shared space, inducing a high sense of spatial presence, as if being physically present alongside the local users. As the system aims to be designed for real-time and live solutions, it will particularly be well-suited for interactive applications such as games, meetings, live performances, movie productions, and concerts. Similarly, the real-time capabilities will offer immense time-saving benefits for collaborators, designers, and participants. The research topic for this thesis lies at the intersection of XR and CSCW, as well as XR, Visual Analytics, and CSCW (see Figure 1.4). The fundamental research question of the thesis can be formulated in the following question.

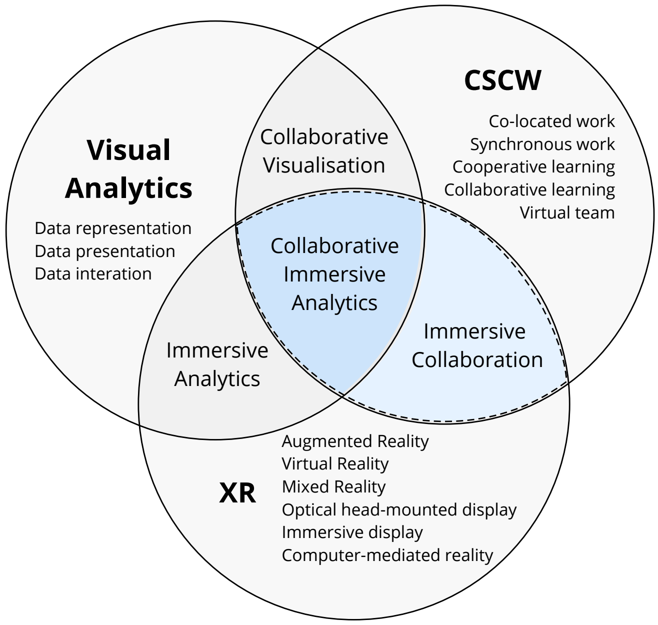


Figure 1.4: The research topic for this thesis lies at the intersection of XR and CSCW ( ), as well as XR, Visual Analytics, and CSCW ( ). Diagram adopted from [22, 156, 6].

RQ: How can multiple remote (VR) users effectively connect and collaborate with local (AR) users in the Mixed Reality Collaboration space?

To clarify the main research question, the overall aim has been divided into several additional research questions. These questions center on *teleportation, telepresence, representation, interaction, viewpoint sharing*, and *mixed-perspective* collaboration in multi-user MR collaboration. In summary, this thesis will address the following questions:

RQ1: How can multiple remote users have a high sense of presence while maintaining spatial awareness and understanding in the Mixed Reality Collaboration space? (*teleportation*)

RQ2: How can multiple remote VR users be effectively represented in the Mixed Reality Collaboration space to achieve co-presence? (*representation*)

RQ3: How can multiple VR users interact effectively in the Mixed Reality Collaboration space? (*interaction*)

RQ4: How can users' viewpoints be managed effectively to support various types of collaborative tasks in the Mixed Reality Collaboration space? (*viewpoint perspective*)

RQ5: How can multi-perspective collaboration be effectively achieved for supporting collaborative tasks? (*multi-perspective collaboration*)

To address the research questions outlined above, this thesis aims to establish a framework that facilitates effective connection, communication, and collaboration among multiple remote VR users and local AR users within a collaborative shared physical space known as MRC space. Hence, the thesis sets the following specific *objectives* for investigation:

Obj1: To conduct a systematic literature review to establish a knowledge foundation, identify gaps, and understand existing challenges and solutions in the field.
This objective serves as the foundational step for your research by thoroughly examining existing literature related to Mixed Reality Collaboration (MRC), spatial cognition, presence, representation techniques, interaction modalities, viewpoint management, and immersive analytics. The systematic review will help establish the current state of knowledge and identify gaps in the field. This directly contributes to addressing RQs 1 and 2, as it provides a comprehensive understanding of the challenges and existing solutions related to spatial awareness and representation in MRC spaces.

Obj2: To design, implement, and usability test of multi-user mixed reality remote collaboration systems (*MRMAC*)
MRMAC uses a 360° camera to capture the physical collaboration space. When designing a computer-supported collaborative system, WYSIWIS [220] may not be a proper choice in many cases. Previous research has shown how to use 360° video streaming for one-to-one collaboration [193, 232, 228], but this is not implemented for multi-user applications. Therefore, in this first step, it is necessary to develop *MRMAC* that will allow multiple users to join a collaboration session. This will also allow multiple people to collaborate together to solve a specific problem. This objective directly addresses **RQs 1, 2, and 3** by creating a system (MRMAC) that allows multiple remote VR users to collaborate effectively with local AR users. The usability evaluation assesses how users understand and interact within MRC spaces, aligning with the investigation of spatial awareness, presence, representation, and interaction modalities.

Obj3: To design, implement, and usability test a context-aware viewpoint selection method called *Vicarious* that allows multiple viewpoint sharing within *MRMAC*.
Vicarious will address the challenges of efficiently managing and selecting users' viewpoints during collaborative tasks within the MRC space. This objective is closely linked to **RQs 4 and 5**. Vicarious offers a solution for efficiently managing users'

viewpoints and provides an evaluation of various view selection methods to facilitate collaborative tasks. Through the user study, the effective manipulation of viewpoints will be explored, directly addressing research questions related to viewpoint perspective and mixed-perspective collaboration.

1.4 Contributions

The contribution of this thesis directly addresses specific concerns related to the research objectives:

- **Design and development of a novel Mixed Reality Multi-user Asymmetric Collaboration system MRMAC**, which enables multiple users to collaborate in a shared mixed reality environment. MRMAC enhances communication, interaction, and teamwork by enabling users to engage with virtual elements and interact with each other seamlessly.

- **System evaluation and two user studies in multi-user telecollaboration scenarios.** System evaluation demonstrates its low-latency synchronized communication for multiple AR and VR users. User Study 1 examines different user roles and interactions within the system, providing insights into user behavior and preferences. In User Study 2, the Multiuser MR Collaboration System is compared against a baseline system or existing collaborative tools to assess its effectiveness and advantages.

- **Design and development of a novel context-aware viewpoint sharing method Vicarious**, addresses the challenge of sharing perspectives and viewpoints among multiple users and leverages contextual information, such as user actions, object interactions, and scene context, to dynamically select optimal viewpoints.

- **User study to evaluate the impact of the context-aware viewpoint selections method** on collaboration performance and user experience. user study was conducted using an asymmetric AR-VR setup where users performed remote collaboration tasks under four distinct conditions: No-view, Manual, Guided, and Automatic selection. The results provide insights and recommendations for design implications and directions for future research.

In summary, this thesis makes significant contributions to the field of MR collaboration by designing and implementing novel practical systems (MRMAC and Vicarious), and providing empirical insights and solutions. These contributions directly address the research objectives and associated research questions, advancing our understanding and capabilities in enabling effective connectivity, communication, and collaboration between remote VR users and local AR users in MRC environments.

1.5 Thesis Structure

This chapter, **Chapter 1**, has presented a brief introduction to multi-user MR telecollaboration and the broader context in which this thesis is situated. It introduced the motivation for and the focus of the research and outlined the scope of the study. The following chapters of this thesis directly address the research problem discussed earlier. In total, this thesis comprises six chapters. This section will lay out the structure of this thesis and provide valuable suggestions for readers to navigate and comprehend its contents. The remainder of this thesis is organized as follows:

Chapter 2 serves a twofold purpose. Firstly, it provides the necessary background to understand the thesis and establishes a solid foundation for understanding the state of the art in MR collaboration, laying the groundwork for the subsequent chapters. Secondly, it presents a systematic review of existing literature on mixed reality (MR) collaboration and related topics. This chapter describes key topics, identifies challenges in the field, examines existing solutions, and discusses future trends.

Both **Chapter 3** and **Chapter 4** collectively focus on the Multi-User Mixed Reality Remote Collaboration System (MRMAC). In **Chapter 3**, the emphasis is on the novel design and implementation of MRMAC, providing detailed insights into the technical components and infrastructure that make up its hierarchical architecture. **Chapter 4** shifts the spotlight to the evaluation of MRMAC. The evaluation includes both system evaluation, critically assessing MRMAC's performance and functionality, and user studies. User Study 1 examines different user roles and interactions within the system, shedding light on user experiences and behavior. User Study 2 compares MRMAC against a baseline system to assess its effectiveness and advantages, providing empirical evidence of its capabilities.

Chapter 5 describes the design and evaluation of Vicarious, a context-aware viewpoint-sharing method built as a feature inside MRMAC, showcasing its effectiveness and potential impact on MR collaboration. The chapter also discusses the implementation process and the functionalities of Vicarious, outlining the methodology for usability analysis that was conducted to evaluate it. The chapter reports the user study results, with an emphasis on viewpoint management and collaborative task performance. Additionally, the chapter explores how Vicarious addresses challenges related to viewpoint perspective and mixed-perspective collaboration in the MRC space.

Finally, **Chapter 6** concludes the thesis by summarizing the key findings and contributions, addressing the limitations of the thesis, and discussing future directions for the use of multi-user asymmetric telecollaboration in collaborative MR applications. The chapter also discusses the evidence for a multi-user asymmetric telecollaboration system designing support for collaborative physical tasks and provides answers to the research questions posed. Furthermore, this chapter reflects on the broader implications of the research and its impact on both the research community and the wider audience, while also laying out potential avenues for possible future work.

Supplementary videos that illustrate the contributions and feature animated examples can be accessed at the following URL: https://zaps.one/thesis

Chapter 2

Background and Related Work

This chapter serves as the foundational backdrop to the core themes and concepts explored within this thesis. It provides a succinct overview of the key technological and theoretical components that underpin the research. The aim is to equip the reader with the necessary understanding of immersive teleportation, mixed reality, extended reality, 360° capture, the Time/Space Matrix, and related concepts, setting the stage for an in-depth examination of their significance and implications in subsequent sections. This foundational knowledge is critical in appreciating the context and relevance of the research presented in the following chapters.

2.1 Immersive Telepresence

Bill Buxton introduced the concept of telepresence in the domain of telecommunication [38]. Buxton's concept of shared person space encompasses the idea of mutual awareness and copresence between remote partners, which involves the use of technologies to convey non-verbal cues such as gaze, facial expressions, gestures, and body language. This notion emphasizes the importance of enhancing communication fidelity and social presence in telecommunication systems, enabling more natural and immersive remote collaboration experiences.

Some researchers argue that the perception of presence has multiple dimensions [209], while others propose three primary dimensions: spatial presence (feeling located in the virtual environment), social presence (feeling connected to others and able to communicate), and self-presence (perceiving oneself in the virtual environment through avatars or representations) [29, 205, 136]. These dimensions are interrelated and can interact with each

	Single Meeting Sites (same place)	Multiple Meeting Sites (different place)
Synchronous (same time)	face-to-face interaction e.g. decision room, shared table	**Multi-User 360° Video Collaboration** teleconferencing e.g. Skype, Zoom
Asynchronous (different time)	team rooms, large public displays, work groupware etc.	communication & co-ordinations e.g. email, bulletin boards, blogs

Table 2.1: The collaboration scenarios targeted by this research lie in the *same time, same place* and *same time, different place* quadrant of the Time/Space Matrix [101].

other to influence the overall sense of presence experienced by an individual. The adoption of 360° video technology presents a compelling opportunity to immerse users in rich and interactive experiences [154]. One key feature of 360° video is its ability to capture and present a panoramic view of the surroundings. This expansive field of view allows users to observe the entire environment, including crucial details for situational awareness, which means users are better equipped to perceive and interpret their virtual surroundings. They can gather comprehensive information, assess spatial relationships, and make informed decisions based on a deeper understanding of the context presented in the virtual environment. This capability is particularly valuable in scenarios where accurate and timely information is essential, such as in training simulations, emergency response, or complex teamwork [202]. This technology aligns with spatial and social presence dimensions, offering heightened sense of immersion and interaction in virtual environments.

Due to technological advancement, most commodity 360° devices the cost of devices that capture 360° video continues to fall, and the number of devices that support VR content is increasing. This provides the basis for much recent research into the use of this technology for immersive telepresence [192, 191, 228, 193].

2.2 Computer-Supported Cooperative Work

Computer-Supported Cooperative Work (CSCW) is concerned with how the use of computer systems in terms of software tools and technology, often referred to as groupware, can support group interaction, sharing goals, tasks, and knowledge among teams, and assist people in their collaboration, communication, and coordination [60]. 360° mixed reality collaboration falls into the “Same Time, Same Place” and “Same Time, Different Place” quadrant of the Time/Space Matrix, shown in Table 2.1. It involves real-time, immersive collaboration where participants interact with each other and digital content in a shared virtual or mixed-reality space, even though they are physically located in different places. This technology leverages mixed reality hardware and software to create collaborative environments that mimic physical presence, making it suitable for remote teamwork, virtual meetings, and design collaborations, among other applications. Synchronous collaboration means that individuals or teams are working together in real-time, while remote collaboration implies that they are not physically located in the same place. This type of collaboration is common in today’s globalized and digital world, where people can work together from different parts of the world using various online tools and technologies. The work of this thesis, which aims to explore these forms of interaction and design technologies to support them, is situated within the sphere of CSCW. This area of research, located at the intersection of computing and the social sciences, has traditionally endeavored to understand how technology can be designed to effectively support collaborative endeavors [12]. Within this field, this thesis focuses on the study of Video-Mediated Communication, with a specific emphasis on enhancing Video-Mediated Communication systems to facilitate distributed interactions, particularly in collaborative physical tasks.

2.3 AR Telecollaboration

AR Telecollaboration involves individuals working together or communicating remotely while utilizing AR to enhance their interactions. This approach is positioned on the left side of the Reality-Virtuality Continuum, as collaboration takes place within the physical environment, as defined by Milgram’s taxonomy [147] (see Figure 1.2).

In general, AR displays can be categorized into three primary types: 1) See-through hand-held displays, which are handheld devices that allow users to view the real world through the device’s screen while overlaying digital content onto it. Examples include smartphones and tablets equipped with AR capabilities where users can interact with AR content by holding up the device and observing the combined view of the real world and virtual objects on the screen [256]. 2) Head-mounted displays where the user sees both the real world and virtual content simultaneously, or video see-through, where cameras capture the real world and display it along with digital content device like Google Glass [24], and 3) Projection displays project digital content onto real-world surfaces, creating an augmented reality experience without the need for wearable devices. For example: SAR, or Spatial Augmented Reality, is a form of projection-based AR where digital content is precisely mapped onto physical objects or surfaces in the real world to enhance or alter their appearance [26].

AR telecollaboration encompasses activities such as sharing AR-enhanced visual information, collaborative work on augmented digital content, and team projects in augmented environments. This concept merges augmented reality technology with collaboration and communication across various domains.

There has been relatively limited research focus on asynchronous AR collaboration, and it poses several challenges, including understanding the role of time, capturing and re-visualizing annotations and different inputs, and examining how other forms of communication influence collaboration dynamics. Successfully addressing these challenges in asynchronous AR can enable multiple users to collaborate effectively and iteratively [96].

2.4 VR Telecollaboration

VR Telecollaboration involves individuals or groups interacting and working together using virtual reality technology. VR Telecollaboration leverages VR technology to create a sense of physical presence in the same virtual space for participants, even when they are geographically separated. VR is situated on the right side of the Reality-Virtuality Continuum, signifying a computer-generated pure virtual environment (see Figure 1.2). In immersive virtual environments, users experience high-level, multi-sensory immersion, contributing to a more realistic user experience, often achieved through the use of avatars [197].

In immersive virtual environments, users benefit from high-level, multi-sensory immersion, facilitated by a combination of advanced features such as avatars, blending virtual objects, dynamic lighting and rendering [192], visual cues, and spatial audio [229]. These elements collectively contribute to enhancing the visual fidelity and overall realism of the virtual experience in immersive virtual environments.

Avatars play a crucial role in inducing a sense of embodiment within collaborative spaces, varying from basic 2D representations to sophisticated 3D models [67]. The quality of avatar representation also impacts user immersion in VR environments, affecting how users perceive themselves and interact with others [145].

Furthermore, the importance of spatial audio cannot be understated. Spatial audio enriches the virtual experience by providing realistic auditory cues that enhance situational awareness and immersion [208]. Blending virtual objects seamlessly into the environment, along with dynamic lighting and rendering techniques, contributes to enhancing visual fidelity and overall realism within the virtual space.

2.5 Asymmetric AR-VR Telecollaboration

AR-VR Telecollaboration is a form of collaborative interaction that blends both AR and VR technologies in an asymmetric manner. Asymmetry refers to differences or variations in the capabilities, devices, or experiences among participants involved in the telecollaboration. One aspect of asymmetry in AR-VR Telecollaboration is the disparity in technology or equipment available to different participants. For example, one participant may have access to advanced VR headsets and controllers, while another may be limited to using a smartphone or tablet for AR interactions. This can impact the level of immersion, interaction capabilities, and overall user experience within the collaborative environment [116]. Another form of asymmetry difference in the roles and perspectives of participants. For instance, participants may take on different roles based on their expertise or objectives. This asymmetry in roles can lead to varying levels of engagement and contribution within the telecollaboration process [240].

Despite differences in hardware or capabilities, participants can feel fully immersed in asymmetric AR-VR Telecollaboration through the integration of interactive 3D visuals, spatial audio, and real-time interactions within the virtual environment. This capability is particularly beneficial in training scenarios, where experts can remotely guide learners using AR or VR devices, enabling interactive instruction and hands-on practice in diverse environments. For instance, in industrial settings, experienced technicians equipped with AR glasses can provide remote assistance to field workers, facilitating complex procedures and troubleshooting tasks effectively [61].

Similarly, in remote assistance applications, asymmetry can be leveraged to bridge expertise gaps. For example, a specialist equipped with advanced tools or access to a comprehensive virtual environment can collaborate with an onsite technician using simpler AR tools. This collaborative setup enables effective problem-solving, knowledge transfer, and real-time support across distances, showcasing the practical and inclusive nature of asymmetric AR-VR Telecollaboration in enhancing real-world interactions and professional experiences.

2.6 360-degree Panorama

360° cameras possess the unique capability to capture a comprehensive view of their surroundings in a single shot. 360° video technology captures immersive scenes from all directions, allowing viewers to explore virtual environments as if they were physically present within them. While the development of a true single-camera lens capable of achieving this

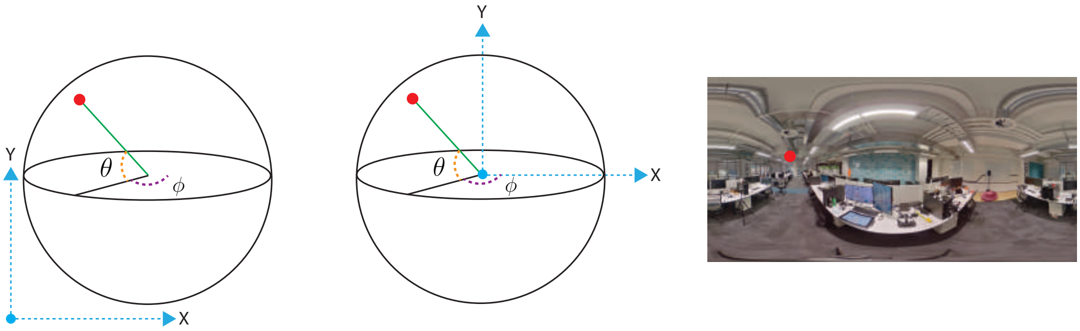


Figure 2.1: Illustration of global and local coordinates for a 360° camera with θ and ϕ to indicate the direction of a point on a sphere.

feat remains a challenge, the creation of a 360° panoramic image can be achieved by stitching together photos taken by multiple cameras. Over the past decade, researchers have introduced various stitching algorithms [132, 222, 103], enabling the creation of high-quality 360° panoramic images, even using everyday devices such as smartphones [261, 272, 15]. Furthermore, the democratization of consumer-level 360° cameras, exemplified by products like the Insta360 GO [1] and Ricoh Theta [2], has made 360° panoramic image capture accessible to the general public.

Figure 2.1, which depicts the coordinate transformation from global to image pixel coordinates. Given a point in space with global coordinates $\mathbf{x}_g$, the conversion to local coordinates $\mathbf{x}_l$ can be efficiently performed using the following equation:

$$\mathbf{x}_l = \mathbf{x}_g - o \tag{2.1}$$

Here, o denotes the central reference point of the 360° camera. This further transform the local coordinates $\mathbf{x}_l$ into spherical coordinates:

$$\begin{bmatrix} \theta \\ \phi \\ d \end{bmatrix} = \begin{bmatrix} \arctan\left(-\frac{(z_l)}{(x_l)}\right) \\ \arcsin\left(\frac{(y_l)}{d}\right) \\ ||\mathbf{x}_l|| \end{bmatrix} \tag{2.2}$$

Finally, these spherical coordinates can be seamlessly converted into image pixel coordinates by scaling them according to the dimensions of the image:

[1]https://www.insta360.com/
[2]https://theta360.com/

$$\begin{bmatrix} I_x \\ I_y \end{bmatrix} = \begin{bmatrix} I_{\mathbf{width}}(\frac{\theta}{2\pi+0.5}) \\ I_{\mathbf{height}}(\frac{\phi}{\pi+0.5}) \end{bmatrix} \tag{2.3}$$

Since 360° panoramic images encompass scene information from all directions, the 360° camera lacks rotational information, typically represented as R. Additionally, the concept of focal length f is non-applicable in the context of 360° cameras. Consequently, objects captured by a 360° camera exhibit scale changes solely in response to variations in their distance from the camera.

2.7 Review of Existing Literature

A systematic review of 557 papers detailing collaboration research utilizing augmented, virtual, and mixed reality (AR/VR/MR) across various publication venues spanning 2000 to 2023 is presented. The aim is to evaluate the approaches used in Extended Reality (XR) collaboration and explore future opportunities within this domain. A taxonomy has been developed to provide a structured understanding of the current research landscape. Based on this taxonomy, publications are categorized by input type (e.g., 360 video, point cloud), device type (e.g., HMDs, HHDs), interaction modality, collaboration mode (e.g., one-to-one, one-to-many), and evaluation type (e.g., Presence and Immersion). This classification helps in systematically analyzing the diverse approaches and technologies employed in XR collaboration research. Further analysis of 395 papers from the top ten publication venues provides insights into current trends in publication types (e.g., System Design, User Study) and application areas (e.g., Immersive Learning, Healthcare). Key findings from the review highlighted the importance of low-latency, real-time communication in XR collaboration, which was essential for inducing a sense of presence and immersion among users and enabling natural interactions like interacting with and manipulating virtual objects. Supporting multiple users, beyond traditional one-on-one interactions, with asymmetric compatibility ensured diverse teams could engage effectively within the XR collaboration space. In terms of collaboration styles and user experience principles, the focus was on spatial, remote, and asynchronous interaction, enhancing user experience through presence, intuitive interaction design, and considerations for comfort and ergonomics. The main gaps and opportunities in this research area have been identified by examining current trends in XR collaboration. These trends underscore the need for adopting multi-user platforms, integrating AI, and developing industry-specific applications. However, challenges such as network optimization, privacy concerns, and the refinement of interaction models persist. Addressing these issues is crucial for advancing the capabilities and usability of XR collaboration tools.

The landscape of remote collaboration has undergone significant transformation, particularly driven by the COVID-19 pandemic [10]. Technological advancements and the huge widespread adoption of remote work have made remote collaboration a necessity rather than an option for many organizations [264]. This necessity is further augmented by Extended Reality (XR) technologies encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), which provide a comprehensive framework that enables

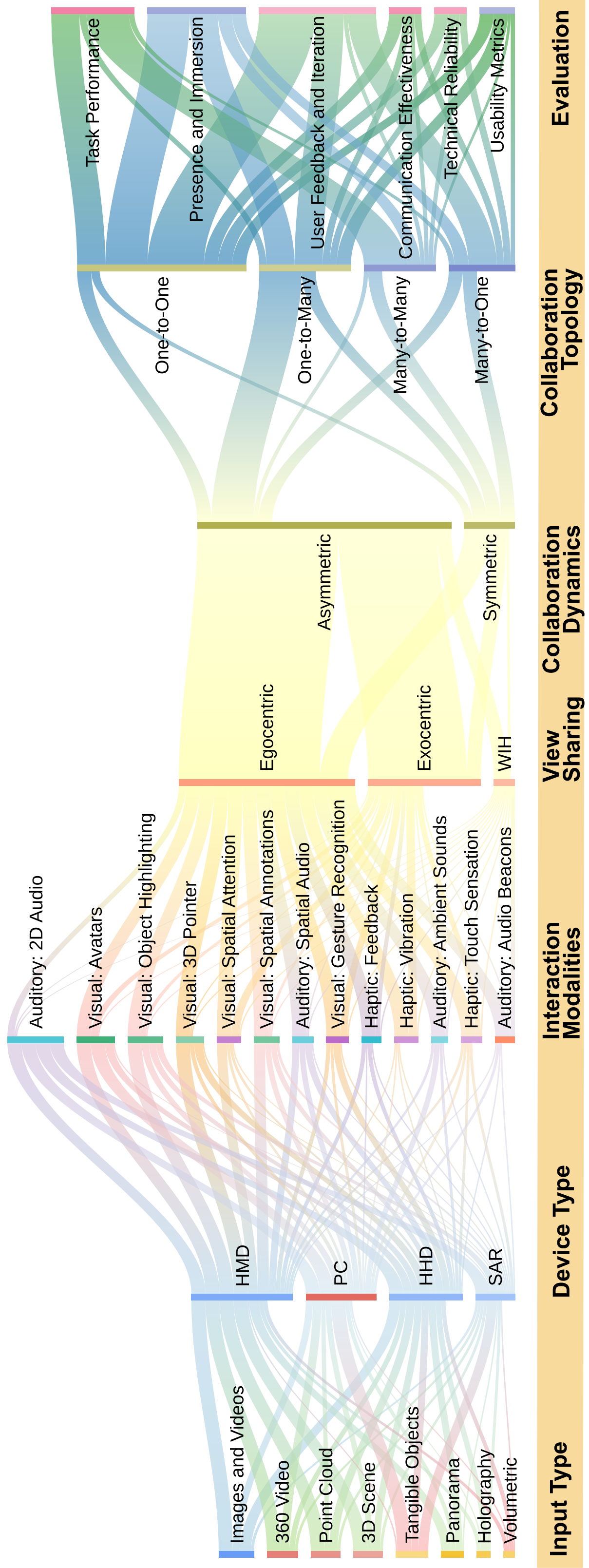


Figure 2.2: A total of 557 papers that employed XR collaboration research between 2000 and 2023 were classified by (left-to-right): system input type, device type, interaction modalities, viewpoint sharing, collaboration dynamic and topology, collaboration topology, and evaluation method in user studies. An interactive version of this classification is available at https://zaps.one/collabxr.

individuals or groups to collaborate, communicate, and interact in shared virtual spaces, transcending physical distances beyond traditional video conferencing [185].

The lack of comprehensive guidelines and methodologies for designing and evaluating XR collaboration solutions presents a notable challenge, especially in establishing structured pipelines. This challenge is particularly evident when it comes to evaluating the usability and performance of such solutions, as identifying suitable evaluation strategies, variables to examine, and appropriate data collection methods becomes inherently complex, making it difficult to address specific research questions effectively. As a result, designers and developers encounter significant challenges in creating XR collaboration experiences that meet both user needs and performance standards.

Prior surveys have investigated such guidelines and methodologies, mostly focusing on particular areas such as remote assistance and training [66], education [164], security and privacy [53], and cultural heritage [17], typically drawing from very focused data sources. Borhani et al. [32] examined the use of annotation in XR environments from 103 publications not specific to collaboration. Fidalgo et al. [66] conducted a systematic literature review to survey remote assistance and training in MR environments analyzing 62 publications not fully capture the complexity and diversity of XR applications and use cases. Wang et al. [253] explored remote collaboration in AR and MR with a specific focus on physical tasks. Merino et al. [144] conducted systematic review of 458 publications to provide guidance for future evaluations of MR/AR approaches. Lukosch et al. [138] explores research articles that highlight and explore unique collaborative experiences using AR systems, also outlining a future research agenda. None of the aforementioned approaches specifically focus on guidelines and methodologies for designing, developing, and evaluating XR collaboration solutions that includes AR, VR, and MR technologies.

This survey addresses this by meticulously analyzing a large volume of publications sourced from diverse data repositories. Through this analysis, the aim is to elucidate a structured pipeline for the design, development, and evaluation of XR collaboration solutions. By synthesizing insights from a broad spectrum of research, the survey endeavors to provide actionable guidance for practitioners and researchers alike, offering a roadmap to navigate the complexities inherent in XR collaboration. Additionally, the survey sheds light on the implications for future research and practice in this domain, facilitating the forecasting of future trends and charting the research trajectory for XR collaboration. Through these efforts, there is an aspiration to contribute to the advancement of XR technologies and their application in fostering collaborative environments across various sectors.

The contributions of this survey are as follows:

- A **comprehensive taxonomy** of methods and approaches for XR collaboration.
- A **comparative analysis** of XR collaboration platforms in terms of features, compatibility, scalability, user interface, and communication tools.
- The **identification and discussion** of challenges and future trends, including technical constraints, user comfort, data privacy, and integration with existing workflows.
- An **interactive visualization** tool for exploring the research papers and a publicly available data set of the analysis.

2.7.1 Scope of Survey

Collaboration generally is the process of individuals or groups working together towards a common goal to achieve desired outcomes [41]. The field of Computer-Supported Cooperative Work (CSCW) has focused on prototyping solutions for knowledge sharing between distributed collaborators known as *groupware* [60, 80, 13] and the space and time that the collaboration takes place [9]. One major issue of remote collaboration is the fact that collaborators do not share a common space/world, reason for the interest in using XR technologies (i.e., real-and-virtual combined environments). Interpretations of XR differ among researchers. Mann et al. [141] perceived XR is in line with the MR continuum [147] (see Figure 1.2), while others, like Al-Adhami et al. [4], consider XR to encompass both MR and VR. Paradiso and Landay [163] introduced the concept of cross reality, which utilizes sensor/actuator networks to influence both RE and VE, blurring the boundaries between physical and digital environments.

Despite variations in interpretation, all these concepts share a common objective: expanding or augmenting reality to provide experiences or perceptions that are not present in the physical world, thus facilitating reality expansion through immersive technologies.

Over the years, numerous compelling research studies have explored augmented reality and virtual reality. Various independent reviews in the literature that covered significant aspects of AR, VR, and MR and inspired this survey are listed in Table 2.2.

Wallace et al. [243] reviewed how Computer-Supported Cooperative Work (CSCW) research has changed over time, especially with the emergence of new technologies like smartphones and wearables. They analyzed ACM Conference on CSCW proceedings from 1990 to 2015 to understand these changes. The findings show that CSCW research now emphasizes describing collaborative work environments in practice rather than developing new systems in labs. There's also a lack of bibliographic work and focus mainly on single-device studies. The study suggests challenges for CSCW research in adapting to technological changes effectively.

Boletsis et al. [31] reviews VR locomotion techniques from 2014 to 2017, identifying 36 relevant articles covering 11 methods such as real-walking and joystick-based locomotion. It highlights a predominant focus on VR technology over user experience since the VR revival, with most techniques relying on physical interaction for navigation. The study introduces a typology categorizing VR locomotion into four types: motion-based, room scale-based, controller-based, and teleportation-based.

A 10-year survey by Dey et al. [54] reviewed 369 AR user studies from 2004–2014. They found an increasing trend of involving handhelds in AR user studies. They also provided more application areas, adding Education and Medical as areas with a high number of publications with user studies. They have confirmed that most user studies conducted in a laboratory setting as 75% were lab-based evaluations.

A comprehensive review of collaborative systems based on AR and MR from 2013–2018 has been proposed by de Belen et al [52]. Their study focuses on how users in remote asynchronous MR environments can create digital information and re-visit project meetings. The meta-review of collaborative MR systems shows that 103 of 259 systems use a local collaboration of multiple users in a shared physical environment. They also pointed out that the provision of non-verbal cues such as gaze and pointing cues were identified as two important factors in enhancing collaboration quality in MR spaces.

A more recent systematic exploration of the mixed-reality space by Ens et al. [61] confirms that "collaborative MR technology is now mature enough to focus squarely on human needs." This includes looking at how hand-held AR can be used to look at complex collaboration in space as well as time and symmetry. They pointed to mixed presence as an issue that still needs to be addressed to match the complexity of real-world collaboration. Furthermore, the roles of the users in groups were highlighted as another issue that is yet to be addressed in groups.

Merino et al. [144] reviewed 458 papers on evaluations in mixed and augmented reality (MR/AR) from 2009 to 2019. It categorized them by type, research topic, evaluation scenario, cognitive aspects, and context. Two main groups were identified: technology-centric and human-centric studies. Of the 458 papers, 248 involved user studies with 5,761 participants, predominantly in laboratory settings. This survey primarily centers on the influence of technological innovation on collaborative XR research, while incorporating traditional CSCW concepts. Relevant papers have been gathered from major conference proceedings, including CHI, ISMAR, IEEE VR, VRST, and CSCW.

Expanding the survey to include a structured pipeline for XR collaboration design, development, and evaluation could provide actionable guidance for practitioners and researchers.

		Test Conditions		
Paper	*Year*	*Time Period*	*Summary*	*# of Papers*
Wallace et al. [243]	2017	1990-2015	analyzed publications across four classification schemes	1209
Boletsis et al. [31]	2017	2014–2017	identified four distinct VR locomotion types	36
Dey et al. [54]	2018	2005–2014	identified two areas with high publications	291
de Belen et al. [52]	2019	2013-2018	overviews on collaboration in MR	259
Ens et al. [61]	2019	1995-2018	focuses on the design possibilities of collaborative AR	111
Merino et al. [144]	2020	2009-2019	provided guidance for future evaluations of MR/AR approaches	458
This work	2023	2000-2023	analyzed publications focuses on collaborative XR	557

Table 2.2: Summary of existing review papers involving collaboration in AR/VR/MR.

2.7.2 Methodology

A systematic literature review approach was conducted based on the PRISMA methodology [3]. A detailed data collection method was established to gather information from the selected studies, accompanied by a coding scheme to efficiently categorize and analyze the data. The following sections describe the methodology in detail.

Research Question

To assess the current state of XR collaboration research, this survey aims to address the following research questions:

SRQ1: How do users effectively connect and collaborate with other users in the XR collaboration space?

SRQ2: What type of collaboration style, user experience, and evaluation techniques have been used in XR collaboration?

SRQ3: What are the current trends of XR collaboration and the main research challenges?

Data Collection Method

This process involved multiple stages, including identification, screening, eligibility, and inclusion. A comprehensive search was conducted across several electronic databases, including PubMed, Scopus, IEEE Xplore, and Google Scholar. Keywords and search terms were carefully selected based on the research questions and objectives. The search involved combining various search queries related to AR/VR/MR (“augmented reality” OR “mixed

[3]https://www.prisma-statement.org/prisma-2020-statement

reality" OR "virtual reality" OR "AR" OR "MR" OR "VR") with keywords associated with specific collaboration topics to create regular expressions across different data sources. A selection of the keywords used is outlined in Figure 2.3.

The initial search results yielded **1,425 articles**, which were imported into reference management software where duplicates were removed. The titles and abstracts of the remaining articles were screened against the inclusion and exclusion criteria.

In total, **1,403 abstracts** were retrieved and assessed for eligibility. Of these, **22 abstracts** were not accessible due to paywalls or other access restrictions imposed by publishers. Articles that did not meet the inclusion criteria were excluded. The final set of studies that met all the criteria were included in the review. Data extraction was then performed using a standardized form to ensure consistency and comprehensiveness.

Inclusion and Exclusion Criteria

During the screening process, articles were filtered using specific inclusion and exclusion criteria to ensure the relevance and quality of the selected studies.

Inclusion criteria included studies published in peer-reviewed conferences or journals, articles written in English, and studies that focus on the specified research questions. Additionally, research involving empirical data or experimental results, and publications from the years 2000 to 2023, were considered.

Exclusion criteria encompassed grey literature such as conference abstracts, theses, and dissertations. Studies not available in full text, articles not related to the primary research focus (i.e., publications containing search terms not intended to understand, develop, design, or implement XR collaboration, but used exclusively for another purpose), and duplicate publications appearing in multiple libraries or formats were excluded. For instance, if identical results were presented in both a poster or extended abstract and a more comprehensive journal article or collection, the latter was prioritized to avoid redundancy. Furthermore, publications that were not primary studies, such as demos, posters, review articles, or commentaries, were excluded. After applying these criteria, the final selection comprised **557** papers.

2.7.3 Taxonomy

A taxonomy was developed to systematically classify and organize the key concepts and findings from the included studies (see Figure 2.2). The taxonomy was designed to cover various aspects relevant to the research topic, including:

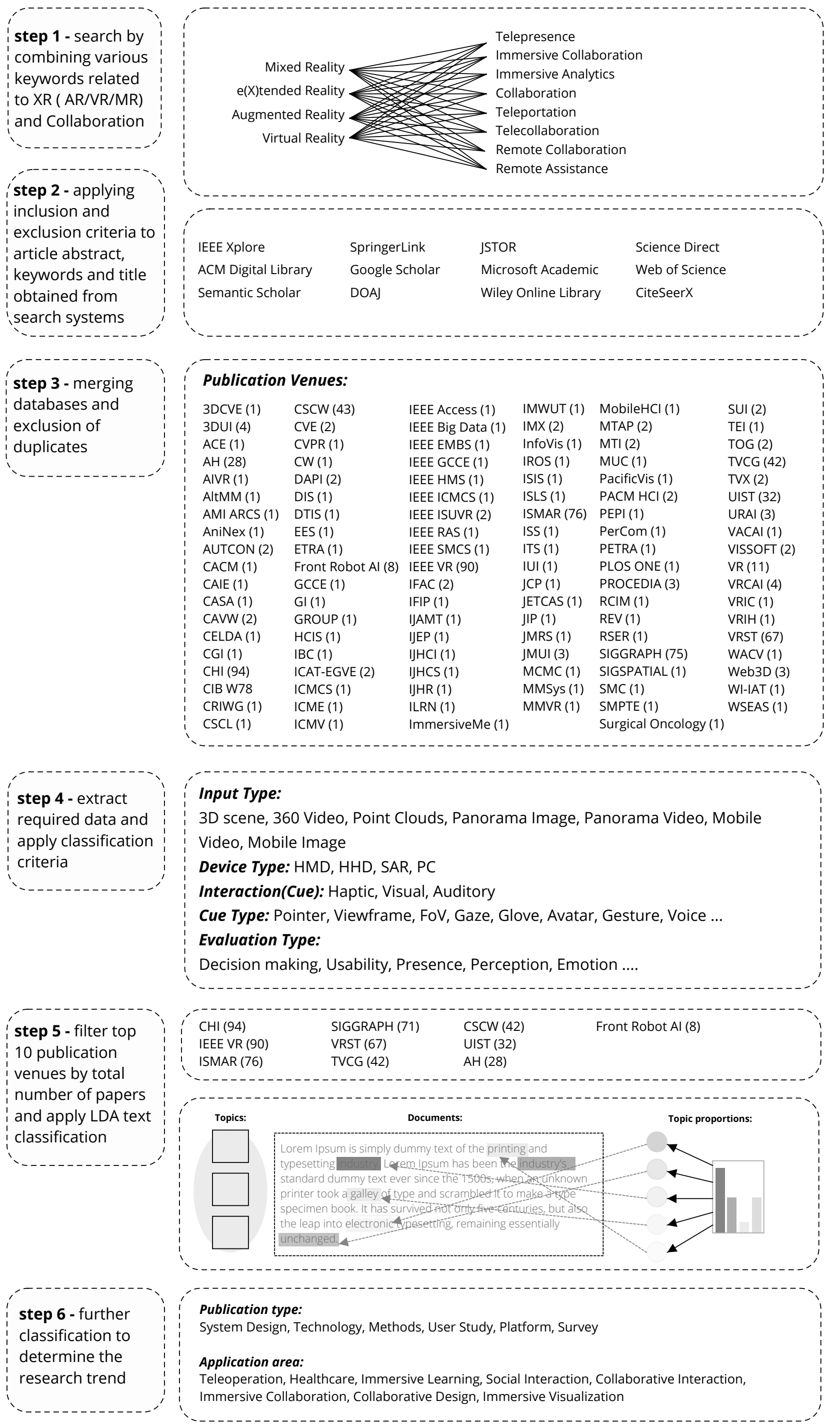


Figure 2.3: The workflow for the systematic literature review.

Input Type

Refers to the method or medium through which users interact with and input data into the system. This includes the ways in which data is captured, represented, and input into mixed reality systems to create immersive experiences.

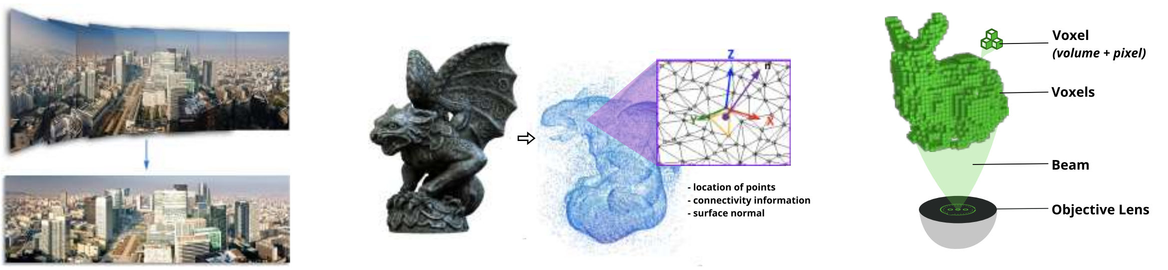
Images and Videos Most of the earlier collaboration systems were developed using 2D images and video. These systems typically utilized cameras connected to common mobile, desktop, or embedded devices [3, 102, 252, 39]. One example is Zoom, where a collaborator employs a camera's video feed as input and shares it with a remote user in real time [73, 39].

Panorama Panoramas, created by stitching images together to provide a wide-angle view, do not necessarily offer a full 360-degree field of view (see Figure 2.4a). These panoramic views enhance interactivity and control over the viewing experience, allowing for seamless exploration of immersive environments [165, 149].

360-degree Video 360-degree videos provide full panoramic views, enhancing immersion and allowing users to explore and interact with environments using 3 degrees of freedom (3DoF). Due to the limited field of view of most webcams, which restricts the capture of remote environments on telepresence platforms, 360-degree video-based telepresence platforms are becoming increasingly popular [228, 193].

Point Cloud Point clouds are collections of data points sampled from real-world surfaces, typically generated by 3D scanners or LiDAR, capturing detailed information about physical objects or spaces (see Figure 2.4b). They enable real-time mapping of environments, allowing remote collaborators to visualize and interact with a shared physical space [123, 71, 242].

3D Scene Virtual worlds serve as environments where users can teleport to simulate interactions with participants through communication and collaboration. These environments are often pre-created using digital graphics software or 3D reconstruction software [186, 187, 157, 249]. Consequently, numerous studies have been conducted within this setup, which is well-suited for immersive learning and collaborative design integration to enhance the sense of presence [211, 131].



(a) (b) (c)

Figure 2.4: (a) Stitching a Panorama using multiple images, (b) Point cloud sampled from the real surface, (c) Holograms, also called volumetric displays, are composed of voxels (volume + pixel).

Holography This involves interacting with holographic representations or displays, where users manipulate virtual objects that appear as holograms within a mixed reality environment. For example, Holobox is a multi-user holographic display system that enables participants to interact with holographic projections in real time [110]. It is designed for collaborative settings where multiple users can view and manipulate the same 3D data simultaneously.

Volumetric Representations that involve 3D models or volumetric capture, often used to create detailed, interactive digital environments or objects that users can manipulate (see Figure 2.4c). Users can manipulate and interact with volumetric representations using gestures, controllers, or other input devices within immersive environments.

Tangible Objects This involves using physical objects or props as input devices within mixed reality settings. Tangible objects can be tracked and integrated into virtual environments, allowing users to interact with digital content or control virtual elements through physical manipulation of tangible objects [107].

Device Type

Refers to the hardware setup or display environment used to deliver mixed reality experiences (e.g., HMDs, projection-based systems).

Desktop or PC Desktops or PCs allow users to both actively and passively experience MR content [64]. However, these setups typically provide limited immersion.

Immersive Rooms/Caves Immersive rooms or CAVEs are physical spaces equipped with multiple displays or projection surfaces that surround users, creating a fully immersive environment. Users can interact with virtual content or simulations within these spaces using specialized input devices, such as motion trackers or handheld controllers [75].

Projection-Based Projection-Based Systems project XR content onto physical surfaces, such as walls or screens, creating immersive environments for collaborative interactions and presentations. Users can gather around the large screen display to view and engage with XR content simultaneously, making it suitable for group collaboration and shared experiences [257]. Spatial Augmented Reality (SAR) is one example of Projection-Based Systems, where virtual objects are seamlessly integrated into the physical environment through projection, enhancing users' perception of reality and enabling interactive experiences [166].

Hand-held Devices (HHD) Handheld devices, such as smartphones and tablets, equipped with cameras, sensors, and powerful processors, can deliver MR experiences. These devices utilize cameras to capture the real world and then overlay virtual elements on the device's screen. Users can interact with these virtual objects through touch gestures, offering a more accessible entry point into MR. This technology has been employed as a collaboration tool in various systems [102, 188, 152, 76].

Head-mounted Display (HMD) HMDs immerse users in digital environments or overlay digital information on the real world. Three types of different head-mounted display technologies exist to superimpose graphics onto the user's view of the real world:

(1) Video See-Through (VST) Displays: VST displays use video cameras to capture the real-world environment and overlay digital content onto the user's field of view, eg: Microsoft HoloLens, Magic Leap.
(2) Optical See-Through (OST) Displays: OST displays use optical components, such as transparent displays or mirrors, to superimpose digital information onto the user's natural field of view e.g., Google Glass.
(3) Fully Immersive or Opaque Displays: block the user's view of the real world and present a completely virtual environment eg. Oculus Rift, HTC Vive.

Interaction Modalities

Interaction cues play a vital role in shaping how users engage with the XR environment and collaborate with others. These cues provide guidance and feedback to users, enabling ef-

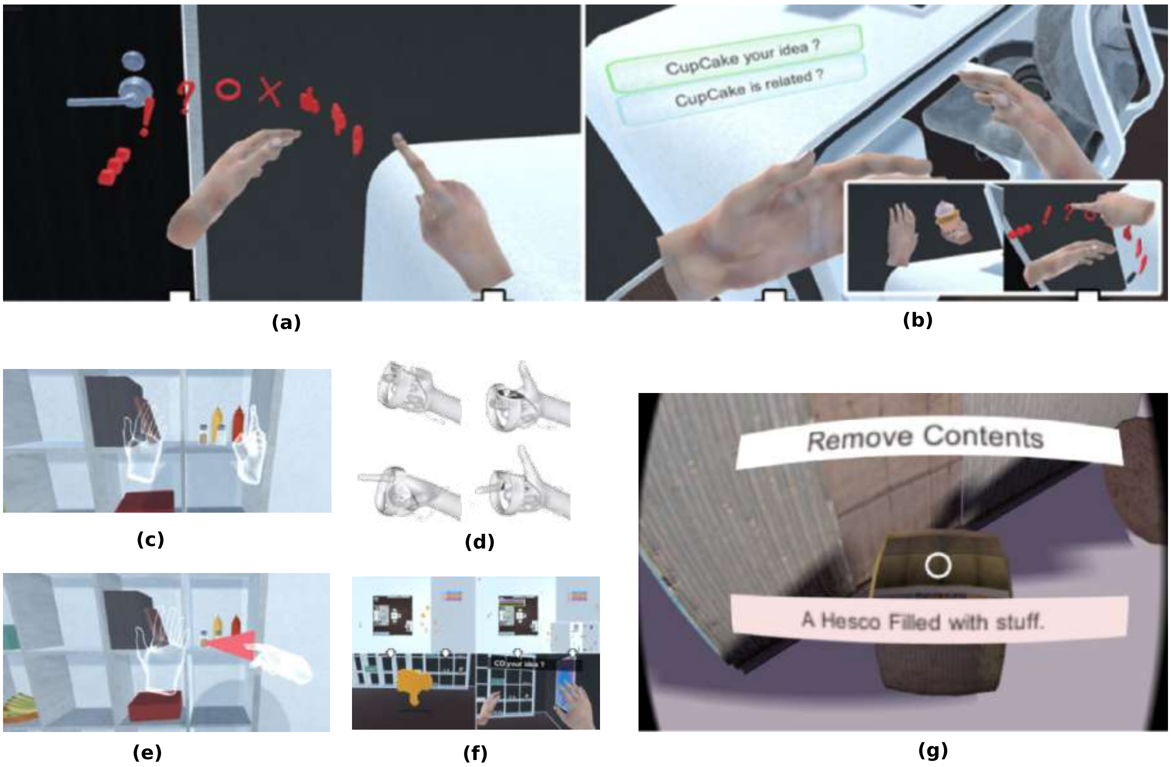


Figure 2.5: Interaction techniques in XR collaboration: Drawing annotation (a), viewpoint conversion (b), pointing (c), hand gestures (d), ray to objects (e), overlay icons and label text on top shared objects (f), gaze (g). Diagram adapted from [99].

fective communication and collaboration within the mixed reality space. This encompasses visual cues, auditory cues, haptic feedback, and other forms of interaction (see Figure 2.5).

Visual Cues Visual cues, such as highlighting certain objects or elements within the XR environment to draw attention or indicate significance, also assist users in manipulating virtual objects by providing visual feedback on their position, orientation, and manipulation axis.

Avatars are virtual representations of users, which can range from simple icons to full-body representations or highly detailed, customizable digital versions of the users. They enable users to interact with each other and with the environment, enhancing the sense of presence and social interaction.

Object Highlighting serves as a visual cue to draw attention to specific elements within a scene. This technique enhances the user's ability to focus on relevant objects, facilitating better understanding and interaction within virtual environments.

Spatial Annotations also known as "see-what-I-see" capability, allow users to draw or highlight objects in real space, providing a way to visually communicate and share insights within a shared environment. This feature enhances collaboration by making virtual notes or marks that are contextually tied to the physical space.

3D Pointer used in virtual environments to indicate positions or objects within the space, assisting users in pointing out details or directing attention.

Gesture Recognition involves the use of sensors and software to interpret and respond to the user's hand or body movements, allowing for intuitive interaction within mixed reality environments.

Spatial Attention ability of a system to detect and respond to where a user is focusing their attention within a 3D space. This can involve tracking eye movements, head orientation, or pointing gestures to determine what the user is looking at or interacting with.

Auditory Cues Auditory cues including ambient sounds, feedback sounds, or notification sounds that provide auditory feedback.

2D Audio refers to traditional stereo or mono audio without spatial positioning.

Spatial Audio is positioned within the virtual environment to create a sense of direction, distance, and location relative to the user's position.

Ambient Sounds enhance the immersive experience by simulating real-world auditory cues, such as nature sounds or city ambiance.

Audio Beacons utilize distinctive audio signals or markers to guide users toward specific locations or objects through auditory cues.

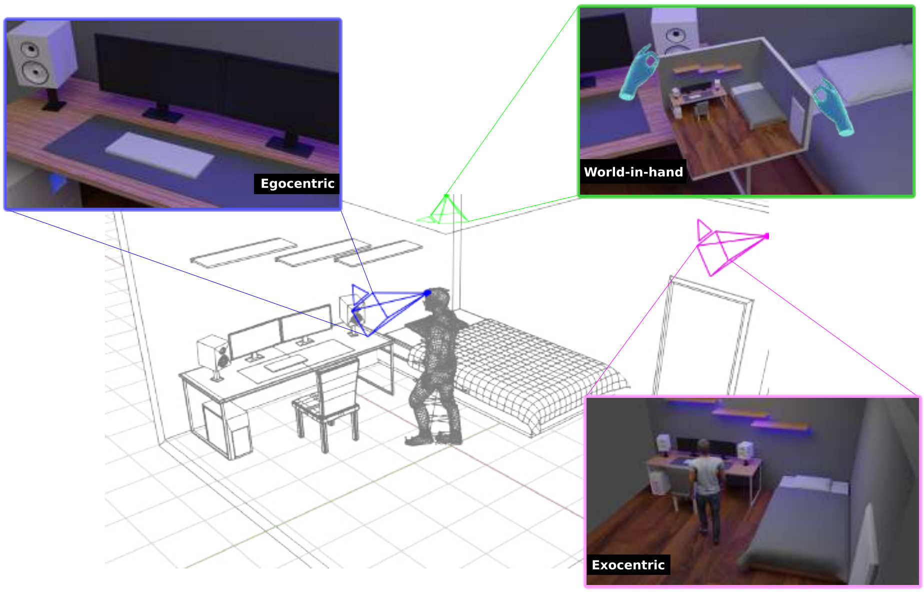


Figure 2.6: Visualization of Ego, Exo, and World-in-Hand viewpoints.

Haptic Cues Haptic cues provide tactile feedback to users, enhancing the sense of immersion and realism in virtual environments. These cues simulate the feel of objects, textures, or forces, allowing users to interact more naturally with digital content.

Vibration simulates physical sensations or notifications through tactile feedback provided by input devices, such as controllers or wearables.

Touch Sensation simulates textures, surfaces, or interactions with virtual objects to enhance the sense of touch within the virtual environment.

Feedback (force, Kinesthetic) simulates forces, resistance, or physical interactions (e.g., pushing, pulling) experienced by users interacting with virtual objects, enhancing realism and immersion.

View Sharing

Viewpoint sharing in an MRC environment allows participants to view each other's perspectives, align actions, provide immediate feedback, and interact with digital content in ways that enhance the collaborative experience (Figure 2.6).

Egocentric This perspective represents the user's point of view, often centered around the user's head and hands. It provides a first-person view, giving users a sense of being within the environment and interacting directly with the objects around them.

Exocentric This perspective offers an external viewpoint, often from a third-person perspective. It provides an overview of the scene, showing the user's interactions with the environment from outside the user's immediate viewpoint, which is useful for navigation and understanding spatial relationships.

World-in-Hand This perspective enables users to manipulate and interact with a miniature view of the virtual world. It resembles a "bird's-eye view" or "god view," providing users with an overview of the environment. This allows them to control or alter aspects of the collaboration space, facilitating tasks such as positioning, organizing, and editing within the virtual environment.

Collaboration Scenarios

XR collaboration setups can vary depending on the combination of devices, tools, and the type of collaboration style supported by the platform. A "collaboration place" refers to a local space, which could be an office, a computer-generated scene, a construction site, or even a space station. The location from which a user joins the collaboration place is termed a remote space. **Participants role** in the collaboration place is defined as follows:

Local: This refers to a participant who is physically present in the collaboration space where the interaction primarily occurs. This could be someone using VR/AR equipment in a specific physical location, such as an office, construction site, or designated virtual meeting space.

Remote: This refers to a participant who joins the collaboration from a different location, typically using XR devices or tools to connect virtually to the shared environment. Remote users interact with the local user space and other participants through digital means, enabling virtual presence and participation despite physical distance.

The interaction of participants within MRC environments, based on their roles, responsibilities, and capabilities, significantly dictates the **Collaboration dynamics** of XR applications. Whether participants have equal or varying levels of involvement influences these dynamics, which can be characterized by:

Asymmetry In asymmetric collaboration, the participants have different roles, responsibilities, or capabilities. One party might have more power, knowledge, or resources than the other. For example, in a mentorship relationship, the mentor typically has more experience and knowledge than the mentee.

Symmetry Symmetric collaboration occurs when all participants have equal roles, responsibilities, and capabilities. Each party contributes to the collaboration in a similar manner. For instance, in a team project where every member has an equal say in decision-making and contributes equally to the work, it's a symmetric collaboration.

Once roles and dynamics are assigned, the way individuals engage in collaborative activities defines the **Collaboration mode**. This term categorizes and defines different approaches to working together, whether in real-time or over extended periods. Examples of collaboration modes include:

Asynchronous Asynchronous collaboration doesn't require all participants to be present or active at the same time. Instead, individuals can work independently on their own schedules and then share their contributions with others. Email, discussion boards, and shared documents are common tools used for asynchronous collaboration.

Synchronous Synchronous collaboration involves real-time interaction between participants. All parties are active and engaged simultaneously, allowing for immediate feedback and communication. Video conferencing, instant messaging, and live document editing are examples of synchronous collaboration tools.

The communication and interaction between collaborators, influenced by their roles, the chosen collaboration mode, and the overall dynamics of the interaction, require the establishment of a specific **Communication topology**.

One-to-One each collaborator interacts directly with another single collaborator, for example, describes a typical scenario for a meeting between two people, involving bi-directional communication.

One-to-Many one collaborator communicates with multiple other collaborators simultaneously, for instance, a single remote user interacts with multiple hosts at a single location.

Many-to-One multiple collaborators communicate collectively or individually with one central collaborator. For example, a team of engineers collaborates on designing a complex system in XR. They each contribute their ideas and modifications to a shared virtual prototype, which is overseen and reviewed by a project manager.

Many-to-Many multiple collaborators (remote users and local users) interact freely with one another without a central communication hub. For example, architects, interior designers, and clients participate in a virtual design review meeting in XR. They explore and discuss different aspects of a building design together, providing real-time feedback and making collaborative decisions.

Evaluation

After the development of each MR application, evaluation of the implemented system to understand its potential shortcomings is important. This evaluation is conducted from several perspectives, e.g., based on a user survey, a comparison with a common system, or based on an environmental reference. Evaluation based on user surveys shows the user satisfaction level with the system output. A comparison of results with a common system helps to determine the potential advantages or disadvantages of the MR application in practice, and evaluating the system output by observing environmental references is a practical MR-based navigation application.

Presence and Immersion Evaluate the sense of presence and immersion experienced by participants in the XR environment. Evaluation includes measures such as the Igroup Presence Questionnaire (IPQ) and the Networked Minds Measure (NMM).

Usability Metrics Assess the overall usability, user satisfaction, and engagement with the XR system. This is measured through standardized questionnaires such as the System Usability Scale (SUS), the User Experience Questionnaire-Short (UEQ-S), and the Usability Metric for User Experience (UMUX-Lite).

Task Performance Measure the efficiency and effectiveness of task completion within the mixed reality environment. This includes both subjective and objective measures such as completion time, the NASA Task Load Index (NASA-TLX) or RTLX, and error rate analysis.

Communication Effectiveness Evaluate the quality and efficiency of communication between participants in the mixed reality setting. This can be assessed by analyzing the clarity and fidelity of audio and visual communication, conducting conversation analysis (e.g., turn-taking, clarity of speech), and performing observational studies.

Technical Reliability Assess the reliability, scalability, and flexibility of the mixed reality platform and associated technologies. This involves monitoring frame rate and latency, as well as generating heatmaps from user tracking data to understand usage patterns and system performance.

User Feedback and Iteration Gather qualitative feedback from participants to inform iterative improvements of the system. Methods include conducting qualitative interviews,

most common words	Abs Freq.	Rel Freq.
social interaction	5.0	0.04065
visualisation	4.0	0.03252
immersive	3.0	0.02439
design	3.0	0.02439
f0	2.0	0.01626
analytics	2.0	0.01626
allows	2.0	0.01626
well	2.0	0.01626
several	2.0	0.01626
user	2.0	0.01626

most common words	Abs Freq.	Rel Freq.
interaction	3.0	0.022727
whiteboard	3.0	0.022727
side	3.0	0.022727
training	2.0	0.015152
remote	2.0	0.015152
digital	2.0	0.015152
robot	2.0	0.015152
heathcare	2.0	0.015152
direction	2.0	0.015152
training	1.0	0.007576

most common words	Abs Freq.	Rel Freq.
telepresence	4.0	0.031496
surrounding	3.0	0.023622
collaborative	3.0	0.023622
medical	2.0	0.015748
sharing	2.0	0.015748
training	2.0	0.015748
collaboration	2.0	0.015748
persons	2.0	0.015748
training	2.0	0.015748
places	2.0	0.015748

tabletop	6.0	0.043796
occupant	4.0	0.029197
designer	4.0	0.029197
teleportation	3.0	0.021898
allows	3.0	0.021898
xr collaboration	2.0	0.014599
cooperation	2.0	0.014599
system	2.0	0.014599
communication	2.0	0.014599
data	2.0	0.014599

teleoperation	3.0	0.028846
environments	3.0	0.028846
f0	2.0	0.019231
telepresence	2.0	0.019231
phones	2.0	0.019231
local user	1.0	0.009615
co-presence	1.0	0.009615
haptic	1.0	0.009615
learning	1.0	0.009615
film	1.0	0.009615

user	4.0	0.030534
remote	3.0	0.022901
system	3.0	0.022901
sharing	3.0	0.022901
natural	3.0	0.022901
cues	3.0	0.022901
gaze	3.0	0.022901
gesture	3.0	0.022901
social	2.0	0.015267
wearable	2.0	0.015267

Figure 2.7: Clustering results from the text mining.

organizing focus groups, using think-aloud protocols during tasks, and administering post-task surveys to gather user perspectives and suggestions for enhancements.

Empirical Derivation of Topics

To identify the prominent topics discussed in a substantial number of publications, the Latent Dirichlet Allocation (LDA) [30] method is employed. LDA automatically interprets and clusters words and documents into various topics. This text-mining technique is commonly used in recommendation systems, such as web search engines and advertising applications. LDA is a generative probabilistic model for a corpus, where it treats documents d as random mixtures over latent topics (t), $p(t|d)$, where every topic is characterized by a distribution over words (w), $p(w|t)$. The method uses the following formula:

$$p(w|d) = p(w|t) \star p(t|d) \tag{2.4}$$

Here, $p(w|d)$ represents the probability that a certain word will appear in a particular document on a certain topic. Word-topic distribution $p(w|t)$ and topic-document distribution $p(t|d)$ are initially randomly chosen, and the LDA iteratively updates and estimates

Publication type	Summary	Application Areas	Summary
User Study	*user evaluation, user study design*	Healthcare	*medical training simulations, patient therapy, surgical planning and rehearsal*
System Design	*proof-of-concept, prototype*	Immersive Learning	*simulation-based training, virtual classrooms, educational games*
Platform	*system architecture, integration, scalability*	Collaborative Design	*team-based design processes, interactive prototyping, design reviews*
Collaborative Interfaces	*interface design, interaction techniques, usability testing*	Collaborative Interaction	*multi-user simulations, interactive shared environments, co-located interaction*
Rendering Techniques	*visual fidelity, frame rate optimization, real-time rendering*	Social Interaction	*social platforms, online communities, digital networking*
Applications	*case studies, application development, use case scenarios*	Immersive Collaboration	*virtual meetings, remote management, collaborative workspaces*
		Immersive Visualization	*architectural design, scientific visualization, virtual tours*
		Teleoperation	*remote control of machinery, robotic surgery, drone operation*

Table 2.3: Empirical derivation of the topics for the publication type and application areas to improve field encapsulation.

these probabilities until the system converges. This process establishes the relationships between documents, topics, and words (see Figure 2.7). All 395 abstracts were input into the LDA model, and a range of parameters were tested. Through empirical analysis, it was determined that six topics best encapsulate publication types, and eight topics suit applications in the field (see Table 2.3).

To increase the likelihood of capturing prominent and influential research trends and applications within the field of XR collaboration, the papers were organized by their publication venue, and selected those from the top 10 venues. These top venues often feature papers that are considered influential, providing valuable insights into emerging trends, prevalent applications, and innovative approaches in the field. (see Figure 2.8). Through empirical analysis, the following topics were identified.

Publication Type In the empirical analysis of topics within XR collaboration, publications are categorized based on their types (see Table 2.3). Below are the primary publication types identified through this analysis:

System This category includes publications that detail the creation, implementation, and features of new XR systems, frameworks, or platforms.

User Study These studies focus on empirical research involving human participants interacting with XR applications. They evaluate user experiences, preferences, performance, and behaviors within XR environments.

Platform This category involves the analysis and description of XR platforms or ecosystems. Papers in this category explore the features, compatibility, and capabilities of both hardware and software XR platforms.

Collaborative Interfaces Research in this area examines the design and effectiveness of interfaces that facilitate collaboration within XR environments.

Rendering Techniques This category covers studies on the rendering techniques used in XR, addressing how visual content is generated and displayed to enhance immersion and realism.

Applications This includes publications that discuss specific applications of XR technologies, demonstrating their use in various fields such as healthcare, education, and training.

Application Area The specific domains or fields where XR collaboration technologies are being applied or studied were derived through analysis of application areas. These domains represent diverse contexts where XR collaboration solutions are being explored, developed, and deployed to address various challenges and opportunities (see Table 2.3). Following application areas were identified:

Teleoperation Remote control of machinery or robots, often using XR technologies to provide a live, immersive view of the environment being manipulated.

Immersive Visualization Using XR technologies to create visualizations that provide a deeper understanding of complex data, often in scientific, engineering, or architectural contexts.

Immersive Collaboration Environments where participants interact in a shared virtual space, using XR to facilitate communication and collaboration.

Social Interaction Interaction between users in virtual environments, often facilitated by avatars, social platforms, or interactive settings.

Collaborative Interaction Refers to interactive activities in XR settings where multiple users collaborate on tasks or projects.

Collaborative Design Involves using XR technologies to collaboratively design products, structures, or systems, often utilizing shared digital spaces for brainstorming and development.

Immersive Learning Educational methods that leverage XR to provide immersive, interactive learning experiences.

Healthcare Applications of XR in healthcare, including surgical simulations, medical training, patient therapy, and telemedicine.

In addition to analyzing paper topics, the relative impact of each paper was assessed by calculating its citation rate, indicating its relative importance (Equation 2.5), where the total number of citations was obtained from Google Scholar [4].

$$CI = \frac{total\ number\ of\ citations}{number\ of\ years\ since\ publication} \tag{2.5}$$

2.8 Results

The results of the systematic literature review are now reported and organized according to the taxonomy introduced in Section 2.7.2 to answer the research questions, identify gaps in the literature, and discuss opportunities for future research.

A total of 557 papers were analyzed, published across 135 venues. Key publication sources include ACM CHI (109 papers), IEEE VR (53 papers), IEEE ISMAR (46 papers), ACM VRST (45 papers), and ACM CSCW (35 papers) (Figure 2.9). Trends in publication over the years show a significant increase, particularly from 2012 onwards, as illustrated in Figure 2.10. Notably, the number of papers surged in 2019 (69 papers) and 2020 (54 papers), a period likely influenced by the COVID-19 pandemic, which emphasized the importance of remote collaboration.

The types of user interaction and user experience in collaborative XR environments were also evaluated, categorized as follows: Visual + Haptics + Auditory (189 papers), Visual + Auditory (112 papers), Visual + Haptic (75 papers), and Visual only (22 papers) (see Figure 2.11b). Author contribution visualization, represented by bubble charts, provides insights of involvement of different authors highlihghts the level of engagement Figure 2.12.

To address **SRQ1** on how users effectively connect and collaborate with other users in XR collaboration spaces, the review reveals that effective collaboration in these environments hinges on several key factors. Users connect through various interaction modalities that incorporate visual, haptic, and auditory elements. Technologies such as avatars, spatial annotations, and 360-degree video are integral to facilitating these interactions. Additionally, the design of collaborative interfaces and platforms that support real-time interaction

[4]https://scholar.google.com/

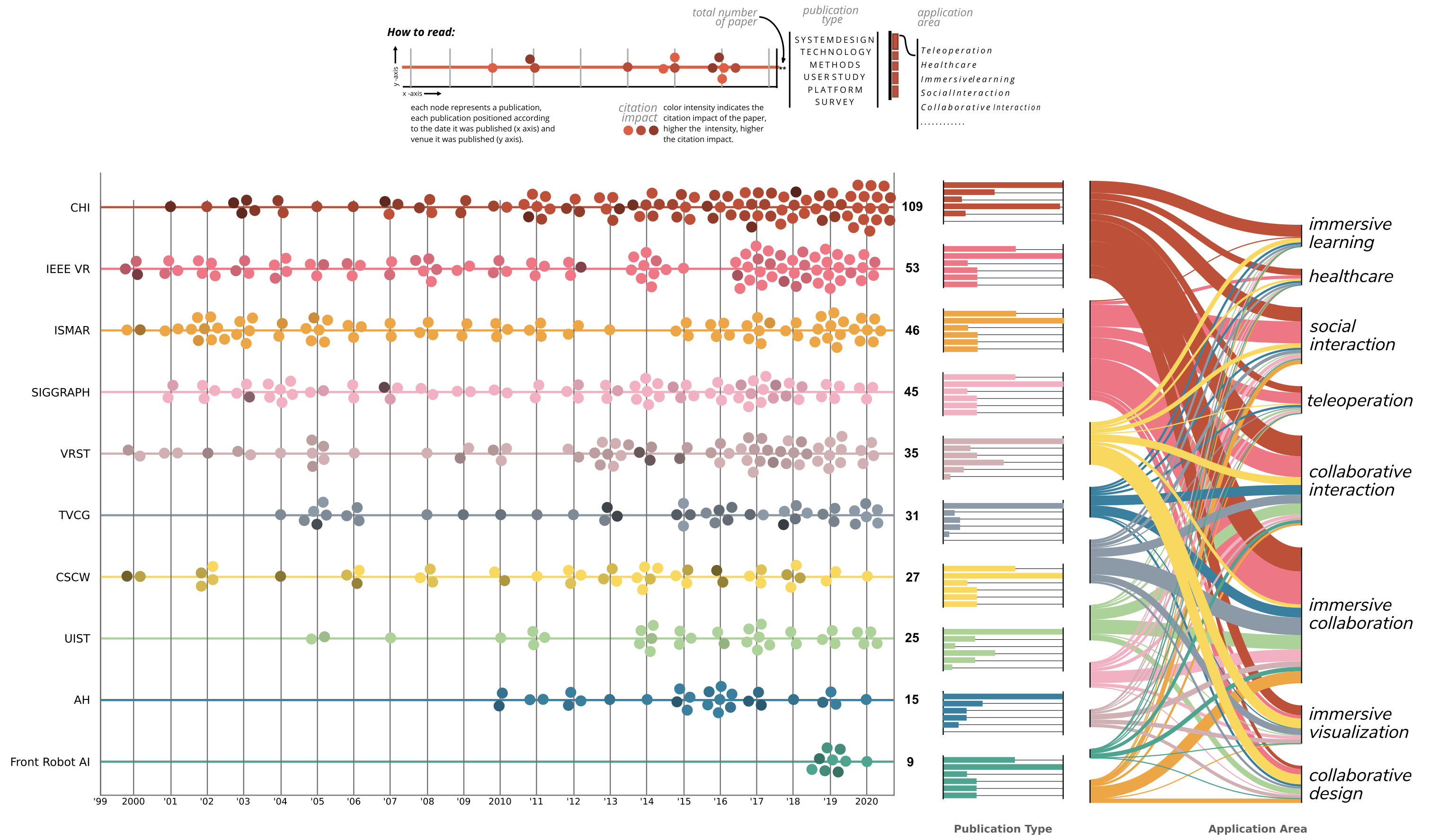


Figure 2.8: Distribution of 395 papers that were further assessed to determine the current XR collaboration trend are organized by year of publication, publication venues, citation impact, type of publication, and area of application. An interactive version of this classification is available at https://zaps.one/collabxr

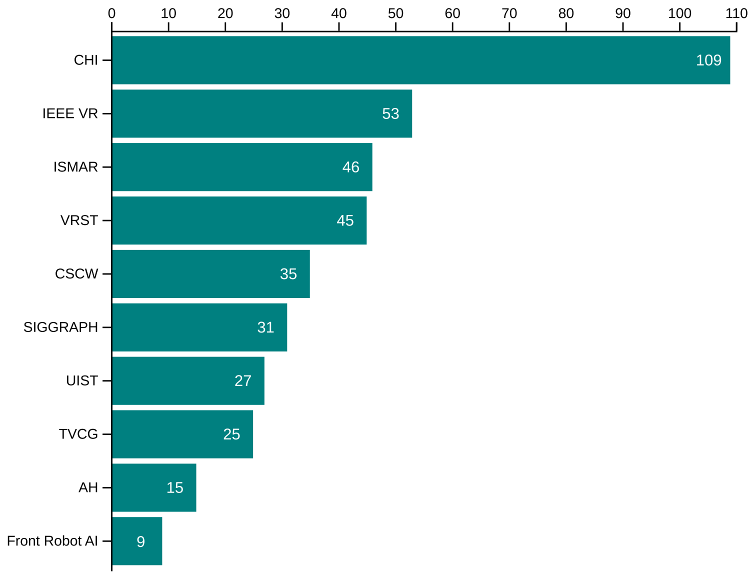


Figure 2.9: Distribution of publications across the top 10 venues.

and multi-user engagement is crucial, enhancing the overall effectiveness of collaboration among users in XR settings.

Regarding **SRQ2** on the types of collaboration styles, user experiences, and evaluation techniques used in XR collaboration, the findings show a diverse array of styles ranging from synchronous to asynchronous interactions, tailored to different tasks and user needs. The user experience is significantly enhanced by immersive technologies that provide realistic sensory feedback, such as haptic cues and spatial audio. Evaluation techniques commonly include user studies, usability testing, and performance metrics to assess the effectiveness of XR applications, focusing on user engagement, satisfaction, and the practical impact of XR technologies on collaborative tasks.

For **SRQ3**, concerning the current trends in XR collaboration and the main research challenges, the review identifies several emerging trends, including the widespread adoption of multi-user platforms, the integration of advanced rendering techniques, and the application of XR in fields like healthcare, education, and remote work. Key research challenges include network optimization, privacy and security concerns, and the need for refined interaction models to better support natural and intuitive user interactions. Addressing these challenges is essential for advancing the capabilities and usability of XR collaboration tools.

The distribution of 395 papers, which were further assessed to determine the current trends in XR collaboration, is organized by year of publication, publication venues, citation impact, type of publication, and area of application, as shown in Figure 2.8.

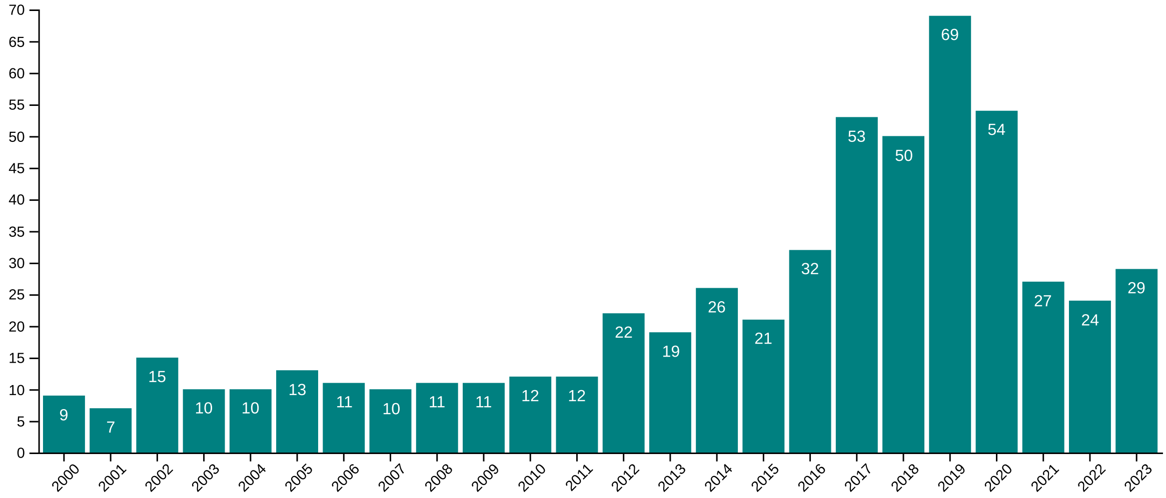


Figure 2.10: Distribution of publications by year.

2.8.1 Discussion

Key findings from the review underscore the importance of low-latency, real-time communication in XR collaboration, which is essential for inducing a sense of presence and immersion among users and enabling natural interactions like interacting with and manipulating virtual objects. Supporting multiple users, beyond traditional one-on-one interactions, with asymmetric compatibility ensures diverse teams can engage effectively within the XR collaboration space. In terms of collaboration styles and user experience principles, the focus is on spatial, remote, and asynchronous interaction, enhancing user experience through pres-

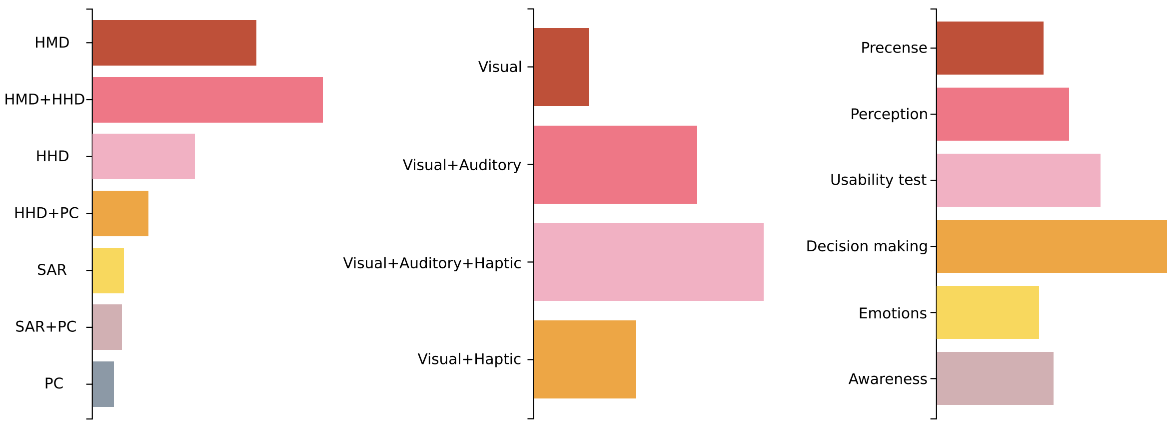


Figure 2.11: Device types (a), User interaction types (b), Evaluation criteria (c)

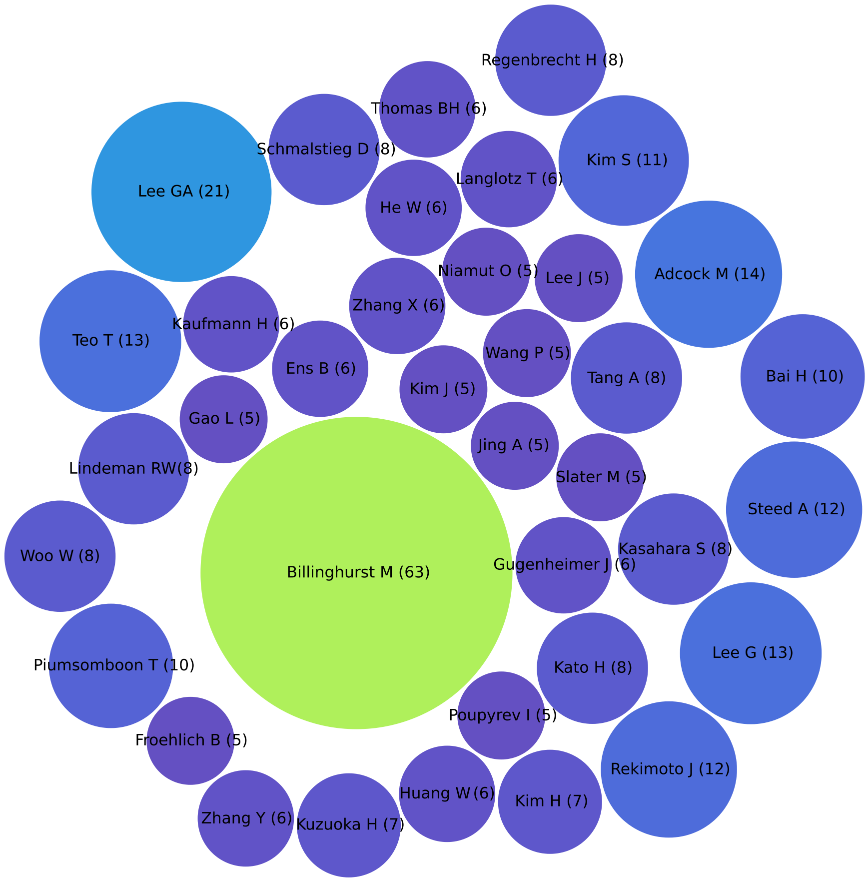


Figure 2.12: Author Contribution Visualization: Bubble radius proportional to number of publications.

ence, intuitive interaction design, and considerations for comfort and ergonomics. Evaluation techniques include usability testing, user studies, performance metrics, and qualitative analysis to assess the effectiveness and usability of XR collaboration tools and interfaces.

Current trends in XR collaboration highlight the adoption of multi-user platforms, AI integration, and industry-specific applications. However, challenges persist, such as network optimization, privacy concerns, and refinement of interaction models to enhance XR collaboration experiences. Addressing these challenges will be critical for advancing the capabilities and usability of XR collaboration tools in various sectors, including maintenance, medicine, engineering, and education.

Challenges

Several significant challenges have been identified that need to be addressed. These challenges span various aspects of XR collaboration, including technological advancements, user interactions, and the dynamic mapping of changing environments.

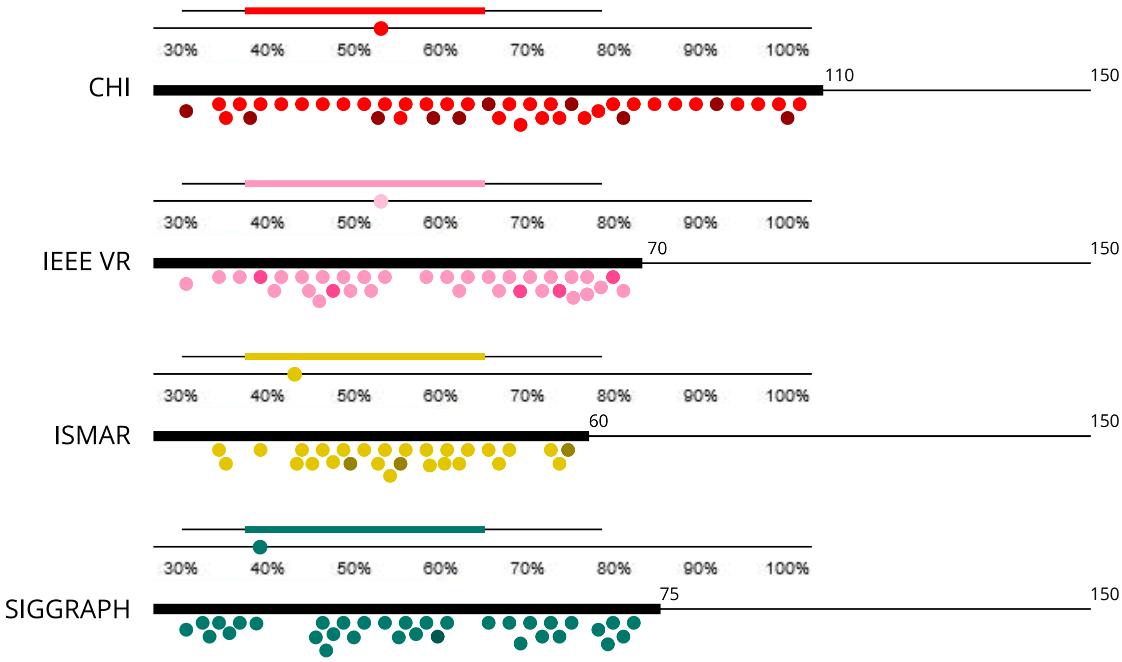


Figure 2.13: Distribution of User Studies Based on Participant Group Size

Real-time Dynamic 3D mapping of Changing Environments The need for real-time dynamic 3D mapping of constantly changing environments, such as those encountered in construction sites. The challenges faced in XR collaboration related to this include:

Handling Dynamic Environments: XR collaboration scenarios often involve real-time interaction in dynamic spaces where the environment is subject to continuous changes. This could include collaborative work in construction, remote inspection of changing manufacturing processes, or collaborative exploration of evolving landscapes.

Data Accuracy and Robustness: XR collaboration relies on precise spatial mapping and tracking of objects and users. Ensuring the accuracy and robustness of 3D mapping technologies in these dynamic environments is a significant challenge. Inaccurate or outdated mapping data can lead to misalignments and disruptions in collaborative XR experiences.

Integration of Emerging Technologies: To address these challenges, there is a growing need to integrate emerging technologies, such as drones equipped with advanced sensors and cameras, into XR collaboration setups. These technologies can facilitate real-time data capture and mapping updates in dynamic environments.

Multi-user Multimodal Capabilities The need for XR collaboration platforms to support multiple users interacting with shared models or environments using a variety of input methods and modes of interaction is crucial. This introduces the challenge of ensuring seamless synchronization and communication among users engaged in collaborative tasks. XR collaboration systems must address the complexities of managing user interactions and data exchange while maintaining real-time synchronization and preserving the integrity of the shared XR space. The challenge is to design and implement XR collaboration solutions that can handle diverse user inputs and interaction modes efficiently, allowing for effective communication and collaboration among participants. Additionally, ensuring a consistent and engaging user experience across different devices and modalities can be a significant challenge in this context.

Merging of Physical and Virtual Worlds The need for XR collaboration systems is to seamlessly integrate both physical and virtual environments. The challenge lies in ensuring that users with diverse XR hardware capabilities can collaborate effectively, even when they may not be fully aware of the differences in their hardware setups. With the increasing availability of various XR devices to the general public, it becomes important to investigate how each user perceives the mediated environments created by different hardware types, such as head-mounted displays, smart glasses, projection-based systems, or CAVE-like environments. This challenge involves understanding how these differences in display and interaction devices can impact users' roles and experiences in collaborative XR settings, ultimately striving for a cohesive and inclusive collaborative environment.

Diminished Reality and Real-time Occlusion Diminished reality and real-time occlusion technologies are increasingly relevant in the context of XR collaboration due to their potential to address certain challenges. In XR collaboration scenarios, the seamless integration of digital content with the physical environment is crucial for enhancing user experiences. Real-time occlusion, which hides virtual objects that are obscured by physical elements in a user's view, is essential for ensuring that digital augmentations do not clash with real-world surroundings. Diminished reality techniques, such as degrading or distorting visual fields, covering real objects with background imagery, overlapping virtual objects on real ones, and generating plausible background images, can aid in creating a cohesive and harmonious XR collaboration environment. By effectively implementing these techniques, XR collaboration can overcome challenges related to the interaction of virtual and physical elements, providing users with a more immersive and productive collaborative experience.

Security and Privacy Security in AR/MR-based remote collaboration is a concern, especially as users transition their collaborative tasks to cloud-based systems over the internet. Existing security solutions predominantly focus on transmission-level security, but there's a growing need to consider additional security measures to safeguard remote collaboration data, including machine structures and problem solutions. Achieving a balance between security and privacy is a complex challenge in MR development. It's essential to ensure users can engage in MR experiences confidently and securely. Robust security measures, such as encryption and secure authentication, are vital for protecting user data. The integration of intuitive interaction methods like gesture recognition, voice commands, and haptic feedback enhances the naturalness of interactions in the MR environment. Data integration with external sources and secure sharing options empower collaboration while ensuring data security.

Cutting-edge technologies like Quantum Key Distribution (QKD) and blockchain offer advanced security options in MR. QKD enhances security by providing theoretically immune encryption keys, making MR communication highly secure and resistant to eavesdropping. Blockchain technology ensures data integrity and transparency, making it suitable for secure data storage and authentication within MR environments. Homomorphic encryption allows computations on encrypted data without revealing the data itself, which is crucial for protecting sensitive information. Additionally, Secure Multi-Party Computation (SMPC) supports collaborative computation over private inputs, preserving data privacy while enabling cooperative tasks in MR environments. These advanced security measures contribute to a safer and more robust environment for AR/MR-based remote collaboration.

New Trends

In XR research, significant updates and changes in trends have emerged. Some of the new research trends and observations were already introduced in the previous section, such as Reconstruction and Collaboration. Here, two important changes in AR research from the 2000–2023 publications are explicitly presented, namely, sharp increases in Evaluation and Rendering research.

Group Size The literature has not thoroughly explored the impact of different group sizes on collaboration mechanisms, communication, and social interaction in XR environments, particularly for remote users. The emphasis in research has often leaned toward face-to-face or dyadic collaboration, mainly due to network bandwidth limitations and the substantial data transfer requirements for smooth collaboration. This section delves into current research trends related to collaborative XR systems, considering how collaboration techniques in communication and interaction vary across dyads, triads, and large groups.

Digital Twin Digital twins are integral to XR collaboration, serving as virtual representations of physical spaces or objects within XR environments. These twins are linked to their physical counterparts through sensors and data collection devices. They empower users to actively observe, analyze, and interact with physical elements in real-time within the XR setting. This capability fosters improved collaboration by offering a virtual space for simulating and refining real-world processes, thus diminishing errors and inefficiencies. Furthermore, digital twins facilitate the development of tailored and eco-friendly products within XR collaboration contexts, aligning with the overarching trend of using XR for more effective and sustainable collaborative endeavors.

Multi-user Supports Supporting multiple users in collaborative XR scenarios, especially in one-to-many setups, is becoming more relevant and reflects real-life collaborative activities. XR technology advancements enable the exploration of more complex and realistic collaborative problems that were previously only discussed theoretically. The trend highlights the increasing importance of one-to-many XR scenarios, which has been amplified by the COVID-19 pandemic, where remote activities like online meetings and virtual events have become common. XR technologies offer an opportunity to enhance traditional one-to-many methods (such as audio and video streaming) by improving user experiences, performance, and decision-making in collaborative tasks.

Tech Advancement Technological advancements are driving an increased interest in using XR technologies for entertainment and collaborative purposes. These advancements include improvements in mobile connectivity, processing capabilities, wearable devices, and tracking technology. The importance of providing immersive and engaging experiences to enhance users' sense of presence is crucial for XR collaboration scenarios. It also emphasizes the need for natural user interactions in XR collaboration, such as gaze-based interaction and non-verbal cues, to create more effective and enjoyable collaborative experiences. The importance of developing visually appealing and engaging content for XR collaboration scenarios, especially in interdisciplinary collaboration between fields like creative arts, psychology, and technology, enhances the overall XR collaboration experience, indicating a trend toward more comprehensive and user-centric XR collaboration approaches.

Research Gap and Future Work

While the outlined research questions provide a comprehensive framework for exploring various aspects of XR collaboration, there are several gaps in understanding how to effectively implement and optimize these strategies in practical XR collaboration scenarios. Specifically, the gap lies in identifying and developing innovative solutions that seamlessly integrate teleportation, telepresence, representation, interaction, viewpoint sharing, and mixed-perspective collaboration to enhance user experiences and facilitate meaningful collaboration in MR spaces.

Real-time communication is crucial for applications such as XR remote collaboration where participants need to interact without noticeable delays. Achieving low-latency communication involves overcoming technical hurdles in network infrastructure, data processing, and protocol optimization. Ensuring consistent performance across different devices and network conditions is complex, particularly as the number of simultaneous users increases. High visual fidelity is essential for these immersive experiences, where detailed graphics enhance user engagement and comprehension. Maintaining high visual fidelity requires substantial computational resources, which can strain devices and networks. Balancing graphical quality with performance and bandwidth constraints presents ongoing challenges in optimization and compression techniques.

Enabling multiple users to interact simultaneously in a collaborative XR space presents a significant challenge, especially as most applications require scaling beyond one-to-one settings. Asymmetric compatibility where users may have varying hardware capabilities, roles, or network conditions introduces complexity in synchronization, data sharing, and interface design. Ensuring a seamless experience for all participants regardless of their setup is a significant challenge.

Supporting multiple viewpoints and collaborative perspectives enhances the richness of interactions, enabling more dynamic and engaging experiences. Integrating and synchronizing multiple viewpoints in real time, especially with high visual fidelity, requires efficient data transmission and processing. Coordinating interactions among participants with different roles or access levels adds complexity to user interface design and interaction models. Maximizing performance and reliability across varying network conditions (like bandwidth limitations, latency fluctuations) demands continuous improvements in protocol efficiency and adaptive strategies. Safeguarding user data and interactions in shared environments requires robust security measures, encryption techniques, and privacy-preserving protocols.

Key findings from the review underscore the importance of low-latency, real-time communication in XR collaboration, which is essential for inducing a sense of presence and immersion among users and enabling natural interactions like interacting with and manipulating virtual objects.

This thesis aims to address some of the research gaps outlined above, especially focusing on identifying and developing innovative solutions that seamlessly integrate teleportation, telepresence, representation, interaction, viewpoint sharing, and mixed-perspective collaboration to enhance user experiences and facilitate meaningful collaboration in MR spaces.

For access to supplementary materials, documents, and interactive visualizations, including Excel files containing the raw survey data used in crafting the taxonomy, please visit the following URL:
https://zaps.one/collabxr

Chapter 3

MRMAC: Mixed Reality Multi-user Asymmetric Collaboration System

This chapter presents MRMAC, a Mixed Reality Multi-user Asymmetric Collaboration system that allows remote users to teleport virtually into a real-world collaboration space to communicate and collaborate with local users. The system achieves telepresence for remote users by live-streaming the physical environment of local users using a 360° camera, while seamlessly integrating 3D virtual assets into the mixed-reality collaboration space. The novel client-server architecture implemented allows for asymmetric collaboration among multiple AR and VR users, incorporating features such as avatars, view controls, and synchronized low-latency audio, video, and asset streaming.

3.1 Introduction

Immersive telecollaboration has the potential to enable remote collaborators to work and communicate effectively as if they were physically present in the task space [159]. While prior work focused on virtual environments created using computer-generated 3D assets [225, 201], recent research has transitioned towards integrating users' real physical environments, thus enabling telecollaboration within real-world task spaces [192, 237, 267].

One approach involves an AR-VR setup where a remote VR user can virtually teleport into a live-streamed real-world environment to interact with a local user [193]. The local user perceives the remote user as a 3D avatar, which is well-suited for remote assistance and training and induces higher social presence compared to 2D video conferencing [193, 128, 194]. To create such a system, it's crucial to capture the local users' physical

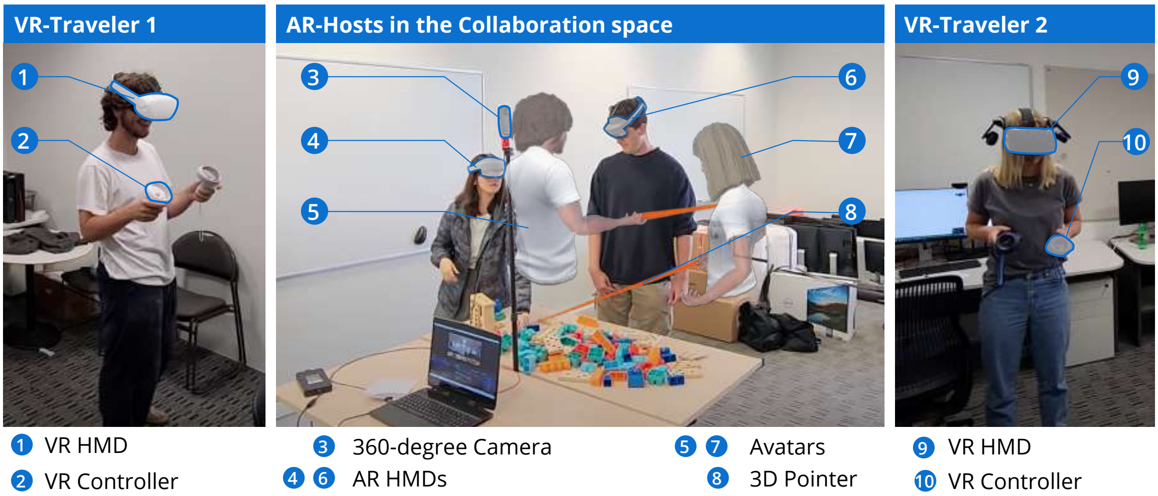


Figure 3.1: The MRMAC system, including the local AR-Hosts and remote VR-Travelers, 360° camera, controls, and display setup.

environment and livestream it to remote users to ensure their telepresence and situation awareness. Earlier work achieved this by either live-streaming remote environment as 3D point cloud [2, 196] or using 360° cameras [228, 231, 193]. While previous designs and implementations of AR-VR asymmetric telecollaboration systems primarily centered around one-to-one interactions (a single user on each side), many applications require multiple users to collaborate together [97, 268]. However, enabling multi-user interaction is challenging due to the large-scale data size that increases overall latency and the need for effective protocols that facilitate seamless interaction among AR-VR users.

This chapter presents MRMAC (Figure 3.1), a Mixed Reality Multi-user Asymmetric Collaboration system enabling multiple remote users to virtually teleport into a real-world task space for collaboration with local users. The system includes a design framework, key features supporting multi-user protocols, and a novel client-server architecture facilitating asymmetrical telecollaboration among multiple AR and VR users.

Parts of the work presented in this chapter and the following chapter have been published in the Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (Zaman et al., 2023) [269].

The contributions in this chapter are summarized as follows:

- A **design concept and protocol** for a multi-user asymmetric remote collaboration system enabling bidirectional face-to-face communication, high situational awareness, and synchronized audio-visual communication between multiple remote and local users.

- The **implementation of MRMAC**, including a novel client-server architecture to enable asymmetric telecollaboration for multiple AR and VR users with avatars, view controls, and low-latency video and asset streaming with synchronization.

3.2 Related Work

Multi-user collaboration has been a research topic for many years and has become increasingly important with the rise of XR technologies, including AR, VR, and MR. Although it can be challenging to implement these systems due to the need to combine different modalities and overcome technical limitations, providing good support for multi-users in XR experiences is crucial for effective collaboration.

3.2.1 Multi-user Collaboration in VR

In multi-user VR collaboration, users can collaboratively experience and manipulate objects in virtual environments. Multi-user VR collaboration systems have been explored in various settings, such as co-located, remote, and both co-located and remote users' environments. For example, DIVE [40], one of the earliest remote multi-user VR systems, allowed users to participate in a shared 3D virtual world and collaborate with one another while maintaining a high level of situational awareness.

Collaborative platforms such as MMVR [143] and mCLEV-R [150] are examples of VR collaboration systems used in a variety of application domains, including conferences, presentations, prototyping, and design. Collaboration in VR has gained popularity in many relevant industries, such as culture [244], communication and collaboration [248], education and training [181], and consultation and therapy [262].

Social VR, another popular area for multi-user VR collaboration, has seen the emergence of platforms such as Facebook Horizon, which enables users to build, explore, and socialize in a virtual world [276]. Another social VR application, Metamorphic [157], allows remote participants to modify their appearance by becoming brush strokes as they move through virtual space in a series of majestically painted worlds. Additionally, CocoVerse [81] allows multiple users to connect in VR to co-create a 3D whiteboard for teaching and learning.

Synchronous asymmetrical interaction for both co-located and remote users has also been explored, as demonstrated by TransceiVR [238], which allows users to freeze and annotate content in the virtual environment. Similarly, ShareVR [82] and CoVR [57] enable external observers to view and interact with the virtual environment independently via projection or touch displays.

Overall, the development of multi-user VR collaboration systems has demonstrated potential in communication across various domains. However, limitations, such as reduced real-world awareness, limited physical interaction, and less accurate spatial awareness, prompted further research into AR and MR technologies.

3.2.2 Multi-user Collaboration in AR

Collaboration in AR enables natural interaction by allowing users to perceive the real environment. Early multi-user AR experiences like AR2 Hockey [160] and AquaGauntlet [224] focused on social communication aspects, with co-located users interacting with virtual objects and characters using tangible interactions. EMMIE [37] integrated tangible interaction into a heterogeneous 3D environment, enabling local or remote users to collaborate using various devices and displays while providing instructions to remote users. Lukosch et al. [138] emphasize the importance of designing AR systems that support natural interactions and enhance collaboration experiences.

In co-located multi-user AR, having multiple users in the same physical space is useful for sharing ideas. VITA [19], an AR system, allows multiple users to interact simultaneously with large-scale and full-size archaeological excavation data. Mahmood et al. [140] and Buschel et al. [36] explore in-situ visualization tools where users can choose from a list of predefined visualizations or create a visualization through a simple menu in an AR application to explore GeoSpatial data.

Collaborative AR has also been explored in areas such as architectural design and urban planning [18, 34], education [11], and collaborative face-to-face meetings [21]. CollabAR [135] explores handheld group collaboration in a co-located environment, offering several features such as mobile AR as a common collaboration tool to support asynchronous collaboration. Early work utilizing point clouds for remote collaboration in AR demonstrated collaborative analysis among crime scene investigators by integrating real-time map-making features [177].

Remote collaboration in AR has been shown to increase social presence [25] and provide a natural way for remote users to interact with the local environment. It also enables unique interaction opportunities, such as using a tabletop projector for multi-scale interactions between remote and local users [217], and sharing AR task information and real-time annotation updates among remote users, as shown in CARS [273]. Remote collaboration in AR also showed to improve situational awareness [139].

Multi-user AR collaboration between remote and local users has proven effective in enhancing remote telepresence through video conferencing [127], training [246], and remote assistance [78, 98]. Addressing cross-platform multi-user synchronization issues [183, 8] has helped reduce spatial inconsistency and latency in remote multi-user AR scenarios [189]. These advancements underline the potential of remote AR collaboration to enable natural interactions and foster unique collaborative opportunities.

3.2.3 Multi-user Asymmetric Collaboration in MR

Asymmetric MR collaboration involves individuals playing different roles in a collaborative task utilizing different device settings and having different levels of authority in the collaborative environment (host/guest) [65]. Multi-user asymmetric MR collaboration has been shown to be useful in remote collaboration situations, where a remote VR user assists a local AR user using hand gestures [71, 169], annotations [74], gaze [245, 170], virtual replicas of task objects [158], haptics [176], and more [73, 105, 61].

One of the earliest multi-user mixed-space asymmetric collaboration systems was presented by Kiyokawa et al. [116], in which users can seamlessly transition from an AR world to the VR world. A similar technique is used in Magic book [23], where users can be in either an exocentric AR view or an egocentric VR view. The use of different viewpoint perspectives in asymmetrical collaboration has proved useful in many application areas, such as engineering prototyping [49], architectural design [95], and visualizing complex virtual content [195].

Tong et al. [240] conducted controlled user studies on collaborative visualization in a distributed asymmetric setting, specifically focusing on a VR setting, and found that a well-designed asymmetric collaboration system as effective as symmetric systems, with PC users perceiving less mental demand and effort in asymmetric PC-VR settings compared to symmetric PC-PC settings, highlighting the importance of considering various collaboration settings.

Live 360° video is becoming increasingly popular for remote collaboration [228, 117] as it allows the remote user to see the entire workspace independently of the local users' positions. For example, Poly [118] mounted a 360° panoramic camera on a backpack monopod to show the wearer's surroundings from an exocentric perspective and allow for real-time remote collaboration on a 360° panorama. To facilitate nonverbal communication between two users using AR or VR displays, Lee et al. [129] developed a system that allows hand gestures and real-time viewing of awareness cues in a 360° panorama captured by a head-mounted camera.

Recent research has focused on collaboration in real user environments, where multiple users can engage in real-time [128, 130, 193]. Various techniques have been developed, such as procedurally generating virtual worlds from 3D reconstructed physical spaces, as demonstrated in studies by Sra et al. [216] and Lindlbauer et al. [134]. Real-time, high-quality 3D reconstructions of physical spaces have also been achieved by Orts-Escolano et al. [161], enabling low-latency communication between remote users as if they were co-present in the same physical space. These advancements have significant implications for remote collaboration, particularly in fields where visualizing and manipulating 3D environments is essential. Incorporating spatial auditory cues has been shown to help guide local workers in collaborative spaces, enhancing their spatial awareness in mixed reality remote collaboration tasks [263]. These findings suggest that integrating spatial auditory cues can enhance local-remote conversations, making them more interactive and intuitive. Spatial auditory cues provide information about directions and distances in a 360-degree range. When combined with visual cues, they create an immersive experience, making participants feel as if they are interacting with a real person.

Advancements in multiuser asymmetric MR collaboration have significant implications for remote collaboration. However, there is a need for more research on the design and implementation of such systems and techniques, especially in users' physical environments, and to support multiple users in these environments.

3.3 MRMAC Design

In this section, the design aspects of the MRMAC framework are described, encompassing several key components, including asymmetric telecollaboration, and multi-user collaboration. The design decisions of MRMAC were derived from the literature review of prior research described in previous section. The section also outlines key design features such as communication, awareness, streaming and syncing, multi-modal interaction, and content management, which are guided by the key insights from the literature review. These components work together cohesively to create an immersive collaborative experience. User interfaces are customized to provide intuitive and user-friendly methods for participants to engage with the MR environment.

3.3.1 Asymmetric Telecollaboration System

An asymmetric telecollaboration system combines Augmented Reality (AR) and Virtual Reality (VR) technologies to provide a mixed reality collaboration (MRC) space [193]. This MRC space facilitates communication and collaboration between physically co-located local users, AR-Hosts, and remote users, known as VR-Travelers, who virtually teleport to the collaboration space from a remote location.

The MRC space serves as a meeting point for AR-Hosts and VR-Travelers, allowing them to interact collaboratively through interactive visual cues, such as 3D virtual objects blended into the physical collaboration space in real-time. In the prior implementation [193], a VR-Traveler can view an AR-Host through live streaming media, such as 360° videos on the VR Head Mounted Display (HMD), while the AR-Host can see and communicate with the VR-Traveler's 3D avatar displayed on the AR HMD.

3.3.2 Multi-user Asymmetric Collaboration (MAC)

A novel multi-user asymmetric collaboration system is proposed, supporting AR-Hosts and VR-Travelers with distinct communication protocols and collaboration activities. The system enables multiple users to participate on both sides and facilitates effective collaboration. Let H_n denote the n hosts in the collaboration space, and T_m represent the m travelers in the remote space. Communication channels between H_n and T_m are established through unidirectional ($\rightarrow$ or $\leftarrow$) or bidirectional ($\rightleftarrows$) connections. The symbol $\circlearrowleft$ indicates communication exclusively within the hosts or travelers. Four scenarios for multi-user asymmetric collaboration setups in the MRC space are categorized as follows:

- **One-to-One ($T_1 \rightleftarrows H_1$; $T_1 \rightarrow H_1$; $T_1 \leftarrow H_1$):** describes a typical scenario for a meeting between two people (Figure 3.2A). It requires a bi-directional communication that allows a remote user (VR-Traveler) to communicate with a local user (AR-Host) and explore the local user's collaboration space [193].
- **One-to-Many ($T_1 \rightleftarrows H_n(\circlearrowleft)$; $T_1 \rightarrow H_n(\circlearrowleft)$; $T_1 \leftarrow H_n(\circlearrowleft)$):** describes a scenario in which a single remote user interacts with multiple hosts at a location, such as a remote lecturer visiting a physical class with multiple students (Figure 3.2B). For this case, the required system is similar to the One-to-One case, as the physical space and multiple users on site can be captured and streamed similarly.
- **Many-to-One ($T_m(\circlearrowleft) \rightleftarrows H_1$; $T_m(\circlearrowleft) \rightarrow H_1$; $T_m(\circlearrowleft) \leftarrow H_1$):** a scenario with a group of remote users interacting with a single local user, such as in a classroom setting with one lecturer and multiple remote students (Figure 3.2C). The required system necessitates additional features to enable efficient communication between multiple VR-Travelers and the AR-Host.
- **Many-to-Many ($T_m(\circlearrowleft) \rightleftarrows H_n(\circlearrowleft)$; $T_m(\circlearrowleft) \rightarrow H_n(\circlearrowleft)$; $T_m(\circlearrowleft) \leftarrow H_n(\circlearrowleft)$):** describes a scenario where multiple remote users interact with multiple local users simultaneously, creating a complex web of communication (Figure 3.2D). The system necessitates a

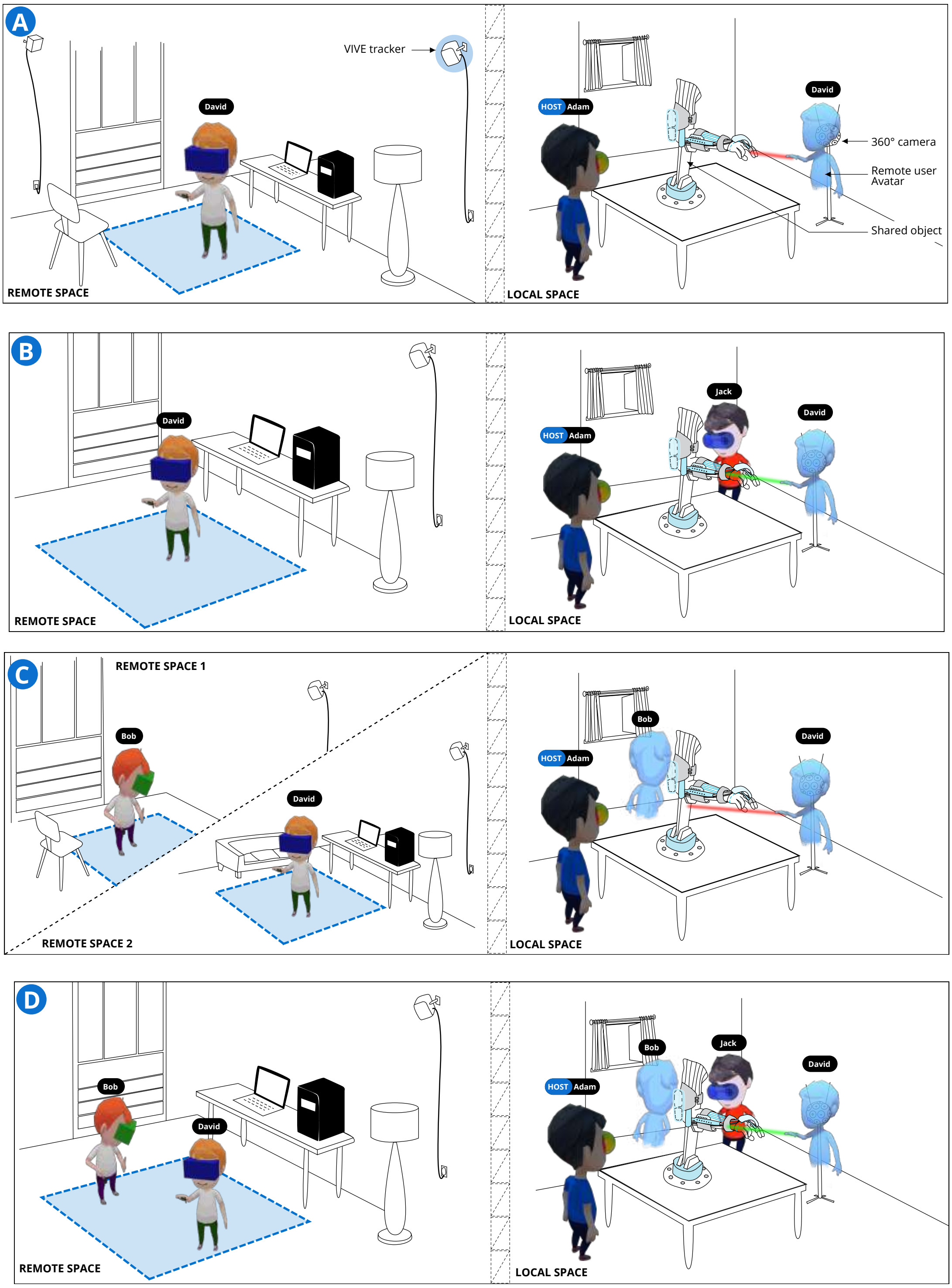


Figure 3.2: Different types of collaboration setup in a multi-user asymmetric MR collaboration. Remote space to Local space **(A)** One-to-One ($1 \rightleftarrows 1$), **(B)** One-to-Many ($1 \rightleftarrows n$), **(C)** Many-to-One ($n \rightleftarrows 1$), **(D)** Many-to-Many ($n \rightleftarrows n$).

communication channel between VR-Travelers and AR-Hosts, as well as among VR-Travelers and AR-Hosts themselves. For instance, consider a scenario where a group of medical students collaborate on a virtual surgery simulation, with multiple students located in both physical and remote sites.

Solutions like Microsoft Remote Assist [1] have emerged, combining the physical and digital worlds to support collaboration in multi-user settings. However, remote users have limited physical presence and interaction capabilities. There are also solutions like Microsoft Mesh [2], which allow collaborators to come together in a shared virtual space for collaboration with multiple users. However, this platform does not enable users to see their collaborators' physical spaces. Thus, telepresence systems, especially those designed for collaboration in physical spaces where users can view collaborators' spaces in real-time, typically facilitate one-to-one communication. Therefore, MRMAC allows multiple remote collaborators to collaborate in local users' physical space as well as allows remote users to see the local user's physical space. To the best of my knowledge, prior work has yet to provide a unified solution and evaluation for these multi-user asymmetric telecollaboration scenarios.

3.3.3 Key Design Features

The MRMAC aims to provide effective communication and awareness cues for all four scenarios. The system addresses the multi-user communication, awareness, streaming, and synchronization needs of the *One-to-Many*, *Many-to-One*, and *Many-to-Many* cases.

Communication

The MRMAC facilitates both verbal and visual communication to enable collaborative work. Verbal cues are crucial for effective communication and collaboration between remote and local users, providing important contextual information that cannot be conveyed through text or visual cues alone. In addition, visual cues such as a virtual pointer, annotation, and hand gestures can facilitate rich collaborative behaviors and efficient instructions.

Awareness

Awareness cues help users understand each other's activities, resulting in better coordination and conflict avoidance in the shared space. Furthermore, these cues enable remote

[1] https://dynamics.microsoft.com/en-us/mixed-reality/remote-assist/

[2] https://www.microsoft.com/en-us/microsoft-teams/microsoft-mesh

users who are not physically present to stay informed and collaborate with other users. For example, multiple 3D avatars represent each user's identity while sharing their spatial communication cues. Additionally, users' viewpoints are essential cues that indicate where their collaborators are looking, which can help discuss spatial distribution or objects in view.

Streaming and Synchronization

Multi-user telecollaboration requires low-latency network streaming and synchronization techniques to ensure seamless communication. This involves ensuring seamless audio and video components streaming through the network and synchronizing devices and data across the remote multi-user network to ensure all users are experiencing the same level of network latency.

3.4 MRMAC Implementation

This section describes the implementation of MRMAC, where the theoretical design is put into practice. MRMAC implementation involves deploying various technical components, including network architecture, visual and verbal cues, awareness and communication cues, software development, and hardware integration, to create a functional mixed-reality collaboration system.

3.4.1 Architecture Overview

A novel client-server architecture is presented to implement MRMAC as a complete system prototype that addresses the four key design features outlined in Section 3.3.3. Unlike traditional architectures that may focus solely on either physical or virtual environments, this architecture seamlessly fuses both realms. The use of a Node.js server as a local gateway introduces a decentralized approach to managing users and objects within the network allows for efficient handling of data, device operations, and API requests in the local environment. The architecture prioritizes real-time synchronization among users and servers to ensure that users experience minimal latency and disruptions. The Interaction Server synchronizing continuous data streams and supporting multiple communication protocols (UDP and TCP). Overall, the architecture's novelty lies in its ability to enhance the user experience through seamless integration, real-time synchronization, and efficient management of interactions across physical and virtual spaces.

The system offers communication and awareness cues, enabling synchronized asymmetric collaboration between an AR-Host site and a VR-Traveler site.

AR-Host site: A 360° camera captures an omnidirectional view of the physical collaboration space. The 360° camera and the camera integrated into the AR-HMD live-stream to VR-Traveler sites and blended visual cues from multiple users at the AR-Host site (left side Figure 3.3). OBS Studio was used to encode and stream the video footage to a designated server endpoint. The videos streamed with OBS Studio were encoded using the x264 library. Parameters were configured as follows: presets - ultrafast, profile - baseline, tune - zerolatency.

VR-Traveler site: Users receive the live-streamed 360° video from the local user and have telepresence at the local user's site through the live-streamed 360° video shown in their VR HMDs. Their perspective views of the 360° video and visual cues are then shared with AR-Hosts and other VR-Travelers. VR-Travelers are represented as 3D avatars, and their visual cues are displayed in users' VR-HMDs.

Network Architecture: The client-server architecture is designed to seamlessly merge physical and virtual spaces, fostering cohesive user interactions and connectivity with surrounding users, devices, and virtual objects. A Node.js server acts as a local gateway overseeing users and objects within the network. It's segmented into three crucial components: the user data handler, object data handler, and API request handler. These components collaboratively manage user and object data, device operations, and API requests within the local environment. The Node.js server plays a pivotal role in orchestrating real-time synchronization among users and servers, managing local device interactions, and facilitating seamless communication of updates across the network. Interaction Server is dedicated to managing and synchronizing continuous data streams for seamless global interactions. It utilizes both UDP and TCP protocols to deliver data streams across users, while access control is managed by the node server. By supporting multiple communication protocols (UDP and TCP), the Interaction Server ensures uninterrupted communication and synchronization of user interactions across the globe, enhancing the overall user experience.

The system architecture and data flow of MRMAC are illustrated in Figure 3.3. The solution is fully integrated into the networking environment and implemented using the Unity 3D game engine with WebXR Exporter. Google's open-source Web Real-Time Communication protocol (WebRTC) [137] is utilized for networking on a dedicated NodeJS server, which receives application requests over HTTP requests.

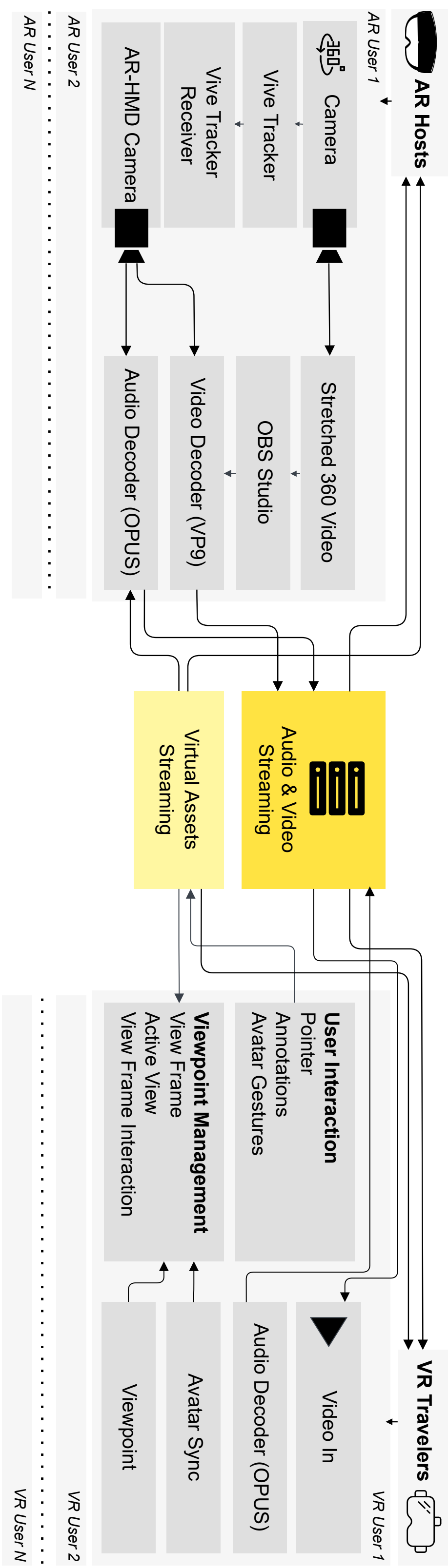


Figure 3.3: MRMAC: overview of the system architecture and the network data flow, demonstrating how the primary processing pieces are mapped to the VR-Travelers and AR-Hosts.

3.4.2 Communication

Verbal cues

MRMAC provides a voice chat feature via audio streaming from each HMD's microphone to all collaborators, allowing remote and local users to speak and communicate verbally in real time. Spatial audio features are utilized to effectively position sound within the collaboration space by tracking remote users' head and device movements. Both AR-Host and VR-Traveler audio are captured at a sampling rate of $44.1kHz$ and encoded using the OPUS voice encoder. To prevent echo, audio played out of the HMD speaker is removed from the signal captured by the microphones. The MagicLeap Soundfield Audio Plugin [3] for Unity is employed to implement audio spatialization.

Visual cues

Visual cues have been integrated to enhance user communication and collaboration. Three key visual cues are shared from the remote to the local users (see Figure 3.4):

Annotation: Free-form 3D drawing annotations are facilitated through 3D hand tracking via VR controllers. Users can create annotations by projecting a pointer from the controller position onto a plane surface, resulting in the generation of a mesh object at the collision point. This mesh object functions as a canvas for drawing, with its position and orientation dynamically adjusted in real-time to mimic the movements of the controller.

3D pointer: the 3D pointer is added to indicate spatial targets accurately. This was implemented in Unity using a raycast to follow the VR controller direction of the remote user.

Hand gestures: the remote travelers' hand gestures are captured using VR controllers, and their transformation is streamed to the local site to control avatars' 3D hands.

To ensure synchronization with other users in the MRC space, both the 3D pointer and the drawn annotations are registered. This means that the position and orientation of the annotation object are synchronized with other users' devices, allowing all participants to see the same annotations in the same position and orientation.

[3] https://ml1-developer.magicleap.com/en-us/learn/guides/soundfield-user-guide-for-unity

Figure 3.4: Communication and awareness cues in MRMAC: Annotation, 3D pointer, and avatar embodiment in collaborative environments.

3.4.3 Awareness

MRMAC enables remote situational awareness by live streaming high-definition video of the physical local space, complemented by augmented visual cues for remote collaboration. Multi-user presence and awareness in collaboration scenarios are further enhanced through the inclusion of multiple 3D avatars and a view-sharing mechanism.

3D Avatars

Using 3D avatars to represent remote users is critical for enhancing their co-presence and awareness in multi-user collaboration scenarios (see Figure 3.1). However, controlling the movements of remote avatars, positioning them within the MRC space, and preserving their distinct identities present challenges. The system addresses these challenges through the following solutions:

Remote avatar control: VR-Travelers are represented by 3D avatars, and their head and hand movements are captured using HMDs and VR controllers. The tracked motion data is used to animate the avatar using Inverse Kinematics (IK). This process involves setting up the character's avatar, animator controller, IK goals, solver, and pass and adding animation clips. Next, the position and orientation of the VR controller are tracked to update the corresponding IK goal, while finger motions are abstracted into predefined gestures. Finally, the avatar's position and rotation are broadcasted and synced across all user displays.

Avatar positioning: In the prior one-to-one setup [193], the remote user's avatar is placed in the center of the 360° camera. However, multiple remote users can cause issues such as overlapping and difficulty identifying collaborators. To ensure that avatars for both users (A and B) face the same direction relative to the world space, the difference in camera rotation between user A and user B ($\Delta\alpha = \alpha_2 - \alpha_1$) is calculated. The initial avatar rotation and position for user B are adjusted by this amount, denoted as $\beta_2' = \beta_2 + \Delta\alpha$, and

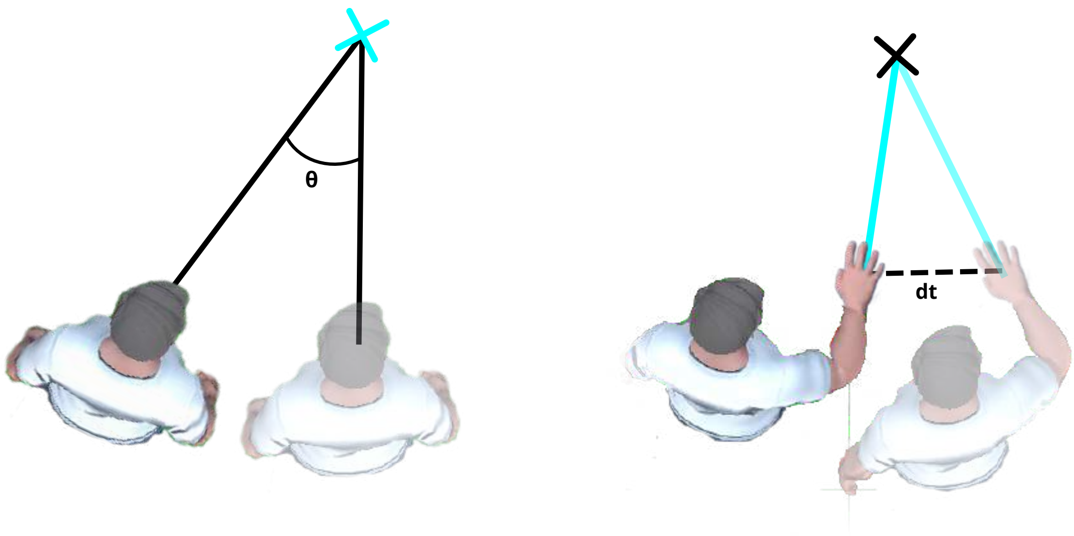


Figure 3.5: Multiple avatar position control includes (a) Applying offset compensation to shift the avatar from its original position, and (b) adjusting the drawing of the 3D Pointer based on the offset compensation.

$p_2' = p_2 - R(\Delta\alpha)(p_2 - p_1)$, where $R(\Delta\alpha)$ represents the 3D rotation matrix for the angle $\Delta\alpha$ around the origin.

An additional offset between the initial rotations ($\theta = \beta_2' - \beta_1$) and positions ($d = \| p_2' - p_1 \|$) of the two avatars allows for manual adjustment of user position. Further adjustment of the avatar rotation and position for user B can be made using this angle and distance: $\beta_2'' = \beta_2' - \theta$ and $p_2'' = p_2' + R(\theta)(d, 0, 0)^T$. Although the avatars and camera are placed in the center of the 360° video, the positions, and rotations offset to make them appear to be far apart from each other. Multiple avatars are positioned in a circle around the camera's center, with each avatar placed at a different position along this circular perimeter. The radius of the circle determines the distance from the camera, while the arrangement of the avatars follows a clockwise order with a specified angular offset between each subsequent avatar.

It's worth noting that the first VR-Traveler to join the session is a special case, placed at the center camera position with no offset. In the experimental setup, the radius is defined as 2m and the angular offset as 30 degrees. In cases where the perimeter reaches capacity with many avatars, a second perimeter with 1.5 times the radius distance is generated, and the clockwise offset pattern continues (see Figure 3.5). All remote users' view orientations are positioned at the center of the 360° video, while their avatar positions are visualized with an offset from the center. The gap between the avatar position and the user's view orientation can potentially introduce false information, impacting communication. Therefore, in the experiments, an offset was carefully selected to separate the avatars from each other while ensuring that the difference in positioning was not noticeable to the participants. Additionally, nametags were implemented to display remote user names above their avatars. These nametags appeared only when remote users were looking towards local users or when they were in close proximity.

Personalized avatar generation: Avatars were generated in real-time using Ready Player Me [4] and the Headshot plugin [5], with a Python wrapper to facilitate the process. Profile pictures of remote users were taken and sent to the API, which then generated personalized avatar configurations based on parameters such as LOD, texture size, mesh, and facial features (see Figure 3.6). Due to time constraints in the user study, participants were not allowed to make further customizations.

View Sharing

The view-sharing feature allows collaborators to see each other's perspective in real-time, which is useful for discussing spatial distribution or objects within view. VR-Travelers and AR-Hosts can share their viewpoint through a picture-in-picture (PiP) window displayed on a 2D plane [148, 247, 125]. This allows all users to have a common understanding of the situation and switch between different views to highlight specific objects. The PiP window can be moved to any relative position and orientation to avoid head movement interference (see Figure 3.7).

3.4.4 Streaming and Synchronization

The MRMAC architecture includes network components on both the server and client sides to facilitate media streaming and synchronization of all related data. Media stream data is transferred directly between connected peers, and establishing a connection involves procedures to control communication and exchange metadata. The client manages media streams and virtual and augmented objects; their properties are synchronized over the network. The server is designed to distribute a large amount of tracking data to each user while keeping latency to a minimum. Figure 3.3 provides an overview of the MRMAC system, including video streaming, data exchange between the server and client, and managing and distributing direct object manipulation.

[4] https://readyplayer.me/
[5] https://www.reallusion.com/character-creator/headshot/

Figure 3.6: Personalized avatar generation process.

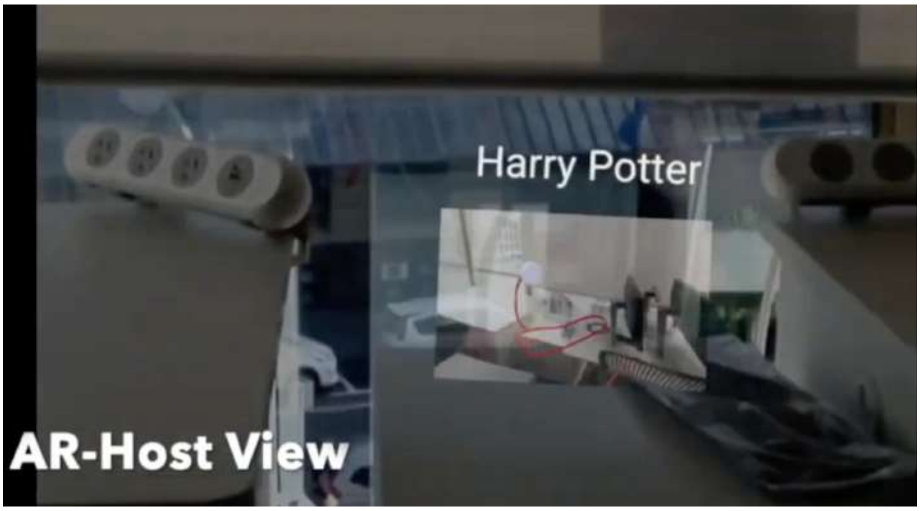


Figure 3.7: View sharing in MRMAC: AR-Hosts can see the viewpoint of VR-Travellers through a picture-in-picture (PiP) window.

Video and Audio Streaming

A 360° camera positioned at eye level within the physical space captures 4K video and utilizes OBS Studio for encoding the stitched video. For video streaming, a WebRTC SFU architecture was implemented using the Janus WebRTC plugin for Unity, supporting VP9 video and OPUS audio codecs, deployed on a local server.

To facilitate network traversal through NAT and firewalls, STUN and TURN servers were implemented. While direct peer-to-peer connections are the preferred method for server-to-Unity communication, certain network configurations may block these connections. STUN and TURN servers provide fallback options to establish connectivity when direct connections are not feasible.

Additionally, a relay service was implemented on the central signaling server using the WebSocket protocol to enable bidirectional communication between the web client and the Unity application. While this relay service ensures seamless communication between clients and the server, it may introduce additional latency and reduce communication quality. Therefore, establishing direct peer-to-peer connections is prioritized whenever possible, with the relay service utilized as a fallback option.

The AR-HMD onboard camera captures live video for first-person perspective, which is down-sampled to 704×396 resolution and transmitted to remote users at 15FPS using WebRTC. An identification tag is embedded in two 8x8 pixel regions in the top left corner of each image for synchronization. The displaced pixels and matching identification tag are stored in the data packet. Upon receiving a frame, the client reconstructs it using the included data before rendering.

The setup was tested over a LAN wired with a 1Gbps Cat6 Ethernet cable. Both VR users and AR hosts were connected to the same network, although the AR hosts connected via the Archer WiFi 6 Router (AX5400 Pro). The average data transfer rate across the network was measured at 16.45KBps (0.58SD). Of this, 1.54KBps (0.31SD) was sent by the local user, and 3.76KBps (0.16SD) was sent by each remote user. The remaining 11.20KBps (0.47SD) is for transmitting the low-resolution video stream from the remote user to the local user. For streaming the 4K video average bandwidth was around 4Mbps. Regarding audio, the Opus codec was employed to encode single-channel 44.1 kHz audio at a target rate of 128 kbps

Virtual Assets Streaming

Non-media data is exchanged between clients using a Node.js [6] server with Socket.io [7], allowing for real-time annotation, avatar movement, and other collaborative features. All associated metadata, such as visual annotation cues, gaze cues, 3D avatar position, rotation and pose, mapping and localization data, and tracking results of the currently active view window, are passed through this central media orchestration server. A REST API parses all requests from a client, including identifying information such as the client ID, the operation to be performed in the Unity scene, and any relevant parameters needed to act. The actions performed by the users using these interactive features are communicated to the Unity machine by sending GET/POST requests.

Synchronization

Synchronization techniques are employed by the system to maintain a consistent state across all connected devices, ensuring seamless collaboration among multiple clients. This involves creating a new pair of threads for each received client of a successful connection. All interactions between the client and the server during the collaboration process are handled by a main server thread.

Whenever a user launches a drawing annotation or pointers to any object, the system automatically connects to all other clients running on the same panel and synchronizes each change in the position of objects and avatars with all connected clients. This allows participants to see the avatar's movement and annotation, not just its final position. Absolute

[6] https://nodejs.org/
[7] https://socket.io/

position and users' motion tracking are performed in the client application, with tracking data applied to the client avatar and distributed to corresponding avatar copies in the server and other clients' applications.

The system employs a series of update messages sent between clients using WebRTC data channels to achieve synchronization. The server relays these messages, forwarding them to all other clients to ensure a seamless collaboration experience. To manage large amounts of data, peer-to-peer networking, time-stamping, and interpolation are utilized for synchronization. Additionally, audio and video synchronization for viewers is ensured through adaptive bitrate streaming.

3.5 Conclusion

This chapter introduced MRMAC, a mixed-reality multi-user collaboration system that allows multiple remote users to virtually teleport into a real-world task space to collaborate with local users. First, a design concept and protocol for a multi-user asymmetric remote collaboration system were presented, followed by implementation details of bidirectional face-to-face communication and synchronized audio-visual communication between multiple remote and local users. The seamless integration of 4k live-streaming video with mixed-reality technologies enables remote users to see the local users' physical environment. By leveraging high-definition video streaming, MRMAC provides remote users with an immersive and detailed view of the real-world task space, allowing them to closely examine objects, gestures, and subtle nuances within the environment. Apart from the real-time video streaming system, spatial audio, another integral component, provides an immersive auditory experience that accurately reflects the physical locations of sound sources.

In order to seamlessly represent multi-users, the system featured tools for managing multiple avatars' positioning and personalized avatars, enabling customized experiences within the MR space. Additionally, to facilitate interaction, integrated tools such as the 3D pointer and annotation drawing simplified object interaction and collaborative tasks.

Moving forward, the next chapter will focus on evaluating all these features. System performance analysis will be conducted to evaluate MRMAC's scalability, overall latency, and performance in collaborative scenarios, particularly in a multi-user context. Furthermore, user studies will be conducted to determine whether MRMAC can induce presence, facilitate collaborators with their tasks, and assess its overall effectiveness.

Chapter 4

Evaluation of MRMAC

This chapter conducts a thorough assessment of MRMAC's performance, conducting a detailed examination of each component within the system. Specifically, it evaluates the system's scalability by verifying its ability to support low-latency synchronized data streaming to facilitate multi-user telecollaboration. Then through user studies, the potential of MRMAC in inducing presence, enhancing task performance, understanding user preferences, facilitating collaborators with their tasks, and ultimately gauging its overall effectiveness will be discerned.

4.1 MRMAC Evaluation: System

This section provides a detailed assessment of the MRMAC system. Each component within the architecture is thoroughly examined to understand its performance characteristics. For this experience to truly provide a real sense of presence to the users, real-time operation is imperative. Achieving this involves maintaining a high frame rate for fluid movement and minimizing latency to provide immediate visual feedback, similar to real-world interactions. Furthermore, the application must seamlessly handle the data load from multiple users, ensuring scalability without sacrificing performance. Benchmarking has been conducted to assess the system's ability to deliver this immersive experience with optimal performance.

4.1.1 System Setup

MRMAC was implemented using the Unity game engine (version 2019.4.17f1) and ran on a machine with an Intel Xeon W-2133 3.60GHz CPU, 16GB of RAM, and a GeForce RTX 2080 Ti GPU. In the physical collaboration space, the AR-Hosts used AR HMDs (Microsoft

HoloLens 2) and a 360° camera (Ricoh Theta Z1), which was mounted approximately 1.7 meters above the floor level. The remote VR-Travelers used VR-HMDs like the VIVE Pro 2 or Meta Quest 2 to view the streamed environment. The system setup is illustrated in Figure 3.1.

The setup was tested over a LAN wired with a 1Gbps Cat6 Ethernet cable. Both VR users and AR hosts were connected to the same network, although the AR hosts connected via the Archer WiFi 6 Router (AX5400 Pro) wireless bandwidth capacity of 3 Gbps. The average data transfer rate across the network was measured at 16.45KBps (0.58SD). Of this, 1.54KBps (0.31SD) was sent by the local user, and 3.76KBps (0.16SD) was sent by each remote user. The remaining 11.20KBps (0.47SD) is for transmitting the low-resolution video stream from the remote user to the local user. For streaming the 4K video average bandwidth was around 4Mbps. Regarding audio, the Opus codec was employed to encode single-channel 44.1 kHz audio at a target rate of 128 kbps.

Testing MRMAC over a LAN with a 1Gbps Cat6 Ethernet cable provided a controlled environment for initial assessment. While the ultimate goal is to support collaboration among more remote users, starting with a smaller-scale setup allowed us to validate the system's core functionality and performance without introducing complexities associated with diverse network environments such as firewall issues, security concerns, and controlling test scenarios to mitigate bias. Amidst the challenges posed by the COVID-19 pandemic, recruiting participants for extensive remote testing proved challenging. By initially focusing on a controlled environment, the system's capabilities can efficiently assess and refine its performance before scaling up to accommodate larger numbers of remote users in varied network settings. This approach has enabled the system's readiness for broader deployment while effectively addressing potential security and logistical challenges.

Motion to photon latency was taken by tracking the time from when a user performs an action, such as moving their head or pressing a button, to when the resulting change is displayed in the VR/AR environment. Approximate latency was recorded when the 360 video was first viewable on the local computer and when viewed on the remote computer in Unity. Latency was measured by performing an action in front of the camera while measuring the time it took for the corresponding changes to appear in the video. The process was repeated ten times, and averages were calculated. Apart from that Unity profiler tools are also used to measure latency at different stages of the rendering pipeline or the interaction between the VR hardware and software components.

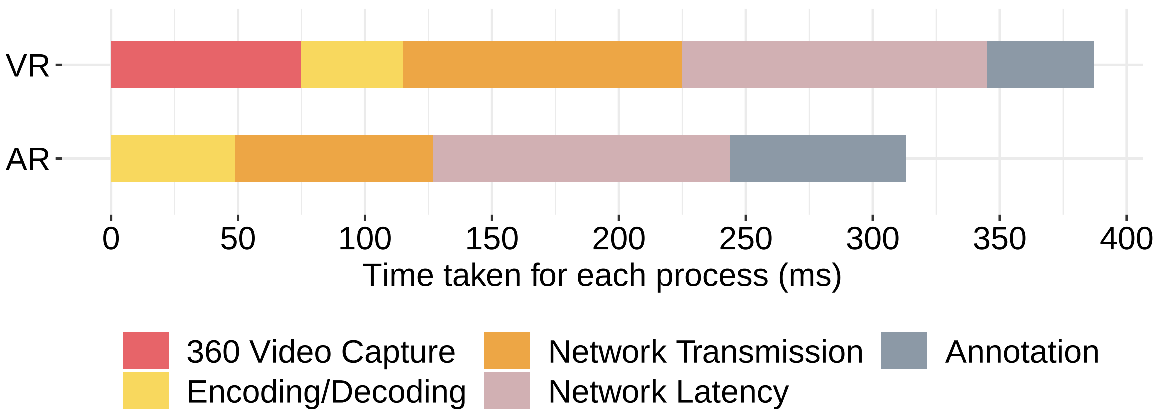


Figure 4.1: System performance metrics in milliseconds (ms) or frames per second (FPS) for each performance metric.

4.1.2 Results

The performance metrics for evaluating the MRMAC system encompass the time taken in milliseconds for 360° video capturing, 360° video processing, encoding/decoding, transmission, and network latency. A scenario was simulated where four users were operating the system (two local, two remote). Each measurement represents an average of ten samples. The results are presented in Figure 4.1.

Table 4.1: System Performance Measurements

360° Video Capture	75ms (σ = 3.11)
Encoding/Decoding	40ms (σ = 1.48)
Network Transmission	110ms (σ = 3.07)
Network Latency	120ms (σ = 2.86)
Total Latency	345ms

The 360° video capture includes video stitching, which takes up 21.7% of the total time. The encoding and decoding handle the video and audio, taking 11.6% of the total time. While the networking transmission and latency take the most time at 66.6% (see Table 4.1).

The average frame rate of MRMAC includes video streaming at 30 FPS, audio streaming at an average sampling rate of $44.1kHz$, and a rendering time of 60±10 FPS. The high frame rate and reduced latency significantly improved the synchronization between audio and visual elements. Processing the view-window sharing onto the display takes 73ms. Remote user actions such as pointing, drawing annotations, or gestures were visible to local users

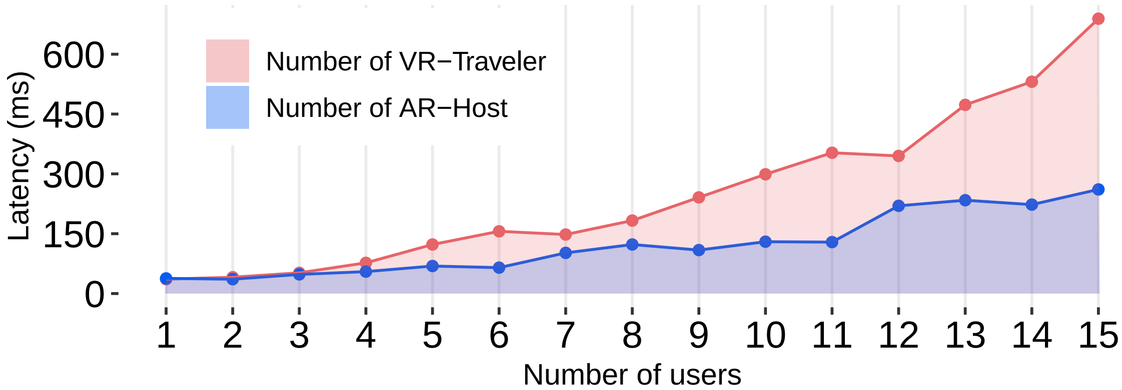


Figure 4.2: Scalability of MRMAC's network latency as the number of users increases.

with an average delay of 106ms.

To demonstrate the scalability of the system, MRMAC was tested by increasing the total number of remote users while maintaining one local user, and then by increasing the total number of local users while maintaining one remote user. The changes in network latency with respect to the number of users are shown in Figure 4.2. The results indicate that a slight linear increase was observed for both VR-Travelers and AR-Hosts. Slightly higher network latency was experienced by VR-Travelers compared to AR-Host, which is believed to be due to 360° video not being streamed to AR-Hosts. However, even with 15 users, the network latency remained under 1 second (650ms).

4.1.3 Discussion

MRMAC is designed to facilitate collaboration between users in mixed-reality environments, where participants are physically present in AR while others join remotely in VR. The collaborative setup aimed to bridge geographical distances and enable immersive, interactive experiences across different locations. The development of MRMAC was influenced by prior research and addressed the need for efficient and seamless communication among users situated in varied physical environments. The network protocols are designed to facilitate efficient video encoding and decoding with low bitrate requirements, enabling live synchronization. Additionally, the network architecture is optimized to prioritize real-time collaboration with minimal latency.

The performance analysis of MRMAC shows that it achieved a total latency of 345ms for a mixed-reality asymmetric collaboration involving four users (two local AR-Hosts and two remote VR-Travelers). The networking transmission and latency accounted for 66.6% of the time (230ms), while the remaining processing time, including video capturing and data encoding/decoding, was less than 115ms.

The overall performance can be further improved through software optimization and advanced network infrastructure with higher bandwidth. However, the prototype shows promising results given that human participants typically perceive conversations as synchronized when the latency is below 250ms [44]. Furthermore, MRMAC demonstrated strong scalability in accommodating up to 15 remote users with a single local user in a simulated experiment while maintaining network latency of less than 250ms with up to 8 remote users. Notably, the 360° video capturing and encoding/decoding times remained consistent across varying numbers of users. With high frame rate audio (44.1 KHz), video (30FPS), and mixed reality rendering (60FPS), MRMAC is well-suited for a wide range of applications with varying numbers of users and devices.

4.2 MRMAC Evaluation: User Role

The section focuses on evaluating MRMAC with respect to two distinct user roles: local and remote. The primary objective is to explore whether the inherent asymmetry in these roles influences the overall user experience. Furthermore, the aim is to find out how users' presence and task performance vary when engaging in collaborative activities within MRMAC while assuming these different roles. By thoroughly examining these dynamics, the goal is to gain deeper insights into how MRMAC functions within collaborative settings and how users' roles may impact their interactions and perceptions.

The evaluation of MRMAC extends to the examination of its performance within distinct user roles: local and remote. This evaluation aims to understand how the inherent asymmetry in these roles influences the sense of presence and overall user experience. As MRMAC allows VR-Travelers to to be in the same place as the AR-Host with full situational awareness, evaluating users' sense of presence became crucial (Figure 4.3). A comprehensive user study was conducted to explore both the social and spatial presence experienced by VR-Travelers and the AR-Host.

Participants in the study also engaged with various metrics and questions aimed at capturing and understanding key aspects of user experience. This included assessments of usability, workload, comfort, and subjective preferences. The insights gathered from this study are invaluable for informing the ongoing design and optimization of the MRMAC system, ensuring that it effectively meets the needs and expectations of its users in mixed reality environments.

The subsequent sections drive into the intricacies of the study's design and hypotheses, the experimental procedure employed, the methods of measurement utilized, the findings obtained, and an in-depth discussion of the results.

4.2.1 Design and Hypotheses

A within-subjects user study was designed where participants solve a collaborative puzzle-building task as VR-Traveler and AR-Host (see Figure 4.3). This study aims to address key research questions (**RQs**) concerning the effectiveness and user experience of collaborative systems in such environments. Specifically, the investigation centers around research question **RQ1**, which focuses on how multiple remote users achieve a high sense of presence while maintaining spatial awareness and understanding in the MRC space, and **RQ2**, which examines how multiple remote VR users can be effectively represented in the MRC

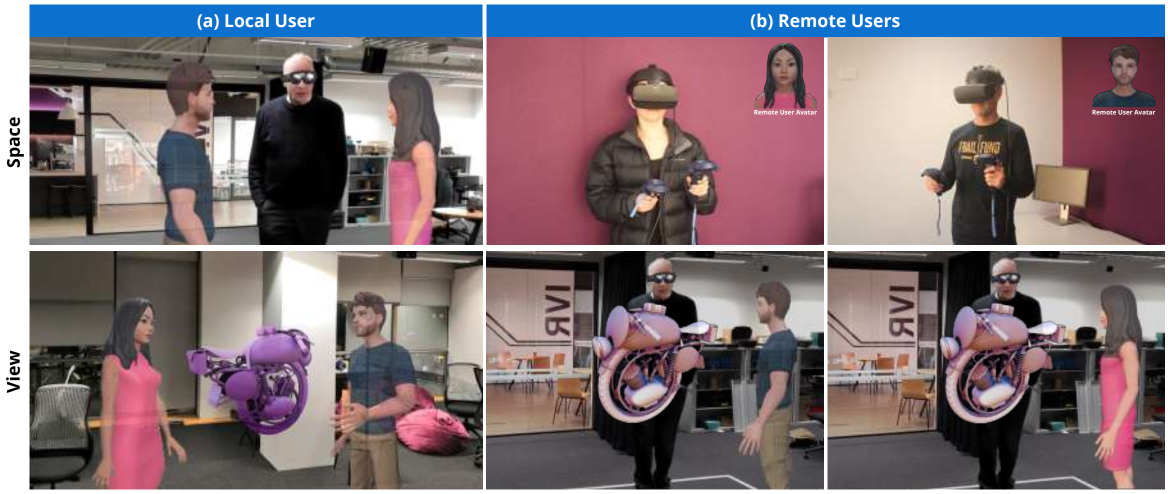


Figure 4.3: MRMAC Scenario: At a physical work site, local AR user (a) Adam (host) initiates a collaboration session using MRMAC in which remote VR users (David and Jane) (b, c) collaborate with Adam. MRMAC broadcasts the captured real-world environment to multiple remote users. Teleporters' identities (represented as avatars) share audio-visual cues to augment their remote collaboration.

space to achieve co-presence (see Section 1.3). To explore these questions, hypotheses were formulated focusing on evaluating the spatial and social presence experienced by users, as well as their perceptions of the effectiveness and preference of the MR Collaboration system. It's important to note that the study was designed to evaluate these aspects specifically in the context of different user roles, ensuring that both VR-Travelers and AR-Hosts experience similar levels of presence and engagement in the collaborative space.

The hypotheses are as follows:

H1 *Spatial presence will be equally high for both VR-Travelers and AR-Hosts* — In MRMAC, VR-Travelers perceive the AR-Hosts' space in detail, and their avatars blending into the AR-Hosts' space will lead to high spatial presence, regardless of their respective roles.

H2 *Social presence will be equally high for both VR-Travelers and AR-Hosts* — MRMAC shares a live stream of the AR-Hosts' space, where both parties engage with each other as if they were physically co-located. This shared experience, facilitated by consistent audiovisual and interaction cues, is expected to create a sense of presence and social connection that is comparable for both VR-Travelers and AR-Hosts.

H3 *Both VR-Travelers and AR-Hosts will find MRMAC effective to use* — MRMAC provides users with the necessary communication and interaction cues, ensuring seamless collaboration and facilitating their respective tasks within the shared collaborative environment.

H4 *Both VR-Travelers and AR-Hosts will equally prefer MRMAC* — MRMAC offers a balanced combination of features and functionalities tailored to users' respective needs, reducing workload and the likelihood of errors to improve task performance.

4.2.2 Study Tasks

For the user study, two tasks (training and main task) were designed with different procedures but with the same collaboration characteristics.

A **training task** was created for participants to familiarize themselves with the user interface and functionality of the system. In the training task, users perform simple activities including the selection and placement of virtual objects in an MRC space with remote partners (Figure 4.4 B). User interfaces for both VR-Travelers and AR-Host include a menu for selecting virtual objects (e.g., cube, sphere, etc.), pointers, and a color palette (Figure 4.4 A). This session's menu and settings were similar to the main task but with fewer options. No data was collected during the training task.

For the **main task**, a remote collaboration task was designed in which the AR-Host makes 3D structures using the physical LEGO and Domino blocks with the help of the

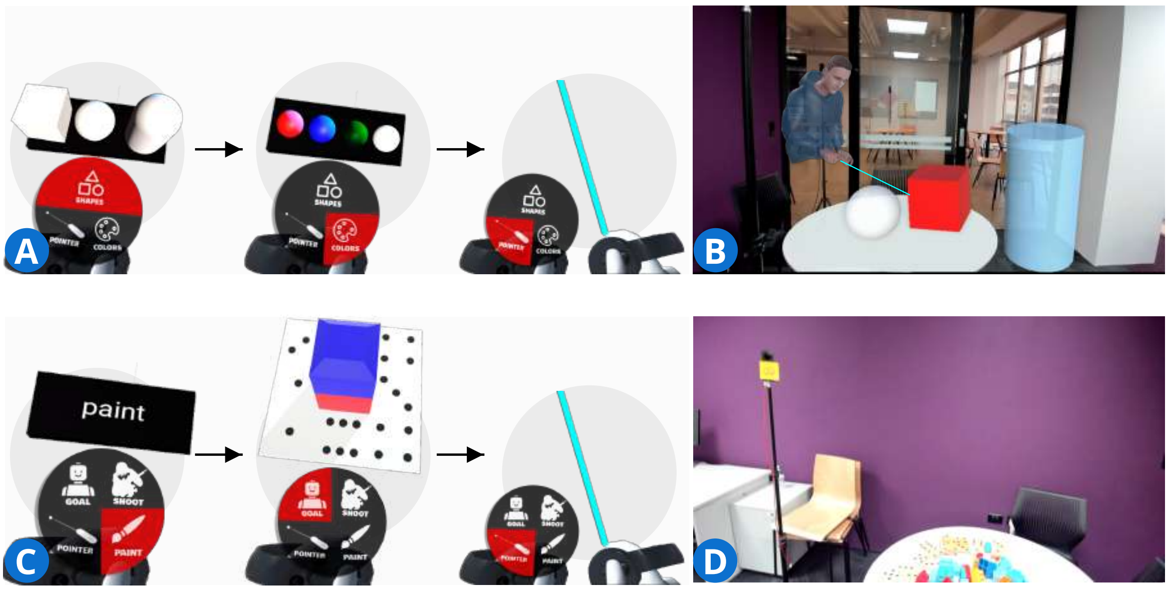
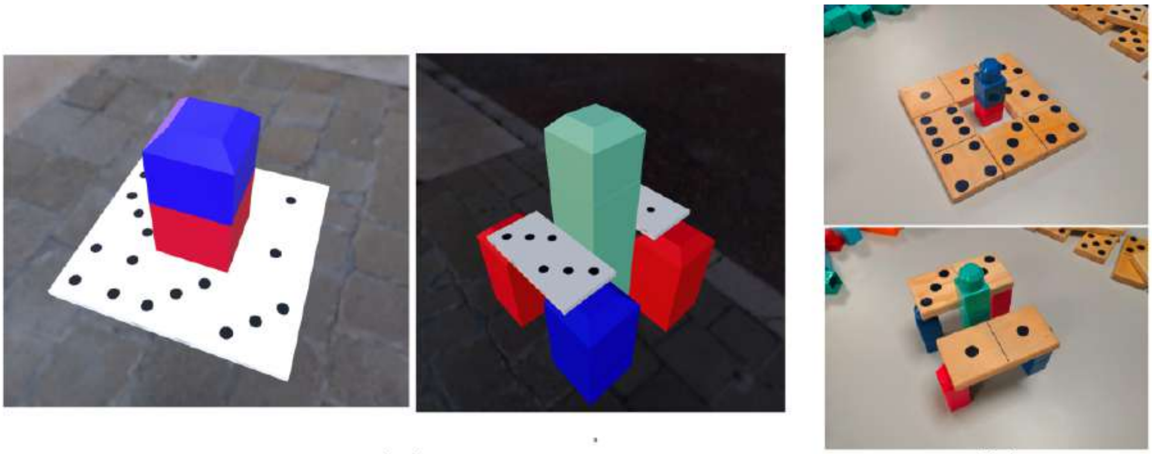


Figure 4.4: User interface and MRC space in the *Training task* (A and B); user interface and MRC space in the *Main task* (C and D).



(a) (b)

Figure 4.5: Lego and domino blocks structures: (a) virtual replicas, (b) corresponding structures made using physical blocks.

VR-Travelers. One of the goals is to encourage the AR-Host and VR-Travelers to naturally communicate and discuss. Various distinct 3D structures were designed using LEGO and Domino blocks, along with their virtual replicas, as shown in Figure 4.5. The LEGO and Domino structures were intentionally kept relatively short to ensure a brief building time in the MRC experiment. This decision aligns with the research objectives, as the primary focus of the experiment is to evaluate communication and collaboration effectiveness rather than complex construction skills. By maintaining a short building time, the emphasis is on assessing participants' ability to work together efficiently within a limited timeframe. Additionally, keeping the tasks short helps to minimize participant fatigue, which could otherwise affect their performance and engagement levels. Therefore, to maintain consistent focus and motivation throughout the experiment, the LEGO and Domino structures were kept relatively simple. Importantly, all structures were designed to be random to prevent participants from relying on memory to complete the task, ensuring that the assessment is based on real-time collaboration and communication skills. In the physical workspace for the AR-Host, LEGO and Domino blocks were randomly placed (Figure 4.4D). The AR-Host follows the instructions of the remote VR-Travelers (voice description, gestures, pointing, and/or drawing annotations) to locate, collect, and assemble designated LEGO and Domino bricks in the correct order.

VR-Travelers can see the 3D virtual replica of the final model on their VR display, which the local user will assemble at their workspace using the physical bricks (Figure 4.4C). VR-Travelers use the 3D pointer, annotation, and verbal instructions to direct the AR-Host. The task is considered complete when VR-Travelers agree that the AR-Host has successfully assembled the desired structure, which matches or resembles the virtual replica displayed on the VR-Travelers' interface.

4.2.3 System Setup

For the user study, a functional prototype of MRMAC was developed using the Unity game engine (version 2019.4.17f1). The study utilized two distinct rooms with different setups: one designated as the Mixed Reality Collaboration (MRC) space for the AR-Host and another as the remote user's space for the VR-Travelers.

In the AR-Host's room, a 360-degree camera (Ricoh Theta Z1) was mounted approximately 1.7 meters above the floor level. This camera was connected to a computer powered by an Intel Xeon W-2133 3.60GHz CPU, equipped with 16 GB of RAM, and a GeForce RTX 2080 Ti GPU.

The AR-Host used an Augmented Reality Head-Mounted Display (HMD), such as the Microsoft HoloLens or Magic Leap, while the VR-Travelers used a Virtual Reality HMD, such as the HTC Vive or Oculus Rift S. The AR-Host and VR-Travelers interact and collaborate within the shared virtual space to perform collaborative task.

4.2.4 Participants

A convenience sample of 24 participants was recruited, comprising 21 males and 3 females (M_{age} = 30.42, SD_{age} = 11.18), through word of mouth, email, and flyers on a university campus. Participants were grouped into triads, with two randomly assigned the role of VR-Traveler and one the role of AR-Host. Most participants (16, 66.66%) did not have any prior experience with AR/VR. Two participants reported using it one or more times per week, two reported using it one or more times per month, and four reported using it one or more times per year. All participants had prior experience using video conferencing software (e.g., Zoom, Skype, Teams). None of the participants had previously used our system, and all reported normal or corrected-to-normal vision and hearing.

4.2.5 Procedure

The overall procedure contains three stages (i.e., preparation, experiment, and interview). First, participants were briefed about the experiment, system features, and the task, followed by a demographic questionnaire. Then participants were randomly grouped into triads where two participants took the role of the VR-Travelers, and the other took the role

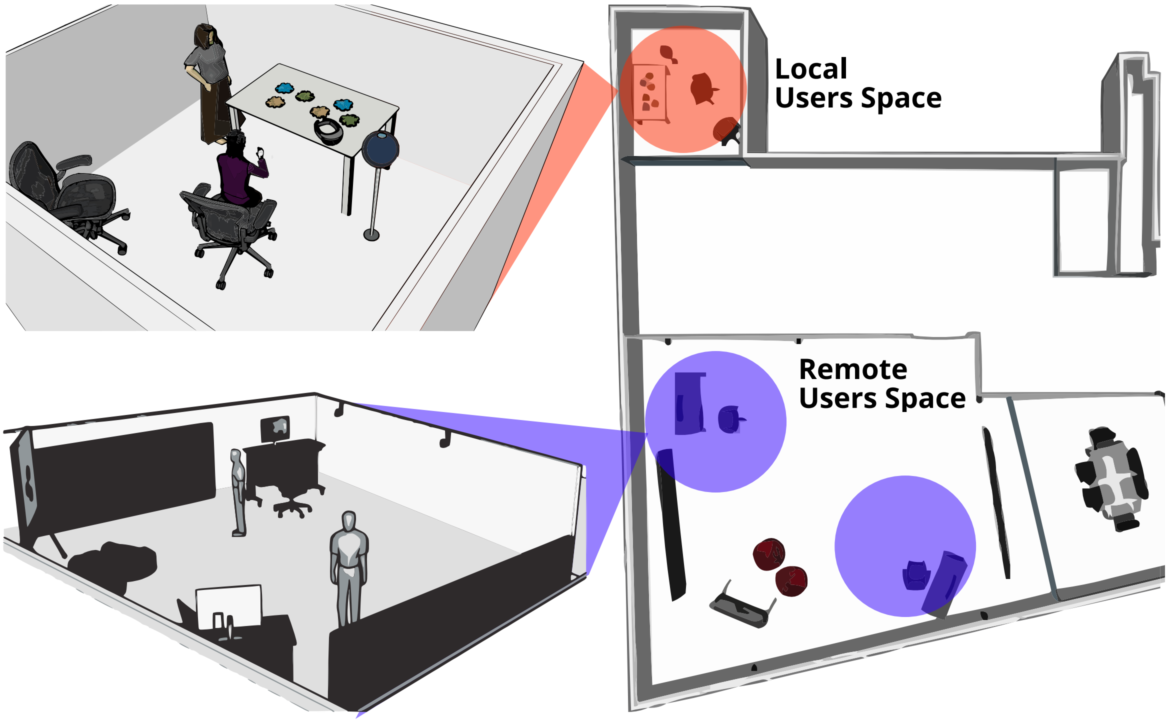


Figure 4.6: Room layout and placement of the VR-Travelers and AR-Host.

of the AR-Host. Participants who took a VR-Traveller role were placed separately in one room, while the AR-Host was placed in another room (Figure 4.6). To familiarize themselves with the task and system, each participant was asked to perform the training task where they could interact freely with the virtual objects in the MRC space together with their remote partners (~ 10 mins).

Then the participants started the main task. The task was completed when the AR-Host put all the pieces in the right places to build the 3D structure with the guidance of the VR-Travelers (Figure 4.5). The total time taken to complete the task was logged. Participants then switched roles, and the procedure was repeated two more times to guarantee that all three members of that group played each role. Between each session, participants were asked to complete the spatial presence, social presence, SUS, and NASA-TLX questionnaires. These are intermediate questionnaires that are completed immediately after each task. Following the completion of the three sessions, participants were asked to complete a post-experiment questionnaire on a 7-point scale and provide additional feedback (if any) and interviewed.

Participants also completed a Simulator Sickness Questionnaire (SSQ) both before and after the experiment. All questionnaires except the SSQ and SUS used a seven-point Likert

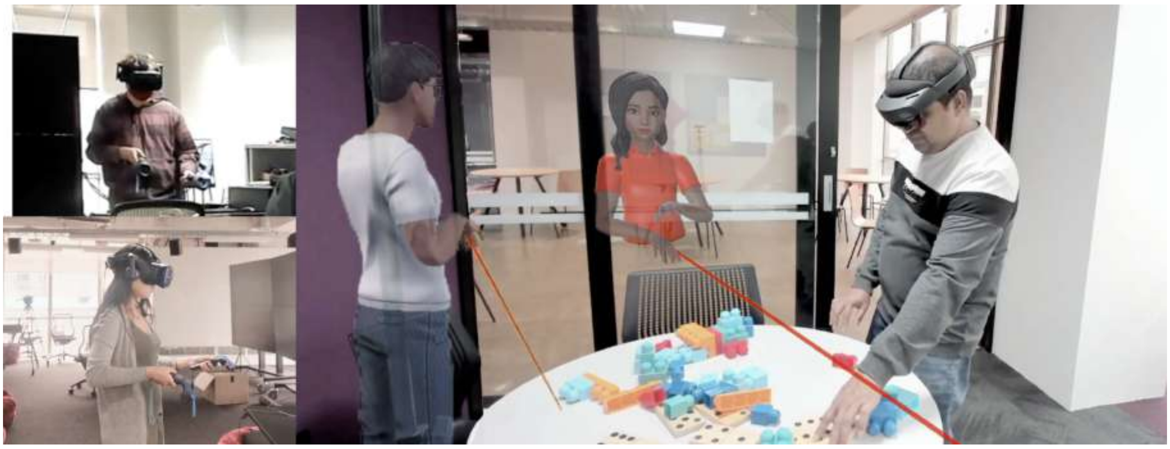


Figure 4.7: User study setup for MRMAC prototype. Left: multiple remote users are exploring the local user environment. Right: the local user in front of the table, identifying the specific LEGO pieces that the remote user instructs to select. The avatar of the remote user positioned within the shared space, guiding the local user using a 3D pointer.

scale, while the SUS used a five-point Likert scale and the SSQ used a zero-to-four-point scale. Overall, the process took about an hour to complete on average. A scene captured during the user study is depicted in Figure 4.7, showcasing the setup and interaction between the AR-Host and VR-Travelers within the MRC space.

4.2.6 Measures

The following measures were taken to analyze the user presence and effectiveness of collaboration of the MRMAC system.

Spatial presence, is a measure of the feeling of being in a certain environment. This measure is particularly significant when examining the experience of VR-Travelers teleporting to an on-site location using the system. To assess this measure, the 14-item *iGroup Presence Questionnaire* (IPQ) [204], which includes a sub-scale specifically focused on spatial presence, was employed.

Social Presence, is a measure of the feeling of being with others, where others are perceived as real people. This questionnaire is compiled based on research by Biocca et al. [28], Bailenson [14], and Hauber [88]. However, to minimize the participant burden during the experiment and considering the collaborative task's nature, only questions from the subscales of co-presence (Bail [14] and Hauber [88]), mutual attention, mutual understanding, and behavioral engagement (Biocca et al. [28]) are included.

Usability, a measure indicating the effectiveness, efficiency, and satisfaction with which the system can be used, was quantified using the 10-item System Usability Scale (SUS) [35] on a Likert scale of 1-5, where 1 represents “Strongly Disagree” and 5 represents “Strongly Agree”.

Task performance, is the measure of task success, logged completion times, and error rate. Subjective workload was evaluated by having participants complete the NASA Task Load Index (TLX) after the task. This includes a set of six rating items within a 100-point range with 5-point steps (0: very low~100: very high, the lower, the better).

Preference was assessed by asking participants to indicate their preferred role during the study via a post-experiment questionnaire on a 7-point Likert scale, aiming to acquire a better understanding of the user experience.

Interviews, provides qualitative insights into participants’ collaborative experiences. Participants were asked open-ended questions about various aspects of their experience, specifically elaborating on some questions asked in the preference questionnaires. Topics covered during the interviews included spatial awareness, communication effectiveness, task coordination, audio-visual quality, avatar representation, and blended virtual content.

4.2.7 Results

All groups successfully completed the remote collaborative tasks. Statistical analysis was conducted to assess differences among groups. Firstly, the assumption of normality and homogeneity of variances was checked using the Shapiro-Wilk test (α = 0.05). Based on the results, either a one-way ANOVA or the Wilcoxon Signed-Rank test (WSR) (α = 0.05) was employed and reported F or $W - values$. In all cases, post hoc tests were performed using a Bonferroni correction to account for multiple comparisons.

Spatial Presence

Overall both VR-Travelers and AR-Host had a high sense of spatial presence, with mean scores above the midpoint of the scale. For AR-Host, the mean spatial presence score was 5.30 (SD = 1.32), while for VR-Travelers, it was 5.27 (SD = 1.29), as shown in Figure 4.8. No significant differences were observed between roles in General Presence (W = 47, p = 0.292), with mean scores for AR-Host and VR-Travelers being 5.00 (SD = 1.07) and 5.50 (SD = 1.10), respectively. Similarly, no significant differences were found between roles in Involvement (W = 250, p = 0.901), Realism (W = 295, p = 0.385), or Spatial Presence (W = 1655.5, p = 0.753).

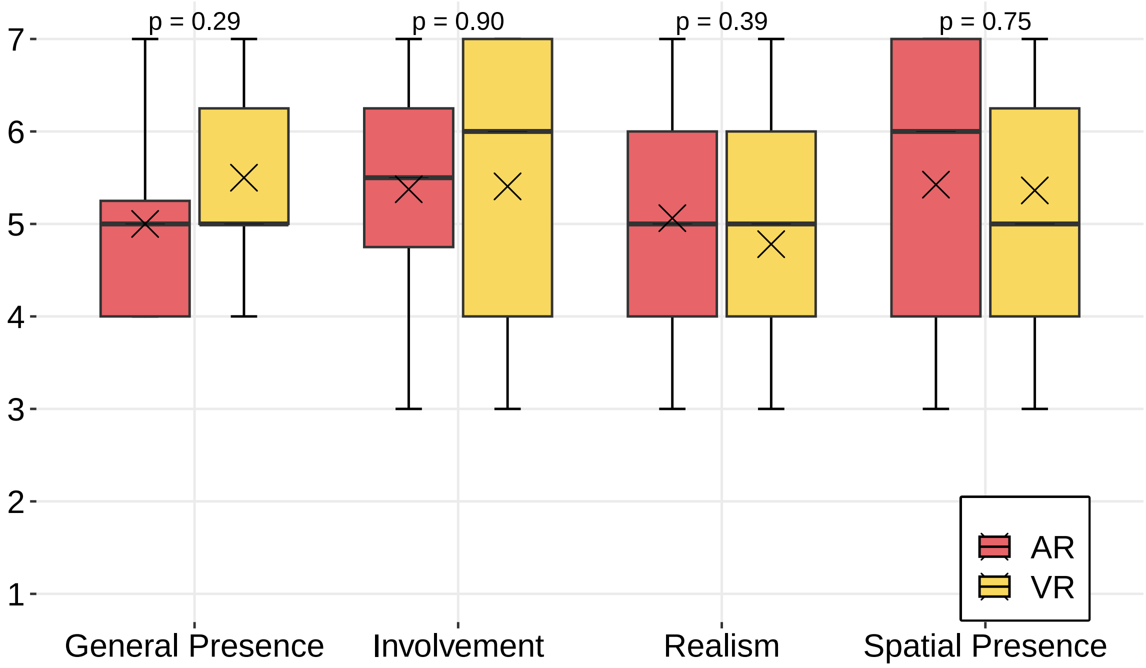


Figure 4.8: Participants' spatial presence scores comparison between AR and VR users.

Social Presence

The results from the social presence questionnaire suggest that both VR-Travelers and AR-Hosts experienced high levels of social presence within the MRC space (Figure 4.9). The mean scores for both groups were above the midpoint. Specifically, the mean score for AR-Hosts was 5.26 (SD = 1.410), while the mean score for VR-Travelers was 5.09 (SD = 1.44). WSR test indicated no significant difference in Behavioral Engagement between AR and VR participants ($W = 270$, $p = 0.760$). The mean score for AR participants was 5.56 (SD = 1.31), and for VR participants was 5.47 (SD = 1.22). Similarly, no significant difference in Co-presence was found between AR and VR participants ($W = 1158$, p = 0.285). The mean score for AR participants was 5.66 (SD = 1.23), and for VR participants was 5.34 (SD = 1.38). For Mutual Attention, there was no significant difference observed between AR and VR participants ($W = 492$, $p = 0.309$). The mean score for AR participants was 4.67 (SD = 1.46), and for VR participants was 5.04 (SD = 1.41). Although there was no significant difference in Mutual Understanding between AR and VR participants ($W = 328.5$, $p = 0.106$), it's noteworthy that the mean score for AR participants was 5.06 (SD = 1.53), while for VR participants, it was 4.31 (SD = 1.57).

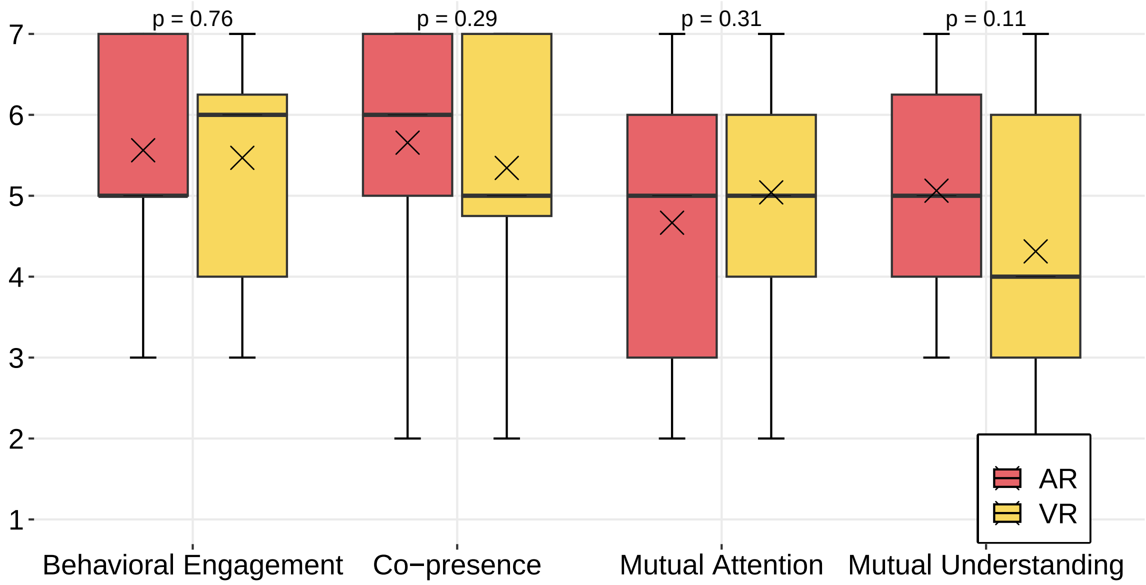


Figure 4.9: Social presence between AR and VR users.

Usability

Both VR-Traveler and AR-Host participants reported relatively high System Usability Scale (SUS) scores, exceeding the standard average SUS score of 68. Specifically, AR-Host participants rated the system usability with a mean score of 72.81 (SD = 14.90), while VR-Traveler participants rated it slightly lower with a mean score of 72.03 (SD = 16.13), as shown in Figure 4.10. Findings suggest that both VR-Traveler and AR-Host experiences were perceived as good according to the interpretation guidelines for SUS scores. The WSR test indicated no significant difference in SUS scores between VR-Traveler and AR-Host ($W = 66.5$, $p = 0.902$).

Additional usability questions (Figure 4.11) where questions about positive aspects (Q1) scored $\mu = 3.83$ ($\sigma = 1.00$) for VR-Traveler and $\mu = 4.11$ ($\sigma = 0.85$) for AR-Host, and (Q2) scored $\mu = 4.04$ ($\sigma = 0.69$) for VR-Traveler and $\mu = 3.66$ ($\sigma = 0.81$) for AR-Host. Similarly, on negative aspects (Q3) and (Q4), the ratings were $\mu = 2.20$ ($\sigma = 1.14$) and $\mu = 2.08$ ($\sigma = 1.01$) for VR-Traveler and $\mu = 2.16$ ($\sigma = 1.49$) and $\mu = 2.29$ ($\sigma = 1.12$) for AR-Host.

Task Performance

Completion time was analyzed using a One-Way ANOVA to verify the balanced complexity of the test across three LEGOs and Dominos structures. A Bartlett's test suggested that the homogeneity assumption of variance was met (Bartlett's K-squared = 0.901, $p = 0.637$). The ANOVA result showed that task complexity is well-balanced and similar ($F(2, 21) = 0.698$,

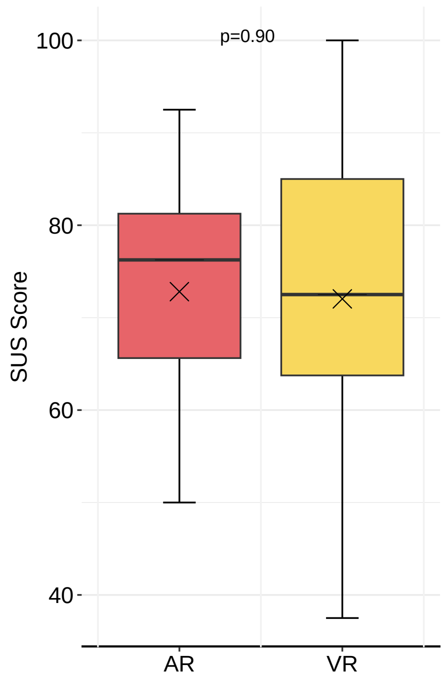


Figure 4.10: Participants' SUS scores comparison between AR and VR users.

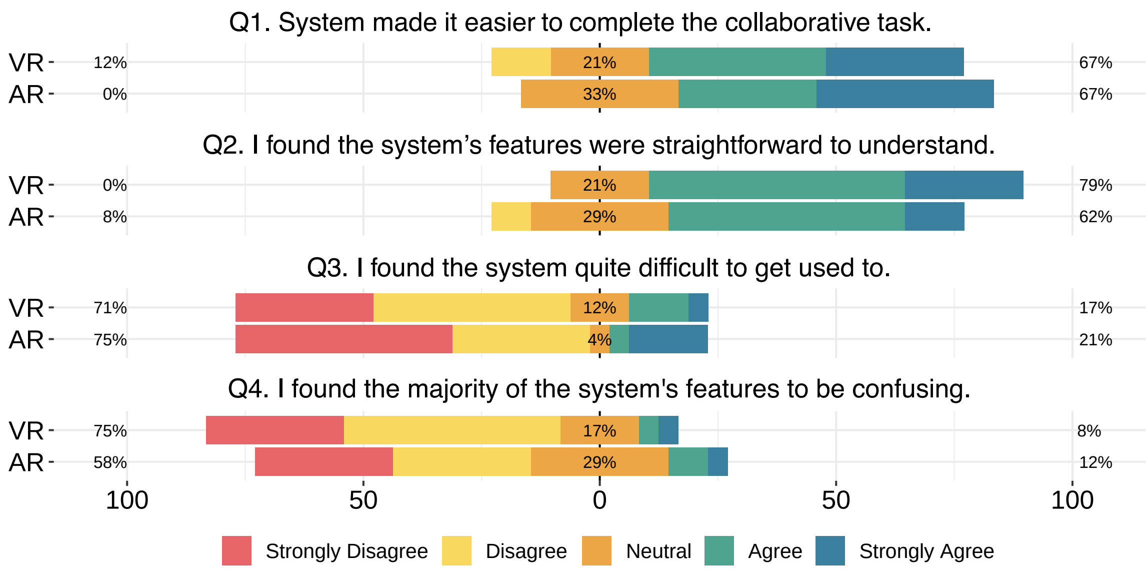


Figure 4.11: Additional system usability questionnaire (Q1 and Q2 are positive aspects while Q3 and Q4 are negative aspects).

Table 4.2: Task completion times (in seconds) and error frequency, each with standard deviation in parentheses.

	Completion time	**Error Frequency**		
	μ (σ)	Color	Shape	Final Output
Session	413.29	1.25	1.20	1.38
	(56.11)	*(0.43)*	*(0.40)*	*(0.86)*

p = .508), indicating that the test was valid across models without a substantial impact on task performance. Table 4.2 shows the completion time for each session. After each task completion error frequency was assessed where errors were defined based on discrepancies between the assembled model and predefined criteria encompassing color, shape, and final output (see Table 4.2).

- **Color Error:** when the colors of the assembled model do not match the intended colors. For example, if the final output has a red block but the participant uses a blue one instead.
- **Shape Error:** involves discrepancies in the shapes of the assembled components compared to the intended shapes. For instance, if the final output require a square block to be used but the participant uses a rectangular one.
- **Final Output Error:** refers to any discrepancies between the assembled model and the intended final output. For example, if the final output should resemble a specific structure or pattern but the participant's assembly deviates significantly from it.

There were fewer errors throughout the session μ = 0.92 (σ = 0.65), with the majority of errors resulting from incorrect orientations and picking up the wrong pieces.

NASA-TLX provided a measure of difficulty in completing tasks with each role. A simplified version of NASA-TLX was used, where the averaged or summed raw subscale ratings (similar to Raw-TLX) were calculated by assigning higher weights to dimensions more relevant to our study. The weighting scheme for the assessment of workload factors in the study was as follows: mental demand (0.3), physical demand (0.2), temporal demand (0.2), performance (0.1), effort (0.1), and frustration (0.1). These weights were assigned to each factor to reflect their relative contributions and ensured that the combined influence of all factors accurately represented the overall workload experience. Results indicated that the overall NASA-TLX score was relatively low (M = 34.75, SD = 9.10), with VR users having

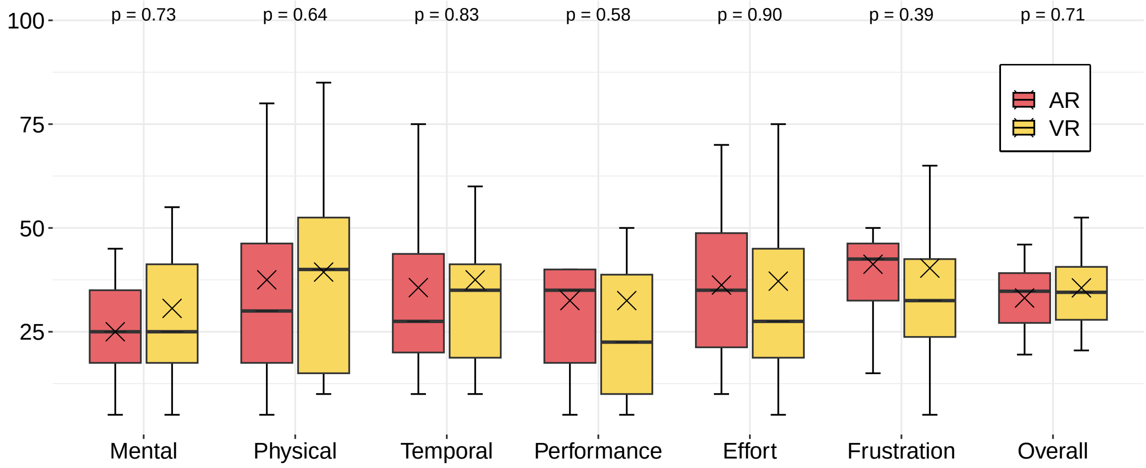


Figure 4.12: Participants' responses on the NASA-TLX questionnaire ($0 : very low$ to $100 : very high$, lower is better).

a higher mean score (M = 35.56, SD = 9.38) compared to AR users (M = 33.12, SD = 8.91) (Figure 4.12). However, this difference was not statistically significant ($p = 0.713$). Which suggest that, on average, participants rated the overall difficulty of completing tasks as moderate. Results also revealed no statistically significant differences in workload scores between roles across any of the dimensions tested (all $p > .05$). Therefore, it can be concluded that there is insufficient evidence to support the presence of differences in workload between roles within the AR and VR conditions.

Simulator Sickness

Figure 4.13 shows the results from the SSQ questionnaire [108], with 16 items rated from 0: none - 3: severe, then calculated the three subscales (nausea, oculomotor, and disorientation) and the total score. The SSQ was administered pre-experiment and post-experiment for each task in each condition. For the AR condition, there was no statistically significant difference in SSQ scores between pre-test (M = 3.27, SD = 2.39) and post-test (M = 2.80, SD = 1.73) conditions, with a WSR test yielding W = 100.5, $p = 0.277$. Similarly, for the VR condition, no significant difference was observed in SSQ scores between pre-test (M = 3.74, SD = 3.05) and post-test (M = 2.57, SD = 2.96) conditions, as indicated by a WSR test yielding W = 29, $p = 0.746$. Findings suggest that there were no significant changes in simulator sickness symptoms for either the AR or VR conditions following exposure to the respective conditions.

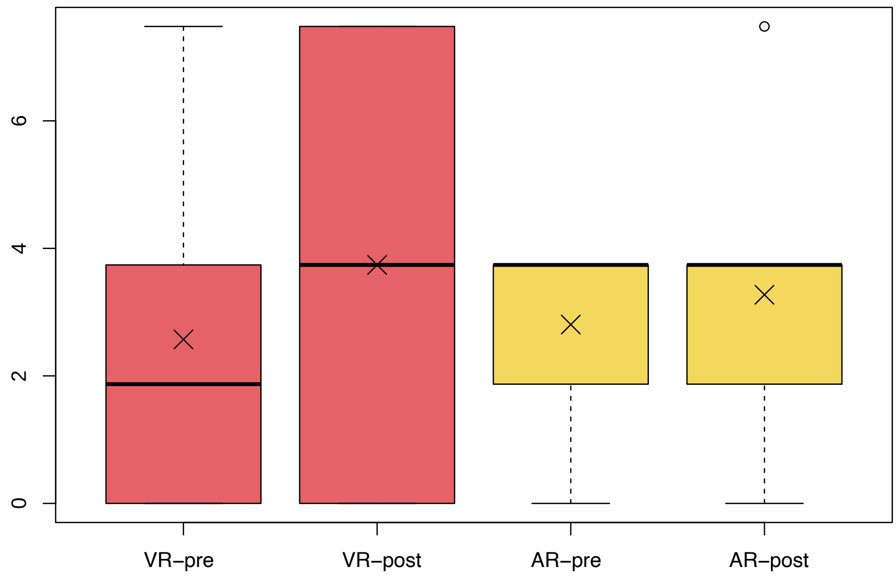


Figure 4.13: Increase in the simulator sickness score (pre- and post-exposure).

User Preferences

Overall preference scores (Figure 4.14) were high for both the VR-Traveler and AR-Host, as the majority of responses were positive. All participants reported strong overall audio quality and had no trouble understanding voice communications (Q3) (VR-Traveler: μ = 6.08 and σ = 0.88; AR-Host: μ = 5.04 and σ = 1.12). Video quality scores were overall high (Q2) (VR-Traveler: μ = 5.38 and σ = 1.35; AR-Host: μ = 5.63 and σ = 1.41). Nearly all participants felt present in the remote environment (Q8) and found it easy to collaborate (Q9). All participants felt the tools were sufficient to perform the overall task (Q7) μ = 4.67 and σ = 1.79. They reported that it was easy to see and point at things on the desk, which was very helpful. All participants reported that MRMAC was comfortable to use (Q10) (VR-Traveler: μ = 5.38 and σ = 1.35; AR-Host: μ = 5.29 and σ = 1.30).

4.2.8 Interview Insights

To gain insights into the dynamics of collaboration, participants provided valuable insights during the short interviews, highlighting various facets of their experiences. These facets included aspects such as spatial awareness, communication effectiveness, task coordina-

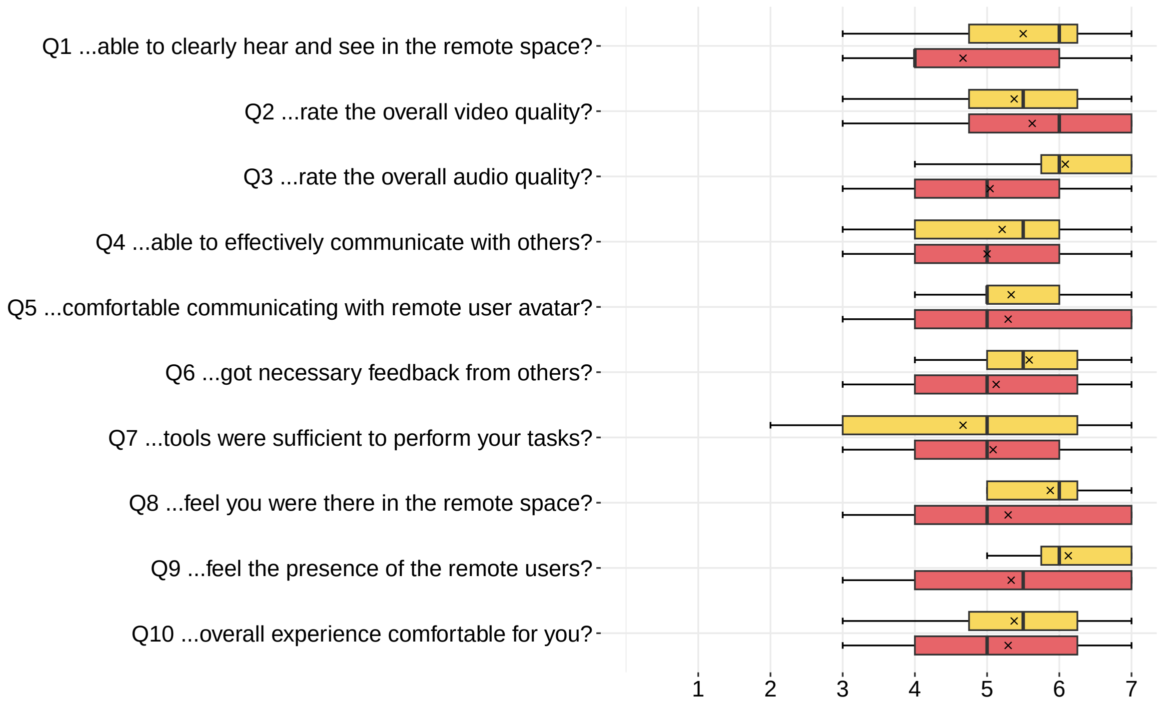


Figure 4.14: Statistical results of user preference questionnaire.

tion, sense of presence, and overall satisfaction.

Spatial Awareness: Participants expressed a heightened sense of spatial awareness. Participants remarked,

> *"[...] I felt like I could truly sense the space around me. It was like I was actually there."* [G3/P8]

> *"I felt more aware of my surroundings in the virtual world. It was like I could visualize the space around me better."* [G4/P12]

This suggests that the immersive nature of the live video stream of 360° in MRC space contributed to participants' heightened spatial perception and awareness. Participants also found it easy to locate objects in the remote space within the MRC environment. Another participant expressed:

> *"I could quickly identify and locate objects, even in the distant parts of the virtual space"* [G1/P3]

.

This highlights the effectiveness of spatial cues and design elements in facilitating object recognition and navigation in the MRC space. When asked about their motivation behind their preference scores on the effectiveness of communication with others during the tasks, they stated that they benefited from receiving feedback from others during the activity. Participants stated:

> *"[...] The feedback I received from others was useful in guiding me to find LEGO pieces and make necessary adjustments."* [G2/P5]

> *"I definitely felt the presence of the other user throughout our interaction."* [G2/P6]

Participants also enjoyed interacting with user avatars during the experience. This suggests that the realism and responsiveness of the user avatars enhanced the social interaction experience, making it more enjoyable. Participant commented,

> *"Talking with the user avatars was surprisingly fun and engaging. It felt like having real conversations."* [G5/P14]

Audio-Visual Quality and Synchronization: Participants provided feedback on the audio-visual quality and synchronization of the MRC space. Some of them were positively impressed by the audio-video quality of the virtual environment. One participant remarked:

> *"The audio-video quality exceeded my expectations. The sound was quite good, and the visuals were really nice and clear."* [G4/P10]

This indicates that the high-quality audio and visuals enhanced the immersive experience, contributing to participant satisfaction and engagement. Another participant noted:

> *"I had no trouble at all with hearing or seeing in the remote space. The sound was crisp, and the visuals were sharp and clear."* [G1/P1]

Participants also mentioned positive experiences with video and audio communication with teammates. This highlights the effectiveness of the virtual environment's technology in delivering high-quality audio-visual experiences, enhancing participant comfort and engagement.

> *"The video quality was clear, and I could hear my teammates well."* [G8/P23]

This suggests that clear video and audio communication contributed to effective teamwork and collaboration within the MRC space. Seamless synchronization between video and audio in the virtual interaction was another aspect that contributed to:

> *"I didn't notice any lag between the video and audio, which really made it feel like I was in the other room with them."* [G1/P2]

This indicates that low-latency audio-video synchronization enhanced the sense of presence and realism during the interaction.

Communication Effectiveness: Participants were able to efficiently locate and identify objects even in distant areas of the virtual space. This indicates the effectiveness of the virtual environment's design in facilitating object recognition and navigation, enhancing user experience and efficiency.

> *"I could quickly identify and locate objects, even in the distant parts of the virtual space."* [G1/P3]

Participants found communication to be more intuitive due to natural gestures, suggesting that the integration of natural gestures enhanced the realism and ease of communication within the MRC space, improving interaction quality. Another participant mentioned:

> *"I could point to things and wave naturally, which made communication feel more intuitive."* [G2/P4]

Participants experienced smooth and hassle-free communication with others. They highlighted the user-friendly interface and functionality of the communication software, enabling effective and enjoyable interactions without technical difficulties. A participant expressed:

> *"Communication with others was sweet as. The software made it easy as to have a yarn, and I didn't run into any hassles getting my message across."*[G7/P20]

Participants preferred the annotation feature and mentioned that the interactive drawing feature enhanced the enjoyment and engagement during the virtual activity. They also found it easy to communicate through pointing gestures in the MRC space. One participant expressed:

> *"Drawing was really fun in the virtual space. It was engaging and intuitive."* [G3/P6]

> *"Pointing was so easy to use for communication. It conveyed my message without needing to explain too much."* [G5/P11]

This highlights the effectiveness of non-verbal communication methods such as pointing and annotation, which facilitated clear and efficient message delivery during interactions. Participants mentioned that using pointing gestures for communication was both convenient and effective, highlighting the usability and communicative value of these intuitive gestures. This streamlined interaction processes and enhanced overall communication within the virtual environment.

> *"Using pointing gestures was straightforward and effective. It made communication simple and direct."* [G2/P9]

Challenges: Participants reported some challenges they faced during collaboration. Some participants experienced discomfort with the VR headset. One participant described the issue, saying,

> *"[...] headset became uncomfortable after a while, especially around the eyes and head."* [G5/P15]

Fitting the VR headset comfortably for long periods was difficult for some participants. Another participant shared,

> *"[...] just couldn't get the headset to make sit comfortably on my face without leaving dents in my face, and it felt heavier on one side."* [G8/P24]

This feedback underscores the importance of ergonomic design in VR hardware. Issues with fit and weight distribution can lead to discomfort and may limit the duration of use, impacting user engagement and overall experience.

Additionally, participants faced difficulties perceiving objects or teammates outside their field of view.

> *"I had trouble seeing things or my teammates when they were outside my field of view in VR. Sometimes, I missed what they were drawing."* [G6/P16]

These challenges can be attributed to the limitations of the current HMDs, which are hardware issues that cannot be entirely avoided with current technology. However, several precautions were taken, such as keeping tasks relatively short and incorporating breaks between tasks to minimize such discomfort.

Furthermore, one participant noted that limitations in visual clarity and resolution affected immersion and realism. This was caused by the MR headset not rendering distant objects with sufficient clarity, especially for those located far away.

> *"The visuals were a bit blurry sometimes, especially when looking at objects in the distance."* [G5/P13]

4.2.9 Discussion

Our results showed that MRMAC supported completing the remote collaboration tasks effectively and efficiently.

Spatial Presence: The mean IPQ score was above the midpoint suggesting that the live 360° video of the MRC space induces a high sense of spatial presence in participants. They reported feeling as if they were there and able to look around and see other people. The high-resolution live 360-video streaming was a key factor in this, as reported by the most of participants with comments such as the following, which supports the hypothesis **H1**.

> *"[...] I felt the resolution was excellent and I was immersed"* [G5/P13]

> *"[...] I felt like I was in a real location."* [G1/P3]

Social Presence: Collaborators experienced a similar sense of co-presence in both remote and local settings, which may facilitate effective cooperation as participants feel physically present in the same location. Participants were also engaged in the MRC space, as indicated by the involvement score of the IPQ sub-scales being higher than midpoint scores for VR-Travelers. The results revealed significant differences in the experience of cooperative social presence (team cohesion and involvement). Therefore it supports our hypothesis **H2** such that both VR-Traveler and AR-Host have an equally high social presence.

Usability: The overall scores from SUS for MRMAC indicated high usability. Both AR-Host and VR-Traveler participants reported the system was efficient and suitable for the collaborative task. Participants were able to become familiar with MRMAC quickly and effortlessly, and it required less mental effort to interact and collaborate with others. According to our results, the total number of errors was fairly low, and there was no significant difference between the two groups. This result strongly supports **H3**.

Task Performance: The average puzzle times show a slight learning effect, as participants became familiar with the interaction and improved their task-oriented language. There were no significant differences regarding puzzle completion times and TLX workload. There was no significant difference in collaborative performance in the puzzle task. The task performance results demonstrate that participants were able to interact with the remote collaborators effectively and efficiently in different places. This supports our hypothesis **H4**. In addition to improving task performance, our results showed improved situation awareness for participants using AR or VR. The results on workload indicated that participants in AR and VR felt they were most successful with tasks, although there is no significant difference.

Feedback and Observations: Throughout the study, verbal communication was predominant using combined spatial and temporal references (e.g."left to the piece you have previously moved."). We observed that participants developed an informal shared protocol to better understand how to complete the task. Explicit line and column indications were rarely used and had a negative impact in all instances. Indications like "third row, second column" were more difficult to decipher than temporal references. Some participants found viewpoint sharing in PiP mode to be extremely helpful and reported that it reduced attention loss and made it easier to focus on remote user interaction while performing model-building tasks:

> *"[...] viewing instructions in a mini-window was more helpful than figuring out by looking at the user Avatar, I could focus on my desk and view instructions at the same time."* [G7/P22]

We found that the VR-Travelers led most of the interactions, as the AR-Host felt the VR-Traveler contributed more, which is to be expected since the VR-Travelers were guiding the AR-Host.

Limitations and Future Work: Throughout the study, a single 360° camera was used. However, employing multiple 360° cameras could prove more beneficial in covering a larger area of the collaboration space, aligning with the future research direction. Further investigation involving a larger number of participants would be the next step. Additionally, integrating other visual cues such as virtual replicas of physical objects and full-body avatars would be intriguing avenues for exploration in future studies. Moreover, revising the user study to emphasize discussion and co-located collaboration could highlight the advantages of multi-user remote communication in addressing cooperation and communication scenarios that may arise in real-world applications.

In the study, the analysis did not include an examination of potential short-term boosts in results between participants with and without prior experience of AR/VR, possibly due to the novelty value of using the technology. However, a small subset of participants had prior experience, constituting less than 20%. With such a limited subgroup, drawing statistically significant conclusions or identifying meaningful patterns in the data becomes challenging. However, future research could consider recruiting a larger and more diverse sample of participants with varying levels of experience in VR/AR. This would allow for a more comprehensive investigation into the influence of past experience on collaboration outcomes in mixed-reality environments and help uncover potential short-term effects of technology novelty on user performance.

Several use cases have been identified that could be explored in the future to demonstrate the need for such collaboration tools. One case is workflows for field workers when they need the help of an expert onsite. With MRMAC experts can collaborate remotely to help field workers solve problems, see what field workers can see, annotate the working space, and make sure tasks are done correctly. Another case is training people on new products, processes, and equipment where spatial context and a sense of co-presence must communicate together. Another case is related to space planning, where people help to effectively design and plan their space. MRMAC can allow people to view their digital asset designs, see them at scale, review them, and iterate them before making financial decisions.

Participants were highly receptive to MRMAC, as both VR and AR users reported high levels of co-presence, as well as increased viewing flexibility in a remote environment and being personally involved in the task. When participants were in different spaces with an asymmetric setup they felt similar levels of presence and had high satisfaction with the system. MRMAC shows strong potential in addressing the needs of multi-user asymmetric remote collaboration by enabling an interaction, cross-device communication platform. The design implications of our work will help improve video conferencing software by incorporating virtual telepresence and teleportation elements to produce a better user experience for mixed reality multi-user asymmetric collaboration.

4.3 MRMAC Evaluation: Baseline Comparison

In Section 4.2, the impact of user roles on collaboration within the MRMAC system was examined. The results of the study were promising, indicating a similar and high sense of presence in the collaboration space for both local and remote users. Participants also reported favorable ratings for system usability, experienced relatively low workload, and demonstrated better task performance. This section focuses on conducting a comparative evaluation of the system against traditional collaboration systems. Building upon the lessons learned from the initial user study, various improvements have been incorporated. The goal of this user study is to understand the effectiveness of the system in enhancing collaborative experiences compared to other baseline conditions.

A user study (N = 36) was conducted to evaluate the key design features: communication and awareness (see Figure 4.15). To measure MRMAC's effectiveness for communication and awareness among participants, MRMAC was compared to two other experimental conditions: conventional 2D and standard 360-degree videoconferencing.

4.3.1 Design and Methodology

A 3 × 2 mixed factorial design was employed, where participants were assigned a role (between-subjects factor: *local vs. remote*) and then experienced three distinct conditions (within-subjects factor: *C1 vs. C2 vs. C3*) in random order. Each group consisted of four

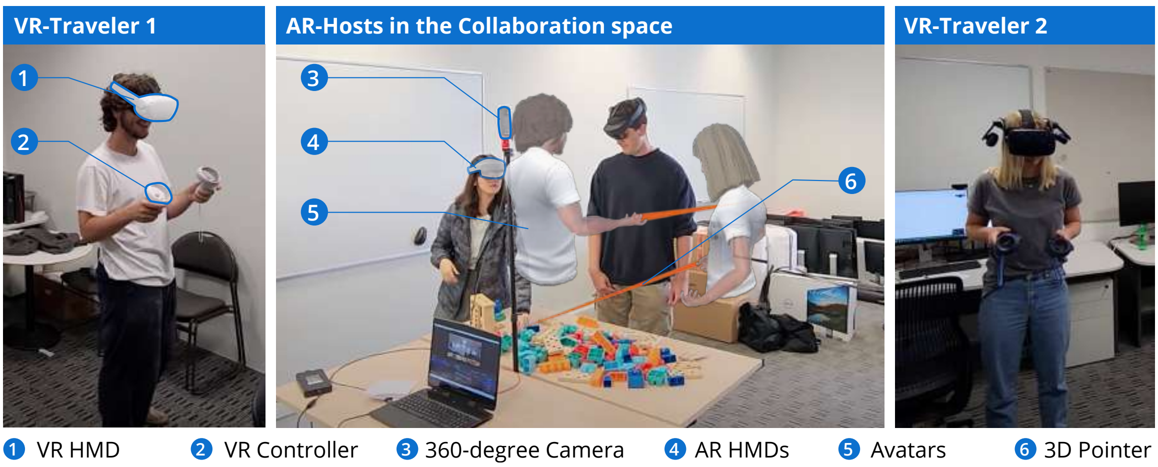


Figure 4.15: Comparison of baseline setups in MRMAC evaluation, including the local AR-Hosts and remote VR-Travelers, 360° camera, controls, and display setup.

participants, with two physically located together (AR-Hosts) and two remotely located (VR-Travelers).

4.3.2 Experimental tasks

The experiment involves a collaborative task using Lego bricks and dominoes. VR-Travelers guide AR-Hosts, in constructing predetermined models. VR-Travelers have access to a visual representation of the final model, while AR-Hosts can only access the physical blocks, which are randomly placed. The task involves searching for specific pieces, identifying their shapes, colors, and dots, and assembling the model based on VR-Travelers' guidance. The task is considered complete when both VR-Travelers and AR-Hosts are satisfied with the final structure.

4.3.3 Experimental Conditions and Setup

The study compares three experimental conditions to explore how different technological setups and modes of communication, interaction, and immersion affect the quality and efficiency of instruction as the interaction and communication between participants (Figure 4.16).

In all conditions, there are variations in communication or awareness cues. For instance, in C2 and C3 for the VR-Traveller, communication and interaction cues differ. However, not all combinations of interfaces were included in the current study due to previous extensive exploration in related research. By including these three conditions (C1, C2, and C3), the study effectively covers the range of interface combinations that were of interest to address specific variations in communication and awareness cues. While it's acknowledged that not all possible combinations were examined, the chosen conditions were selected to focus on key variations and their potential impact on collaboration outcomes. Below are the three experimental conditions designed for the user study:

- *C1. Conventional video with 2D annotation:* Condition C1 represents the conventional video conferencing, using a 2D camera (Logitech Brio) mounted on a desktop computer to stream the task space and participants. Real-time communication occurs between VR-Travelers and AR-Hosts through audio, video, and optional 2D annotations. VR-Travelers use a 3D blueprint of the Lego/Domino model to share their instructions.

- *C2. 360° video without augmented visual cues:* In this condition, VR-Travelers wear VR headsets (VIVE Pro 2 or Meta Quest 2) to view a 360° video stream (Ricoh Theta Z1)

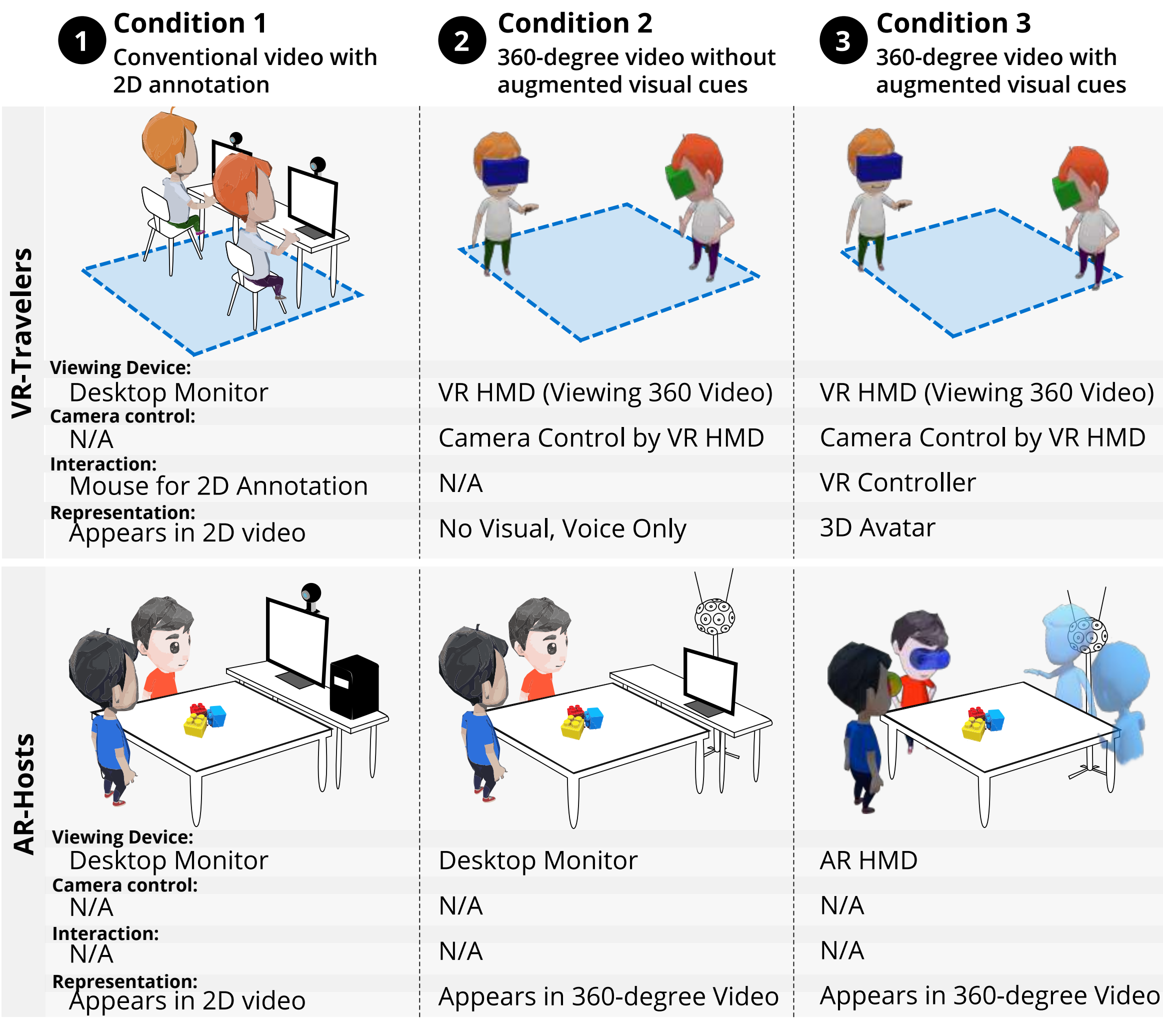


Figure 4.16: Overview of the User Study.

of the AR-Hosts space. On the other hand, AR-Hosts use a desktop computer to view the perspective view of the VR-Travelers. VR-Travelers see the 3D blueprint from their headset while moving and rotating them using a VR controller. Communication in C2 is limited to voice only, but the level of immersion is higher than in C1, as VR-Travelers can freely look around the local participant's space, as in [58, 210].

- *C3. 360° video with augmented visual cues:* Condition C3 is our MRMAC system. Real-time communication and interaction occur through audio, gestures, 3D pointer, and 3D annotations.

To maintain consistency and fair comparison, conditions C1 and C2 were implemented by varying the capture device and interaction options in the MRMAC.

4.3.4 Procedure

Participants were divided into nine groups, each with two VR-Travelers and two AR-Hosts. They completed a demographic questionnaire after being briefed about the experiment. The VR-Travelers and AR-Hosts were in separate rooms and completed the collaborative task under different conditions. Before starting each condition, all participants were given an overview of the application and were given five minutes to familiarize themselves with it. During the experiment, participants completed spatial presence, social presence, SUS, and NASA-TLX questionnaires after finishing each condition as an intermediate assessment. After completing all three experimental conditions, participants were given a post-experiment questionnaire where they were asked to rank the conditions based on their preferences. The conditions were presented in random order to each participant, and participants cannot assign the same rank to two different conditions. On average, the process took about an hour to complete, and participants were given a $20 supermarket voucher as a token of appreciation.

4.3.5 Measures and Hypotheses

The effectiveness of participants' communication and awareness of remote collaboration was evaluated considering key factors, including spatial presence, social presence, usability, task performance, and user preference. Spatial presence was assessed using the 14-item *iGroup Presence Questionnaire* (IPQ [204]) with four subscales: General Presence (GP), Realism (RL), Involvement (INV), and Spatial Presence (SP). For social presence, a questionnaire was compiled with four subscales: Behavioral Engagement (BE), Co-Presence (CP), Mutual Attention (MA), and Mutual Understanding (MU), based on Biocca et al. [28], Bale [14], and Hauber [88]. Usability was evaluated using the *System Usability Scale* (SUS). Task performance was measured by task success and logged completion times, while the *NASA TLX* was used to evaluate subjective workload. In the post-experiment questionnaire, participants ranked their preferred condition under various categories and provided qualitative feedback by answering open questions. The following hypotheses were derived:

H1 Spatial presence and Social presence, particularly co-presence, would be significantly higher in condition C3 than in conditions C1 and C2.

H2 Both task completion time and workload would be lower in condition C3 compared to conditions C1 and C2.

H3 System usability would be significantly higher in condition C3 compared to conditions C1 and C2.

H4 Participants would prefer condition C3 over conditions C1 and C2.

4.3.6 Participants

A total of 36 participants aged 18-81 years were recruited (μ = 30.83, σ = 14.10), comprising 17 males (47.2%) and 19 females (52.8%). Among them, 20 (55.6%) identified as European, 10 (27.8%) as Asian, 3 (8.3%) as Latin American, 2 (5.6%) as Polynesian, and 1 did not disclose their ethnicity. The majority of participants, 27 (75%), reported English as their primary language. Of the participants, 22 (61.1%) reported no prior VR/AR experience. Among those with prior experience, 2 reported using VR/AR weekly, 4 monthly, 4 less than once a year, and 4 once or more per year. All participants reported normal or corrected vision and hearing. Recruitment was conducted through email and flyers on a university campus, resulting in diverse backgrounds and experience levels.

4.3.7 Results

Data normality was assessed using the Shapiro-Wilk test (α = 0.05). For normally distributed data, a two-way repeated-measures ANOVA was conducted with Tukey's HSD as a post hoc test (α = 0.05). For non-normally distributed data, the Aligned Rank Transformation (ART) method [258] with the ART-C procedure [59] was employed. All post hoc analyses applied a Holm-Bonferroni correction for multiple comparisons.

Spatial presence

Overall ratings for condition C3 were notably higher in both local and remote roles (Remote: μ = 5.58, σ = 1.03; Local: μ = 5.54, σ = 1.12) than in the other conditions. C3 also had a higher overall IPQ score: C3: μ = 5.56, σ = 1.08 than C1: μ = 4.29, σ = 1.05 and C2: μ = 4.70, σ = 1.01. Figure 4.17 shows the average spatial presence rating. Significant main effects were found among the conditions (GP: $F_{2,68}$ = 19.30, p < .001; INV: $F_{2,176}$ = 27.58, p < .001; RL: $F_{2,176}$ = 34.39, p < .001; SP: $F_{2,500}$ = 56.79, p < .001). However, no significant effects were observed on the user roles (GP: $F_{1,34}$ = 0.35, p = .560; INV: $F_{1,34}$ = 0.33, p = .572; RL: $F_{1,34}$ = 2.35, p = .135; SP: $F_{1,34}$ = 2.75, p = .106) or their interaction (GP: $F_{2,68}$ = 2.11, p = .129; INV: $F_{2,176}$ = 0.01, p = .991; RL: $F_{2,176}$ = 1.01, p = .367; SP: $F_{2,500}$ = 0.49, p = .614). Post hoc pairwise comparisons indicated a significant difference in four subscales in all three conditions where C3 resulted in higher scores across these factors compared to both C1 (GP, INV, RL, SP: p < .0001) and C2 (GP: p = .0094, INV, RL, SP: p < .0001). Findings suggest that C3 resulted in significantly higher levels of immersion and connectedness to the remote

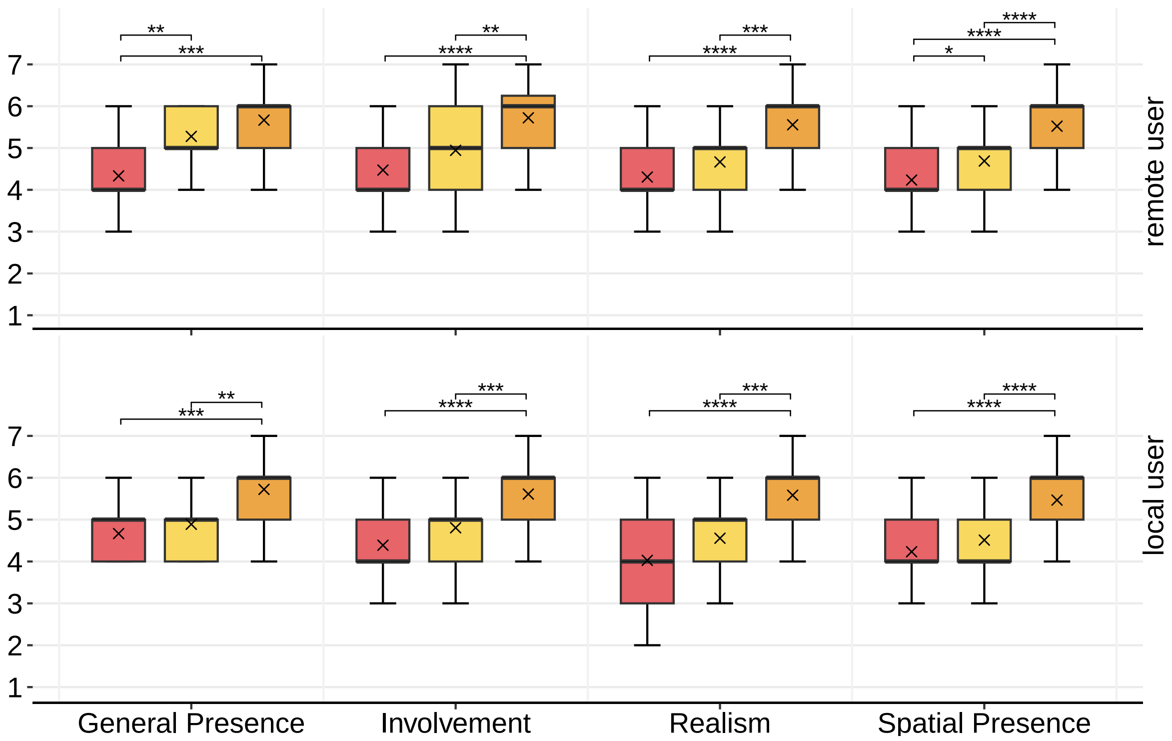


Figure 4.17: Spatial Presence scores across different conditions (C1 C2 C3).

environment than C1 and C2. For VR users, a significant difference was found between C1 and C2 for both Spatial Presence and General Presence scores ($p < .001$).

Social Presence

The average social presence rating is shown in Figure 4.18. Overall social presence ratings for condition C3 ($\mu = 5.80$, $\sigma = 0.89$) were higher than C1 ($\mu = 4.20$, $\sigma = 0.87$) and C2 ($\mu = 4.84$, $\sigma = 0.84$); also remote users rated slightly higher than local users (Remote: $\mu = 5.81$, $\sigma = 0.89$; Local: $\mu = 5.78$, $\sigma = 0.89$) but not statistically significant. Significant main effects were found among the conditions (BE: $F_{2,176} = 63.29$, $p < .001$; CP: $F_{2,392} = 200.25$, $p < .001$; MA: $F_{2,284} = 89.26$, $p < .001$; MU: $F_{2,176} = 35.68$, $p < .001$). However, no significant effects were found on the user roles (BE: $F_{1,34} = 0.52$, $p = .475$; CP: $F_{1,34} = 0.97$, $p = .330$; MA: $F_{1,34} = 0.08$, $p = .135$; MU: $F_{1,34} = 0.93$, $p = .339$) or their interaction (BE: $F_{2,176} = 0.06$, $p = .936$; CP: $F_{2,392} = 0.70$, $p = .495$; MA: $F_{2,284} = 1.57$, $p = .209$; MU: $F_{2,176} = 0.86$, $p = .421$). Post hoc pairwise comparisons indicated a significant difference in four subscales in all three conditions where C3 resulted in higher scores across these factors compared to both C1 (GP, INV, RL, SP: $p < .0001$) and C2 (GP, INV, RL, SP: $p < .0001$). Findings suggest that

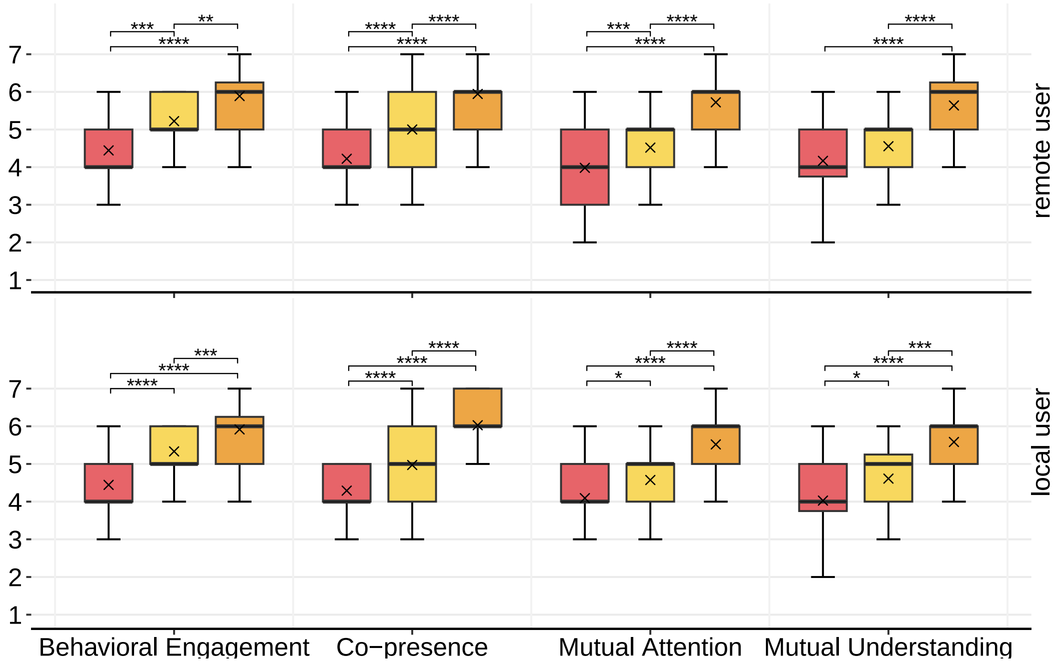


Figure 4.18: Social Presence scores across different conditions (C1 C2 C3).

C3 led to a significant increase in participants' social presence, fostering a stronger sense of connection and engagement with others compared to C1 and C2.

Task Performance

The average task completion time for each of the three conditions shown in Figure 4.19 suggests that condition C3 had the fastest completion time (μ = 396.88, σ = 92.79). Bartlett's test suggested that the homogeneity assumption of variance was met (χ^2 = 1.148, p = 0.563). A One-Way ANOVA suggested that task complexity was well-balanced ($F_{2,24}$ = .48, p = .51) across the three LEGO and Domino structures. Moreover, to control for the potential influence of task order, participants were randomly assigned to different conditions to randomly distribute any learning effect and help reduce its impact.

Figure 4.20 shows the task workload (mental load, effort, frustration, and overall) result where C3 had a lower overall score for both remote and local users (Remote: μ = 25.30, σ = 12.08; Local: μ = 24.79, σ = 11.73, combined: μ = 25.19, σ = 8.69). Significant main effects were found in all three conditions (Mental: $F_{2,68}$ = 13.54, p < .001, Effort: $F_{2,68}$ = 34.74, p < .001, Frustration: $F_{2,68}$ = 19.43, p < .001, Overall: $F_{2,68}$ = 8.50, p < .001), with no significant differences found between the roles or interaction effects. Pairwise comparisons suggest C2 led to a relatively lower overall workload than C1 (p < .001), while condition C3 resulted in

Table 4.3: Error frequency in the user study (mean with standard deviation in parentheses).

	Color	Shape	Final Output
C1	0.78	0.67	0.78
	(0.79)	*(0.47)*	*(0.63)*
C2	0.22	0.78	0.22
	(0.42)	*(0.63)*	*(0.42)*
C3	0.67	0.11	0.67
	(0.82)	*(0.31)*	*(0.67)*

the lowest perceived overall workload than both C1 ($p < .001$) and C2 ($p = .0004$).

Similar to the first user study, error frequency was assessed after each task completion, where errors were defined based on discrepancies between the assembled model and predefined criteria encompassing color, shape, and final output. The results are shown in Table 4.3, and error cases were defined similarly as described in Section 4.2.7. The repeated measures ANOVA revealed a marginally significant main effect of conditions $F_{2,8} = 2.882$, $p = 0.062$, suggesting potential differences in error frequencies across conditions (C1, C2,

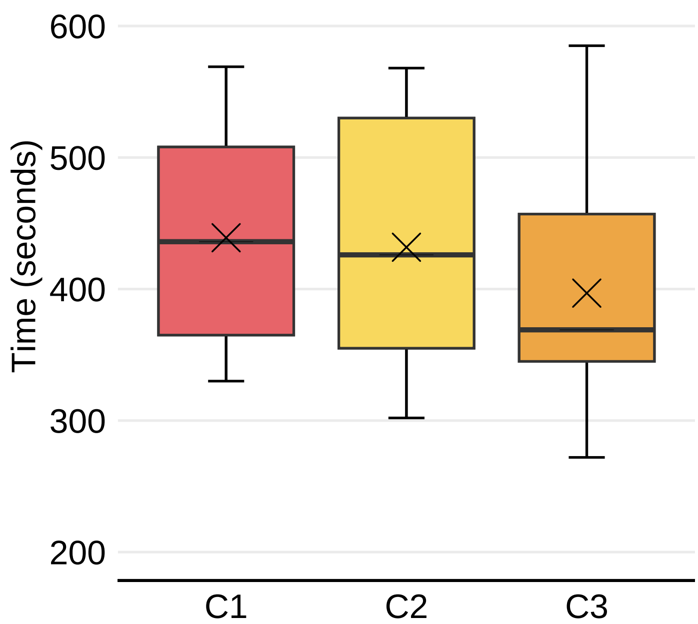


Figure 4.19: The task completion time for each condition.

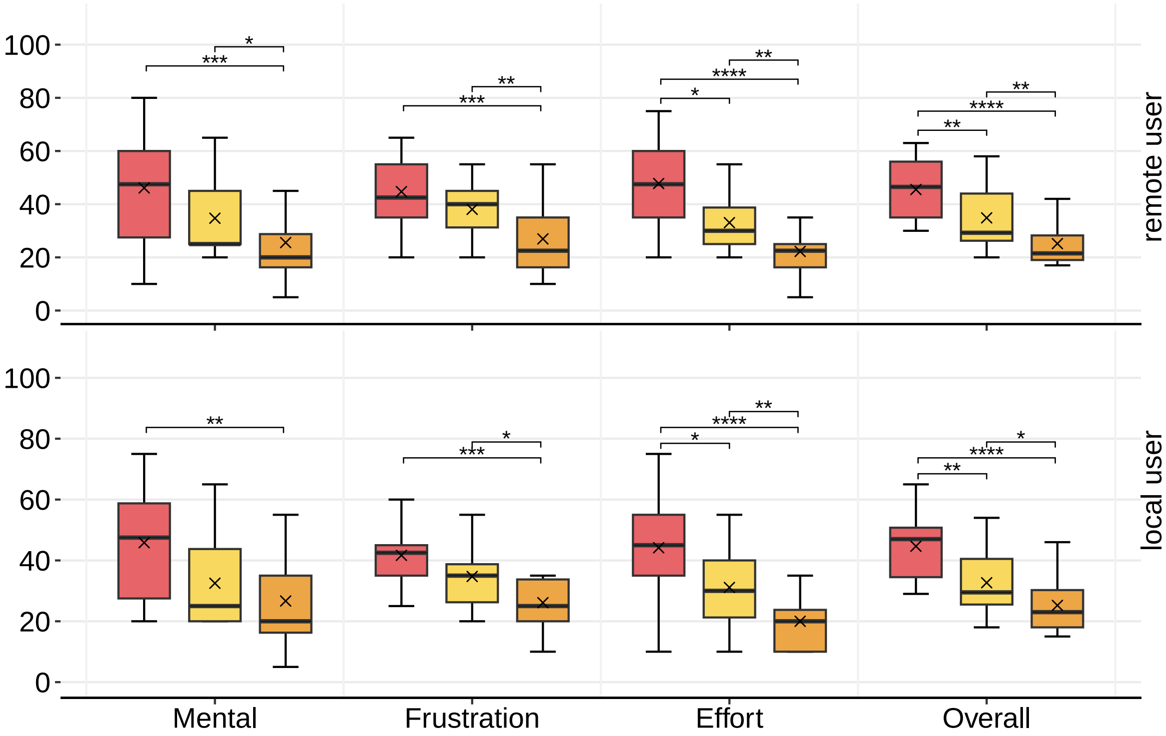


Figure 4.20: NASA-TLX score from 0 (low workload) to 100 (high workload) (lower is better) (Local Remote).

C3) for the varied tasks. However post-hoc comparisons using Tukey's HSD test indicated no significant differences between pairs of conditions. These findings imply that the error frequencies observed during the Lego/domino model building task may not be strongly influenced by the specific experimental conditions (C1, C2, C3).

System Usability

Significant main effects observed among the three conditions ($F_{2,68} = 81.55$, $p < .001$), with no significant differences identified between roles ($F_{1,34} = 0.51$, $p < .477$). However significant interaction effects were found for conditions × role ($F_{2,68} = 3.73$, $p < .028$). Post hoc pairwise comparisons showed significant differences between each condition (C1, C2, C3) ($p < .001$, $p < .001$, $p < .001$) (see Table 4.4). The interaction effect between C1 and C2 was statistically significant ($p = .018$) suggesting that the effect of the C1-C2 factor varies between roles. Figure 4.21 shows the average SUS score for each condition and role. Participants in the C3 provided significantly higher ratings (Local: $\mu = 84.44$, $\sigma = 7.50$; Remote: $\mu = 85.83$, $\sigma = 7.02$) compared to the C2 (Local: $\mu = 75.27$, $\sigma = 9.62$; Remote: $\mu = 72.36$, $\sigma = 8.33$) and

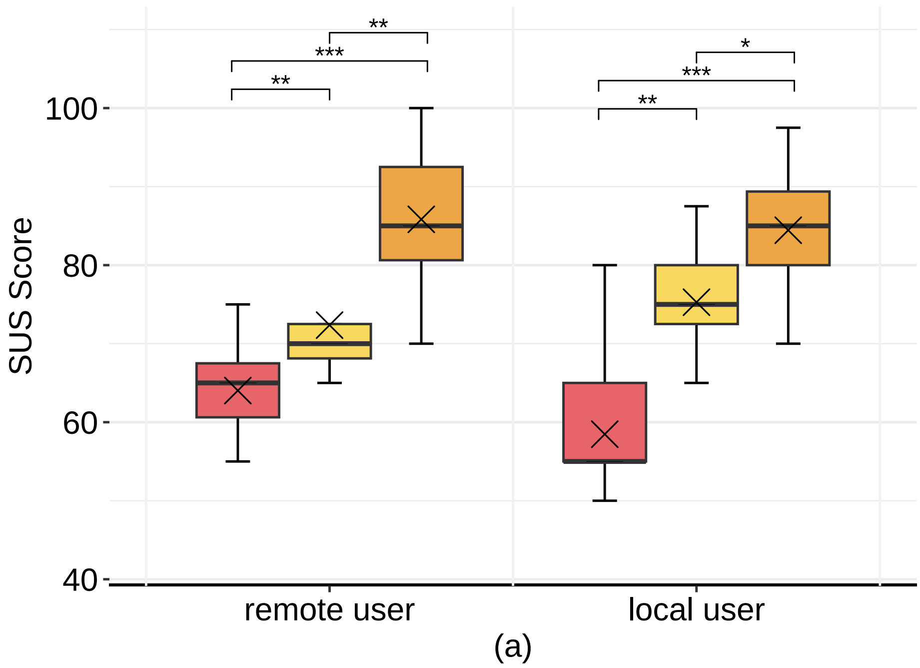


Figure 4.21: SUS results across different conditions. C1 C2 C3

the C1 (Local: μ = 58.47, σ = 8.00; Remote: μ = 64.02, σ = 8.83). The mean SUS score in the C3 condition was above 85, indicating an "excellent" level of usability. On the other hand, in the C2 and C1, both roles were at "good" and "poor" levels, respectively.

User Preferences

User preferences result is shown in Figure 4.22. Most of the participants ranked the C3 condition as their preferred choice, see Table 4.5. To assess participants' preferences across various conditions, both AR and VR were considered together, as there were no significant differences found among roles. A Friedman test was conducted to examine the differences

Table 4.4: Group Statistics and Post-Hoc Test Results of the SUS.

	$\bar{x}$ (Mean)	$\tilde{x}$ (Median)	s (SD)	F	Sig. (Diff)
C1	61.30	61.19	8.20		
C2	73.80	73.91	7.60	F(2,38) = 35.25	C1 ≤ C2 < C3
C3	85.10	85.03	6.90		

in rankings (R1, R2, R3) across the three conditions.

A Friedman test was conducted to assess differences among conditions (C1, C2, C3) for Q1, revealing a significant effect, $\chi^2(2) = 29.4$, $p < .001$. Pairwise comparisons using Wilcoxon rank sum tests showed that the comparison between C2 and C3 was significant ($p < .001$), while other comparisons were not statistically significant after Holm's adjustment.

For Q2, the Friedman test indicated a significant effect, $\chi^2(2) = 8.4071$, $p = 0.01494$. Pairwise comparisons using Wilcoxon tests showed that the difference between C1 and C3 was statistically significant ($p = 0.011$) after Holm's correction.

The Friedman test for Q3 revealed a significant effect, $\chi^2(2) = 38.475$, $p = .001$. Pairwise comparisons using Wilcoxon tests indicated significant differences between C1 and C2 ($p < .001$) and between C2 and C3 ($p = 0.0019$) after Holm's adjustment.

Similarly, for Q4, the Friedman test was significant, $\chi^2(2) = 14.458$, $p = 0.0007254$. Pairwise comparisons showed significant differences between C1 and C3 ($p = 0.00054$) after Holm's correction.

Finally, for Q5, the Friedman test revealed a significant effect, $\chi^2(2) = 40.371$, $p < .001$. Pairwise comparisons using Wilcoxon tests showed significant differences between C1 and C2 ($p < .001$) and between C2 and C3 ($p < .001$) after Holm's adjustment.

Post-hoc pairwise comparisons using Wilcoxon rank sum tests with continuity correction and Holm's adjustment were then conducted to identify specific differences between conditions. For Q1, significant differences were observed between C2 and C3 ($p < .001$). For Q2, a significant difference was found between C1 and C3 ($p = .011$). In Q3, significant differences were observed between C1 and C2 ($p < .001$) and between C2 and C3 ($p = 0.0019$). Q4 showed significant differences between C1 and C3 ($p = 0.00054$), and Q5 revealed significant differences between C1 and C2 ($p < .001$) and between C2 and C3 ($p < .001$).

These results suggest that there are significant differences among the conditions for each of the five questions, with specific pairwise differences observed in some cases after correcting for multiple comparisons.

The test revealed a significant difference in rankings ($\chi^2(2) = 24.6$, $p < .001$). The effect size, as measured by Kendall's W, was 0.467, indicating a moderate effect. To further investigate pairwise differences, Wilcoxon signed-rank tests were performed. The results indicated that for Q1, there was a significant difference between Condition C1 and C2 (Z = -0.323, p = .747, r = .066), suggesting a small effect size. Similarly, a significant difference was found between Condition C2 and C3 (Z = -1.060, p = .289, r = .216), with a small to

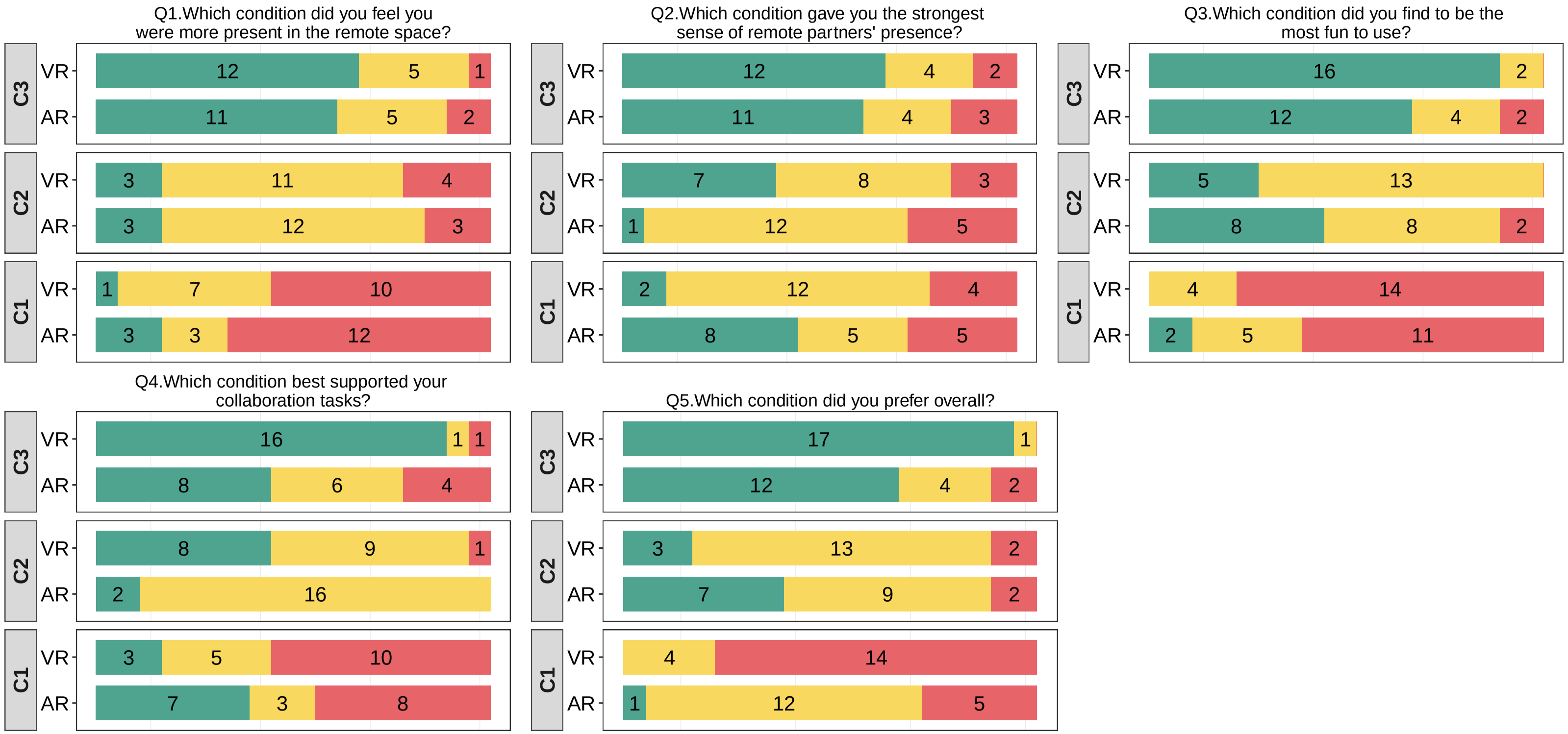


Figure 4.22: Users ranked the three conditions based on preference for presence, partner presence, fun, task support, and overall preference (Q1-Q5). Rank1 Rank2 Rank3 (Rank1 = most preferred rank, and C3 = MRMAC).

Table 4.5: Average ranking value of user preferences for each question, role, and condition.

Q1. Felt more present in remote space

Conditions	Role	Rank 1	Rank 2	Rank 3
C3	VR	**12**	5	1
	AR	**11**	5	2
C2	VR	3	11	4
	AR	3	12	3
C1	VR	1	7	10
	AR	3	3	12

Q2. Strongest sense of remote partners' presence

Conditions	Role	Rank 1	Rank 2	Rank 3
C3	VR	**12**	4	2
	AR	**11**	4	3
C2	VR	7	8	3
	AR	1	12	5
C1	VR	2	12	4
	AR	8	5	5

Q3. Most fun to use

Conditions	Role	Rank 1	Rank 2	Rank 3
C3	VR	**16**	2	0
	AR	**12**	4	2
C2	VR	5	13	0
	AR	8	8	2
C1	VR	0	4	14
	AR	2	5	11

Q4. Best for collaboration tasks

Conditions	Role	Rank 1	Rank 2	Rank 3
C3	VR	**16**	1	1
	AR	**8**	6	4
C2	VR	8	9	1
	AR	2	16	0
C1	VR	3	5	10
	AR	7	3	8

Q5. Overall preference

Conditions	Role	Rank 1	Rank 2	Rank 3
C3	VR	**17**	1	0
	AR	12	4	2
C2	VR	3	13	2
	AR	7	9	2
C1	VR	0	4	14
	AR	1	12	5

medium effect size for Q2. For Q3, most participants preferred to use Condition C2 and C3 over Condition C1. However, the effect sizes for the pairwise comparisons were small to medium, suggesting that the practical significance of these differences might be limited. Moreover, a significant difference was observed for Q4, between C1 and C3 (Z = -1.034, p = .129, r = .065), with an effect size of medium. For Q5, the significant differences observed between the conditions indicate a preference hierarchy, with C3 being the most preferred, followed by C2 and C1.

4.3.8 Discussion

As MRMAC is the first working system that supports multi-user mixed reality telecollaboration in an asymmetric setup with low-latency data synchronization, conducting side-by-side system performance comparisons with prior works wasn't possible, as they only supported one-to-one collaboration scenarios [169, 170, 228, 193]. The user study was carefully designed to address this issue by comparing against two baseline conditions: conventional 2D and standard 360-degree videoconferencing.

Sense of Presence

The study's first hypothesis (**H1**) proposed that users would experience significantly higher spatial and social presence in condition C3 compared to the other two baseline conditions. The experimental results supported this hypothesis, as C3 received higher scores on the overall IPQ and spatial presence subscales compared to C1 and C2. Differences in communication and awareness cues in C3 likely contributed to these variations, thus supporting hypothesis **H1**.

Additionally, the overall social presence ratings were higher in C3 compared to the other conditions, with a significant difference, further reinforcing hypothesis **H1**.

The appearance of avatars also appears to be influenced social presence by allowing users to personalize their virtual representation, enhancing their sense of identification and connection with the virtual environment.

> *"Seeing avatars that look like us makes the whole thing feel more real. It's like bringing a bit of ourselves into the digital world!"* [G6/P22]

Spatial sounds have helped them locate the users' positions in the MRC space and be aware of their partners' activities and whereabouts. Remarks included:

> *"[...] like being surrounded in a rich tapestry of sounds, I can hear where other girls are and what they're doing and saying clearly."* [G3/P11]

> *"[...] the clarity in the remote space was impressive. I could hear every word clearly and see smallest details without any issues."* [G4/P14]

View sharing was another factor enabled participants to see where other participants were looking, suggested a more natural and engaging collaborative environment, thereby enhancing social presence.

> *"[...] to see where my colleagues were looking was quite interesting. It took some time to realize at the beginning, but I quickly got used to it intuitively."* [G5/P17]

Task Performance

The hypothesis (**H2**) stated that both task completion time and workload would be lower in condition C3 compared to conditions C1 and C2. While condition C3 demonstrated the fastest task completion time among the tested conditions, it's important to note that this difference was not statistically significant. This suggests that the observed differences in completion time may not be reliably attributed solely to the additional features of C3. Factors beyond the conditions tested, such as individual differences or random variability, may have contributed to the observed pattern. Further exploration with a larger sample size could provide deeper insights into the nuanced effects of experimental conditions on task completion time.

In contrast to the non-significant findings on task completion time, the results strongly support hypothesis **H2** regarding workload. Condition C3 was consistently associated with the lowest perceived workload for both remote and local users, as indicated by measures of mental load, effort, frustration, and overall workload.

This finding suggests the importance of C3's variations in communication or awareness cues, such as 3D annotation and 3D pointers, which likely facilitated the location of desired LEGO/Domino blocks and minimized the need for explicit visual scanning and communication.

> *"[...] having those 3D sticks (pointers) really helped me during the model making. They made it much easier to point to the right blocks without needing to describe or explain everything."* [G1/P3]
>
> *"I really loved being able to draw in space. It was so cool to sketch out ideas and plans right in front of me."* [G2/P8]

Therefore, hypothesis (**H2**) was partially supported, with condition C3 demonstrating significantly lower perceived workload, but not statistically significant differences in task completion time compared to conditions C1 and C2.

System Usability

Hypothesis **H3** proposed that system usability would be significantly higher in condition C3 compared to conditions C1 and C2. The study's results strongly support this hypothesis, with participants in condition C3 providing significantly higher ratings for system usability compared to conditions C1 and C2. The mean SUS score of above 85 in condition C3 signifies an "excellent" level of usability, reflecting participants' positive experiences with the system. The observed differences in SUS scores across conditions highlight the impact of specific features and design elements present in condition C3, which likely contributed to enhanced user satisfaction and perceived usability.

The higher usability ratings in condition C3 likely attributed to several factors, including the incorporation of intuitive navigation tools, clear visual cues, effective communication features (such as 3D annotation and pointers), and overall system responsiveness, including real-time streaming of 360° video and audio-video synchronization. These elements collectively improved user interaction leading to more positive perceptions of system usability.

> *"I had no trouble communicating with others. The tools were easy to use, and I felt heard and understood."* [G1/P4]

Feedback and Observations

Feedback from participants supports our fourth hypothesis regarding participants' preference for C3 over C1 and C3. Out of 36 participants, 25 ranked C3 higher across Q1-Q5, with 29 out of 36 participants ranking C3 as their preferred condition overall. This strongly supports **H4**, indicating that C3 is the most preferred option. The findings of our study revealed favorable outcomes concerning the system performance and usability of MRMAC. Notably, eye contact and gestures held particular significance in guiding attention and expressing intentions. While vocal and verbal communication proved to be crucial, non-verbal cues also played a significant role. In some cases, one co-located participant understood the instruction faster and guided their counterpart. For communication between VR-Travelers, avatar representation helped spatial awareness and natural communication:

> *"[...] live-streaming the physical environment and blending 3D virtual assets really amps up the collaboration experience"* [G3/P12]

.

Spatial audio, like in AR, was essential for VR users to facilitate communication. Visual cues in VR were also crucial for guiding attention to specific areas or objects:

> *"It feels like we're all in the same room. I can hear everyone's voices coming from different directions, it is even more realistic"* [G8/P31]

.

Voice communication was equally crucial for VR pairs to discuss problems and coordinate actions in real-time.

> *"It felt like I was in the same room with the people in the other room. I could see what they were doing and which block they were picking up, making it easier for me to give instructions."* [G4/P15]

Limitations and Future Works

The number of collaborators per group in the user study was limited to four (two local and two remote) due to space, equipment, and time constraints, which prevented testing with larger groups. Another contributing factor was the COVID-19 pandemic, which made it difficult to recruit more participants as the study was conducted at the tail end of the pandemic.

Therefore, participant groups in the study were convenience samples, which may have introduced biases into the sample population. As a result, to generalize findings of the study to the wider population, conducting more user studies with a broader population could be a future direction. Although efforts have been made to minimize latency, there is room for improvement, especially for larger collaboration scenarios. Additionally, while compression standards offer extensions that can reduce overall bandwidth, these extensions are primarily designed for camera arrays and lack real-time encoding implementations. Adapting these extensions to incorporate contextual knowledge about the location and movement of participants could enhance the system's ability to handle bandwidth limitations and improve overall performance. Lastly, while the system provides high audio-visual fidelity for remote collaboration using 360° video, the absence of depth perception may pose challenges for certain collaborative tasks that require precise spatial understanding or mixed-reality collision handling.

Looking ahead, the aim is to enhance MRMAC by supporting more AR-Hosts and VR-Travelers and incorporating additional 360° cameras. Although multiple avatars were positioned using circular offset, this method has limitations in accommodating many users in a large space. Expanding the collaboration space by utilizing multiple 360° cameras could be a potential direction. This approach would accommodate a larger number of avatars and cover a wider area, offering ample space for research. Exploring optimal positioning between avatars and cameras is also an area for further exploration. In the future, it would be beneficial to explore the use of a 3D VR reconstruction of the workspace as well as 360-degree video to assess the difference in immersion this might create.

Additionally, enabling depth streaming for better spatial understanding and 6-DoF movement is an objective. Moreover, there is potential to incorporate more sensory features, such as haptic and tactile interfaces, to facilitate richer communication and collaboration among users beyond audio-visual cues. By integrating haptic feedback into virtual environments, users can feel virtual objects and interactions, enhancing the sense of presence and realism. Tactile interfaces, like gloves or suits, could enable users to physically interact with digital content, opening up new possibilities for training, gaming, and remote collaboration. Advancement in display technologies could greatly enhance virtual experiences. By upgrading resolution, expanding fields of view, and achieving better color reproduction in HMDs. With higher-resolution displays, users can enjoy clearer and more detailed visuals, reducing the "screen door effect" that can detract from immersion. In parallel, advancements in rendering techniques are crucial for achieving more immersive visuals. Telerobotics integration can facilitate remote users' physical access in future research. These advancements can further enhance MRMAC's effectiveness and usability across various domains.

4.4 Conclusion

This chapter presented a comprehensive evaluation of MRMAC in achieving the different aspects of the primary research questions outlined in **RQ1** (teleportation), **RQ2** (representation), and **RQ3** (interaction). The evaluation was conducted through performance evaluation and two user studies. System performance results helped assess the technical capabilities, including high frame rate, low latency, and scalability of the MRMAC system.

Findings from the system evaluation showcased the system's capability to provide low-latency synchronized communication for both AR and VR users. The scalability results indicated that network latency remained under 1 second (650ms) for up to 15 users.

The user studies provided valuable insights into the effectiveness of MRMAC in enhancing communication and awareness among participants in mixed-reality environments.

In the first user study, which examined how user roles affect presence, interactions, and task performance, it was found that user roles did not significantly affect the immersive experience, with both roles experiencing an equal level of presence and no notable changes in task performance. Participants were highly receptive to the system, as both VR and AR users reported high levels of co-presence, as well as increased viewing flexibility in a remote environment and being personally involved in the task. When participants were in different spaces with an asymmetric setup, they felt similar levels of presence and had high satisfaction with the system.

In the second user study, which compared MRMAC against existing collaborative tools, it was observed that MRMAC induces significantly higher spatial presence than 2D and 360° video alone. Furthermore, the findings revealed that the system reduced the difficulty of collaborative work, leading to a lower workload and enabling participants to perform more efficiently.

Overall, the evaluation and user studies demonstrated that rich real-time multiuser telepresence experiences can feasibly be created. It was also noted that these experiences still pose a significant engineering challenge. Despite this challenge, the findings opened up a plethora of future research opportunities in the field.

In summary, MRMAC has shown that rich real-time telepresence experiences can feasibly be created. MRMAC was able to induce significantly more spatial presence than 2D and 360 video alone. Although multiuser mixed reality collaboration poses a significant engineering challenge, MRMAC has opened up a lot of future research opportunities.

Why travel with hefty backpack when you can teleport with MRMAC?

Chapter 5

Vicarious: Context-aware Viewpoints Selection for Mixed Reality Collaboration

Mixed-perspective, which combines egocentric (first-person) and exocentric (third-person) viewpoints, has been shown to enhance the collaborative experience in remote settings. Such experiences allow remote users to switch between different viewpoints to gain alternative perspectives of the remote space. However, existing systems often lack seamless selection and transition between multiple perspectives that better fit the task at hand. To address this limitation, a new approach called *Vicarious* is presented. *Vicarious* simplifies and automates the selection between egocentric and exocentric viewpoints. It employs a context-aware method for dynamically switching or highlighting the optimal viewpoint based on user actions and the current context. To evaluate the effectiveness of the viewpoint selection method, a user study (n = 27) was conducted using an asymmetric AR-VR setup, where users performed remote collaboration tasks under four distinct conditions: *No-view, Manual, Guided*, and *Automatic* selection. The results of the study showed that *Guided* and *Automatic* viewpoint selection improved users' understanding of the task space and task performance, and reduced cognitive load compared to *Manual* or *No-view* selection. The findings also suggested that the asymmetric setup had minimal impact on spatial and social presence, except for differences in task load and preference. Based on these findings, design implications for future research in mixed reality collaboration are provided, aiming to further enhance user experience and task performance in remote collaboration scenarios.

5.1 Introduction

In MRMAC, viewsharing for multiple users was incorporated to facilitate collaborative interactions. This support for multiple users with their viewpoint sharing motivated the exploration of multi-perspective collaboration, laying the foundation for subsequent research endeavors. This progression naturally evolved to address the nuanced aspects of multi-perspective collaboration. Specifically, it aims to address the questions of how users' viewpoints can be managed effectively (**RQ4**) and how multi-perspective collaboration can be effectively achieved for supporting collaborative tasks (**RQ5**). While traditional video conferencing tools have gained popularity for remote collaboration, they come with limitations, such as a lack of spatial presence and peripheral awareness, which hinder the seamless exchange of information and coordination among collaborators [182]. Immersive collaboration solutions such as VR/AR/MR address these by creating shared spaces through 3D reconstruction using depth sensors and/or photogrammetry [230, 237], or 360° views captured by panorama cameras [193]. However, due to the limited field of view of head-mounted displays, the annotations or actions of remote users may not always be visible.

Researchers have investigated various visual communication cues, including the pointer [112, 126], gaze [83, 91, 124], view frustum [223, 152, 72], and combinations of these cues [114, 113], to guide remote user attention [68, 170]. In complex tasks, a single viewpoint can be difficult to understand and explore the dynamic physical space. Moreover, it becomes difficult to perceive others' actions without constantly shifting one's focus [100, 105]. Therefore, prior work has explored multiple viewpoints combining egocentric and exocentric views to share with remote users [117, 62]. This approach enables users to switch perspectives and share their views in real time, enhancing their overall understanding of the environment.

However, managing multiple viewpoints presents challenges for users to decide which views to focus on, potentially leading to overlooked information or user actions. One solution is for users to physically shift or toggle viewpoints, but this consumes time and effort [89]. Moreover, relying on users to navigate viewpoints can be inefficient, adding to the cognitive load of choosing the right perspective at any given time [1].

To address this, a novel approach called Vicarious is presented that simplifies and automates the selection of ego- and exocentric viewpoints (see Figure 5.1). A context-aware method is employed for selecting and dynamically switching or highlighting optimal viewpoints based on user actions and the current context. A user study (n = 27) was conducted where remote users guided a local user in a collaborative task in the local space under four conditions: *No-view*, *Manual*, *Guided*, and *Automatic* selection. Results indicate that *Guided* and *Automatic* viewpoint selection improved understanding of the local space and task per-

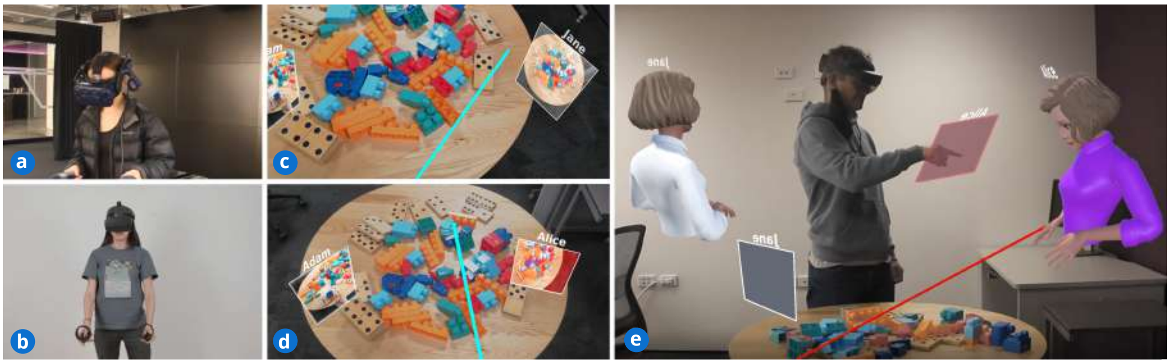


Figure 5.1: Two remote users (a and b) and their corresponding viewpoints (c and d) guide a local user (e) in a physical task. With Vicarious, all users have access to ego- and exocentric viewpoints, and the context-aware viewpoint selection method highlights the optimal viewpoint.

formance and reduced cognitive load compared to *Manual* or *No-view* selection.

> "Vicarious" means experiencing something through another person or as a result of their actions. So, "working vicariously" implies a situation where someone is not directly involved in doing the work themselves but is still benefiting from the work or its outcomes, by delegating tasks or supervising remotely.

Parts of the work presented in this chapter have been published in the Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST) (Zaman et al., 2023) [270].

The main contributions of this chapter are as follows:

- A *context-aware viewpoint selection method* that simplifies and automates the selection of egocentric and exocentric viewpoints based on visual saliency, user actions, and speech patterns.
- A *user study (n = 27)* evaluate the impact of context-aware viewpoint selections method on collaboration performance and user experience under four distinct conditions (*No-view, Manual, Guided, and Automatic*) and two user roles (*local and remote*).
- The results provide *insights and recommendations* for design implications and directions for future research.

5.2 Related Work

This research builds upon earlier work on viewpoint sharing in telepresence and mixed-perspective sharing. In this section, existing research within these areas is reviewed, and the research gap that this chapter addresses is highlighted.

5.2.1 Viewpoint Sharing in Telepresence

Viewpoint sharing has been studied for decades, as it allows sharing remote user perspectives with other collaborators and experiencing the remote environment through others' eyes as if they were physically present at that location [206]. Telepresence systems have explored the concept of out-of-body view where a live 360° video of a local person's surroundings shared with remote collaborators [106, 117, 153]. It allows the user to seamlessly switch between a first-person perspective, and a third-person perspective to explore the remote workspace and improves the sense of presence for remote collaborators [117].

Veas et al. [241] showed that having multiple viewpoints improved spatial understanding and situational awareness during collaboration. Chellali et al. [43] found sharing viewpoints enhanced co-presence and awareness during remote object manipulation tasks. The lack of awareness of the remote user's view direction has been found to diminish users' sense of embodiment [118] while providing independent viewing directions improved the sense of presence for remote users [198, 266]. Integrating point clouds with 360-degree videos enhances the viewing experience by providing depth perception, 3D scene reconstruction, and improved interaction [230, 267]. This fusion offers a more immersive environment and a better understanding of spatial relationships. However, technical challenges like data processing and alignment need to be addressed for effective integration.

Flying or aerial telepresence explored alternative viewing positions based on unmanned aerial vehicles (UAVs) or drones emphasizing the significance of viewpoint in remote collaboration [92, 199], teleoperation [235], and telepresence [172]. Viewpoint manipulation in such applications had a large impact on user perception, control ease, collaboration, and overall system effectiveness [207].

These approaches highlight some of the fundamental challenges related to user attention management, user engagement, and communication clarity when using viewpoint sharing to coordinate with remote collaborators in MR collaboration. While the use of viewpoint sharing in MR-based collaboration is not new, the novelty of this chapter lies in combining visual cues with contextual information to automatically suggest which view users should be focusing on. This aspect has not been explored or evaluated in prior work.

5.2.2 Mixed-Perspectives Sharing

Mixed-perspective representations have been extensively studied in prior work. For example, in Dollhouse VR [95] multiple users are able to navigate an interior design using a combination of exocentric and egocentric perspectives were found useful for gaining an overall understanding of the interior design and spatial layout for users outside the VR. Similarly, ShareVR [82] combines floor projection, mobile displays, and positional tracking to render the virtual world to non-HMD users to enable multiple perspectives of the shared physical and virtual space. TransceiVR [238] leverages screen sharing and spatial annotation to enable external users to explore VR scenes, reducing communication barriers between users in VR and non-VR.

ARgus [62] explores the trade-offs of such mixed representations in an AR workspace, where they combine different view representations (Headset View, External View, and Virtual View). They found mixed representations improve efficiency and spatial understanding, and reduce reliance on verbal instructions.

Mixed perspectives are also explored in multi-scale interactions such as Voodoo Dolls [168] which provides the user working at multiple scales, manipulating both visible and occluded objects, with an additional thumbnail view of the selected object. Snow Dome [171] introduces mixed-perspective representations by placing a remote VR user within a virtual 3D reconstruction of an AR user's space. This setup allows the VR user to experience the environment as a giant or miniature, offering them an overview of the AR user's workplace from different points of view and scales.

CollaVR [155], 360Anywhere [215], and SpaceTime [260] demonstrated viewpoint sharing increases the overall context in synchronous collaboration scenarios. Collaborative MR systems [128, 230] enable remote users to have an independent view of the overall task space through live 360 panoramas and reconstructed 3D models. Remote users can view and annotate the 3D scene from a different perspective, resulting in reduced task completion time.

While these existing systems offer manual view selection, there is potential for automated view selection to aid users in selecting viewpoints that enhance collaboration and provide valuable context. This potential to enhance collaboration by streamlining the process of identifying relevant viewpoints and providing valuable context warrants further exploration. It remains to be seen if this leads to increased efficiency in decision-making, improved spatial awareness among participants, and a more immersive collaborative experience.

Table 5.1: A high-level comparison of previous work outlined based on viewpoint sharing.

	Selection Method	Input	Viewpoints	Annotation	Collaboration Type
Sasikumar et al. [200], Gao et al. [72]	×	Pointcloud	Ego	✓	One-to-One
Sodhi et al. [212]	×	2D, Pointcloud	Exo	✓	One-to-One
Le et al. [123]	×	2D, Pointcloud	Ego, Exo	✓	One-to-One
Muller et al. [151], Kratz et al. [118]	×	2D	Ego	×	One-to-One
Billinghurst et al. [24]	×	2D	Ego	✓	One-to-One
Piumsomboon et al. [174]	×	2D	Exo	✓	One-to-One
Young et al. [266]	×	2D,360	Ego	×	One-to-One
Ryskeldiev et al. [198]	×	2D,360	Ego	✓	One-to-One
Kasahara et al. [105], Komiyama et al. [117]	Manual	360,2D,3D	Ego,Exo	×	One-to-One
Lee et al. [125]	Manual	360,2D,3D	Ego, Exo	✓	One-to-Many
Young et al. [267]	Manual	360, Pointcloud	Ego, Exo	×	One-to-One
Vicarious	Manual, Suggestive, and Automatic	360, 2D	Ego, Exo	✓	Many-to-Many

5.2.3 Viewpoint Sharing and Transition

The earliest work on viewpoint sharing and transition techniques used a dot in the user's field of view (FoV) to indicate the gaze of the remote user to the local users and a picture-in-picture (PiP) mode to display remote users hand gesture [45]. Magic Book [23] presents a collaborative approach, where one person has an egocentric point of view of the inside of a book via VR, while another person can look from an exocentric perspective and allows seamlessly transport users between Reality and Virtuality. This approach has been seen as being particularly helpful for navigation tasks [79, 218].

Phillips and Piekarski [167] explore a possession metaphor in AR-to-VR transition that allows players to quickly switch viewpoints without physically traveling. While in Mobileportation [267], the user can switch between exo- and egocentric views simply by walking up to or away from their partner's avatar. Maintaining visual context is important during the transition as it helps the participant understand where they are in the global context. Fussell et al. [68] use a continuous transition to move from an egocentric perspective to an exocentric one. Pausch et al. [221] allow users of their immersive virtual reality system to place a camera icon on the world-in-miniature (WIM) map of their environment to seamlessly transition to different viewpoints. Lee et al. [125] use fade-in and fade-out effects when switching between a user's view to another's to prevent the expert from feeling sick.

This chapter draws motivation from prior work to address the open challenge of integrating different perspectives, seamless viewpoint management, and assesses whether it reduces cognitive load and minimizes disruptions of engagement throughout the session or improves the overall collaboration experience for users. By addressing this challenge, Vicarious aims to streamline the selection of multiple viewpoints and determine if it improves the collaboration experience and effectiveness. Despite extensive research showing the benefits of providing viewpoints and having multiple viewpoint of the collaboration space, no prior work on automatic seamless selection and transitions between multiple viewpoints during collaboration was found. In Table 5.1, some previous work that is more similar is included and grouped based on input type, viewpoint selection method, annotation, and collaboration type. These categories were chosen based on related work, and more details are discussed in Section 5.3.

5.3 Vicarious

A high-level overview of the design of *Vicarious* is provided, highlighting the key components and outlining the main features and functionalities that enable context-aware viewpoint sharing and collaboration.

5.3.1 Design Overview

To facilitate efficient collaboration between local and remote users, the viewpoints of both users are captured and shared with each other. An asymmetric AR-VR collaboration setup is utilized, where the local user's physical space is live-streamed using a 360° camera. This

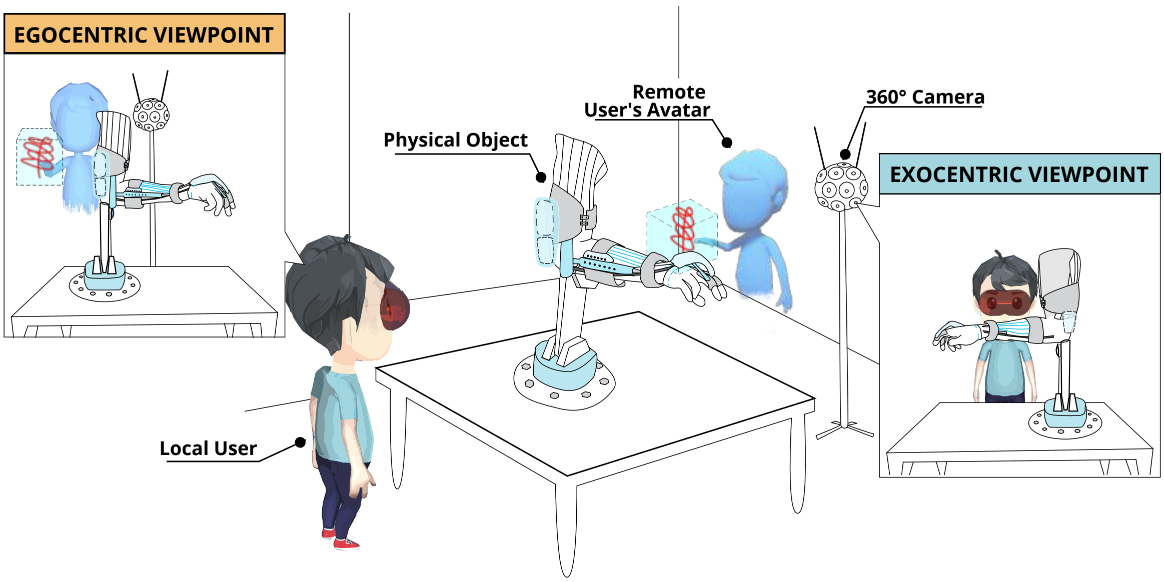


Figure 5.2: Schematic view of egocentric (left) and exocentric (right) viewpoint.

setup provides the remote user with a live panoramic view of the collaborative space, enabling them to offer assistance and support to the local user, even when they are not physically present. The local user utilizes an AR-HMD, while remote users use VR-HMDs. The local user's perspective is captured through the camera integrated into their AR-HMD, while the remote users' perspective is captured based on what is rendered on their VR-HMD viewports.

Therefore, users have the option to choose from two different ways of viewing the environment at any given time (see Figure 5.2):

- *Egocentric View*: Live video stream from the local user's AR camera is shared with the remote experts. The video stream is displayed in picture-in-picture (PiP) mode, which smoothly follows the user's field of view or can be anchored or pinned to a fixed location in the world space [148, 247, 125]. Users can interact with the PiP window by clicking on it (using either a VR controller or their hands in case of an AR user).

- *Exocentric View*: A live 360° camera, mounted in the local user's space, provides a panoramic view of the environment, which is then streamed to remote users. Remote users can view the video through a VR headset and have 3 degrees of freedom (3-DoF). Each VR user is represented as an avatar, and the movement of avatars is achieved by utilizing inverse kinematics from head and hand tracking.

Due to the limited field of view of the human peripheral vision and the constraints of AR-VR display technology, the remote user can only see a portion of the 360 spherical surface at any given time, depending on their viewport orientation (see Figure 5.3). Therefore,

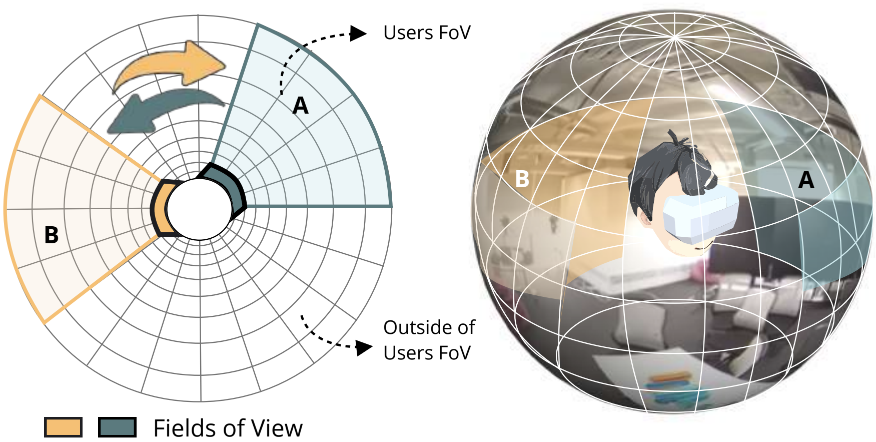


Figure 5.3: Illustration of user FoV. As shown on the sphere.

the FoV of the remote user is also captured and broadcast to other users. This allows users to see where the remote user is currently looking. By combining these different environment representations, incorporating both ego- and exocentric perspectives, a context-aware method is employed to determine the optimal viewpoint.

5.3.2 Context-aware Viewpoint Selection

With several viewpoints available to users at any given time, the goal is to identify the viewpoint that provides the most comprehensive information about the current state of the collaborative task. To determine the optimal viewpoint, the following criteria have been taken into account:

Contextual Information

First, from the list of ego- and exocentric viewpoints (see Figure 5.4a), saliency information is extracted and used as an indication of where the areas of interest are. This information is utilized as part of localizing attention. To achieve this, three attention cues are extracted: global contrast (K), optical flow (O), and Gaussian blur (G_σ).

Global contrast (K) is calculated as the sum of absolute color differences between adjacent pixels, weighted by a factor $f(i)$. This measure captures the overall contrast within the image and is used instead of local contrast for efficiency reasons. A histogram is constructed, and the difference for each color is calculated.

$$K(i) = \sum_{j=1}^{m} f(i)|C_i - C_j| \tag{5.1}$$

Here, C_i and C_j represent the color values of adjacent pixels, and $f(i)$ is a weighting factor. Optical flow (O) represents the change in motion within the frame, providing insight into dynamic elements within the scene. Gaussian blur (G_σ) is applied to smooth out details and emphasize larger-scale features, reducing noise and distractions. These attention cues are combined to create saliency maps, with each cue contributing to the final saliency value ($S(i)$) at each pixel. The combination is achieved through a weighted linear equation:

$$S(i) = w1 * K(i) + w2 * O(i) + w3 * G_\sigma(i) \tag{5.2}$$

Here, $w1$, $w2$, $w3$ represent the mapping weights for each attention cue. By adjusting these weights, saliency maps can be fine-tuned to prioritize certain visual features over others. Global contrast is utilized instead of local contrast for improved efficiency.

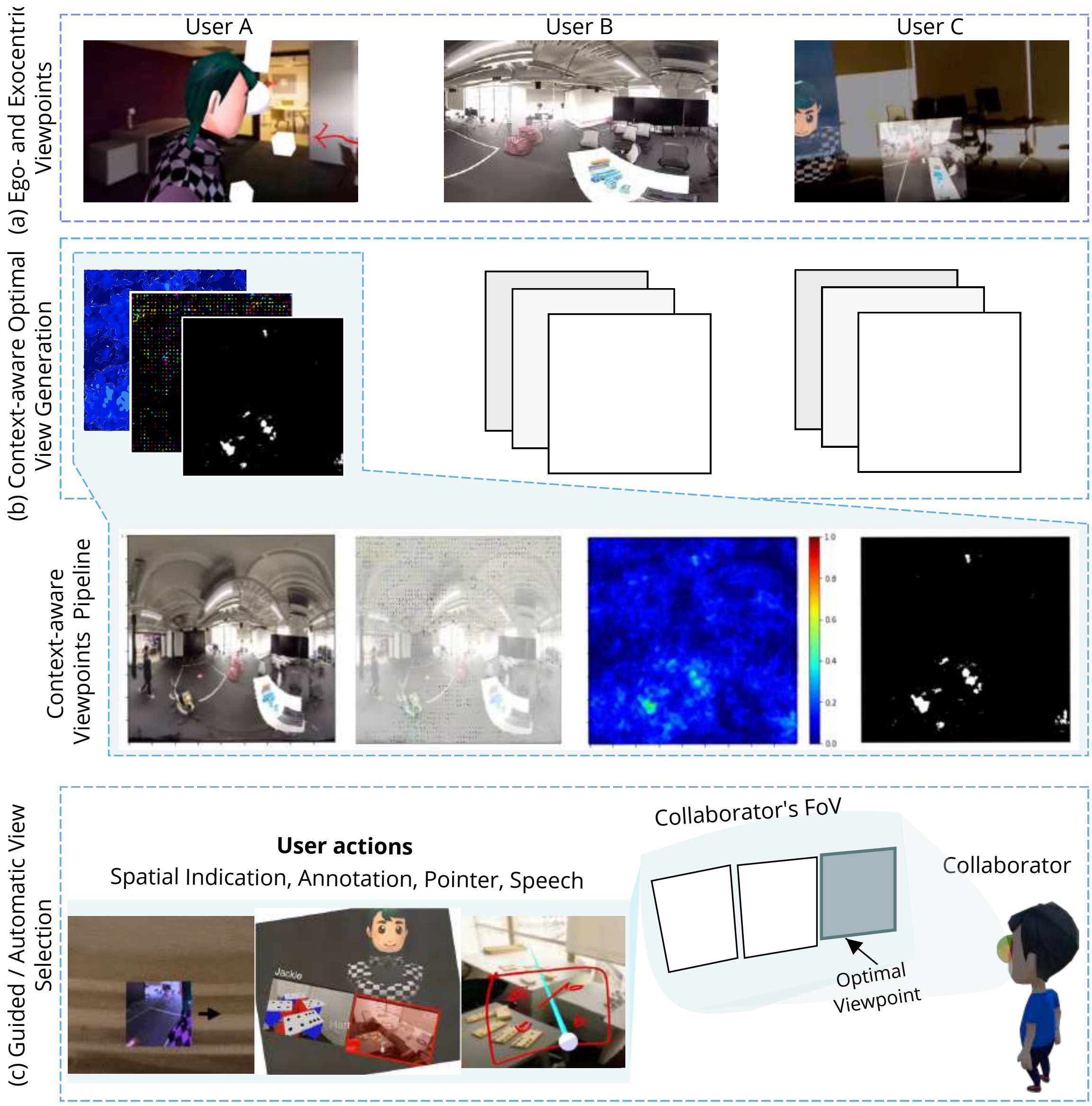


Figure 5.4: Pipeline of the pre-processing step. The pre-processing step first computes optical flow from an input 2D video, estimates saliency based on the optical flow, and finally, the binary threshold of the saliency map

The Lucas-Kanade method was employed to track scene feature points for detecting changes in the layers. When local changes occur, motion vectors are recorded for later registration in the corresponding saliency map of the layer, resulting in a set of motion vectors $O = \{o_i \,|\, i = 1...T\}$ from the stacked video frames, comprising T key frames $F = \{f_i \,|\, i = 1...T\}$. Next, the saliency maps are filled with a weight of 1 to the pixels corresponding to the registered feature points and facial regions for each layer. This step ensures that regions of interest identified by the Lucas-Kanade method are appropriately highlighted in the saliency maps. Furthermore, to account for the probability of salient object appearance decreasing with distance from the center of an image, a Gaussian function $G_\sigma(i)$ is applied to the saliency maps. This Gaussian function assigns higher saliency scores to pixels closer to the center of the image, effectively highlighting the most visually significant regions within the video (see Figure 5.4b).

Predefined Actions

While the saliency score is helpful in identifying visually prominent content, it is not the solely useful factor for determining the optimal viewpoint. Therefore, the user's actions or features they are using at any given time are also considered. In the application, a set of predefined functionalities or features is provided, which users can access through user interfaces. This user-centric approach emphasizes incorporating the users' actions and tasks into the determination of the optimal viewpoint.

Verbal cues. All users can utilize voice chat to communicate with each other. WebRTC's native AnalyserNode is utilized to detect voice activity in an audio stream and calculate the intensity. By calculating the root mean square (RMS) value, which reflects the average amplitude of the audio signal, an indicator of voice intensity is obtained. Periodically retrieving the time-domain data from the AnalyserNode allows for the calculation of the voice intensity metric in real-time and analysis of the audio stream for information about the voice intensity.

Visual cues. Users can point to any object using a 3D pointer, annotate, and perform gestures that are translated through avatars (see Figure 5.5). Changes in these user activities on the scene are tracked, and weight values are assigned accordingly. A pilot study was conducted with nine participants to fine-tune weight values for voice chat, annotating, and pointing actions in the user interface. Each group of three participants performed a Lego building task. This task was specifically chosen because it naturally encourages communication among participants (more details in Section 5.4.4). Participants assigned weights to each action based on perceived importance. Following the completion of the pilot study, the collected data were analyzed to derive consensus weight values for each action. These weight values provided a clear representation of the relative importance of each action as determined by the participants. Based on the experiments, it was found that voice chat, annotating, and pointing (ordered by weight from high to low) yielded better results.

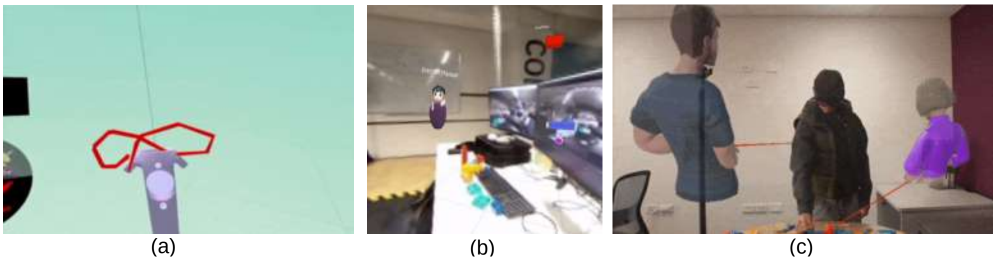


Figure 5.5: Visual cues: (a) annotate (b) gestures (c) 3D pointer.

A decision matrix is used to select the focus or automatically determine the viewpoint during collaboration based on the selected mode. Saliency scores and predefined action weights are assigned to each viewpoint. Weighted scores are calculated by multiplying the ratings with their corresponding weights, and the viewpoint with the highest weighted score is selected. By incorporating the predefined actions with weights determined from pilot studies (m_n, where $n = 1, 2, 3, ..$) alongside saliency scores and voice intensity, the determination of the optimal viewpoint takes into account not only visual prominence but also the dynamics of user interaction and communication. A sample decision matrix is shown in Table 5.2.

Table 5.2: Decision Matrix for Optimal Viewpoint Selection

Criteria	**Participant 1**	**Participant 2**	**Participant 3**
Predefined Actions	x_1m_1	x_2m_2	x_3m_3
Saliency ($S(i)$)	y_1	y_2	y_3
Voice Intensity	z_1	z_2	z_3

Multiple-Viewpoint Awareness

Users can select their preferred view and also have the option to toggle between different views to see alternative perspectives. When a user selects their viewpoint, the presence of other viewpoints is visually highlighted or indicated. This is done by displaying small thumbnail images or avatars of other users who have different viewpoints. Clicking on these thumbnails could provide a quick switch to that user's viewpoint. In Figure 5.4, user actions, saliency score, and active speaker are illustrated, which influence the view selection.

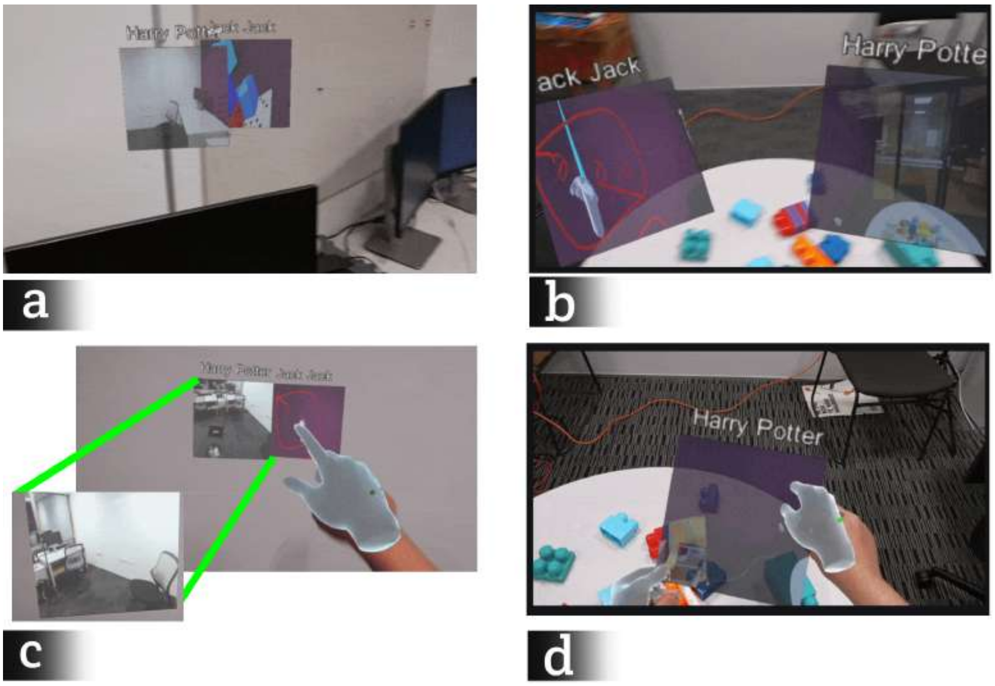


Figure 5.6: Visualization and Interaction options include: (a) Smooth follow (b) Pinned to a fixed location (c) Zoom in/out (d) Resize.

5.3.3 Visualization and Interaction

To visualize contextual information about important content outside of the user's current view, other users' points of view are overlaid on the main thumbnail view. Users can manually or automatically select them to bring them to the top.

Interaction Interface

To ensure smooth transitions between viewpoints, the video gradually fades in when switching to a new viewpoint, avoiding sudden changes and providing a seamless transition. Similarly, when moving away from a viewpoint, the video fades out gradually to prevent abrupt visual changes that could cause disorientation.

In addition, the Picture-in-Picture (PiP) window adjusts dynamically to match the user's head movements. It smoothly follows the user's field of view (Figure 5.6c), maintaining a consistent relative position within their visual perspective. This enables a natural and immersive viewing experience as the user looks around.

Alternatively, the PiP window can be anchored or pinned to a fixed location in the world space (Figure 5.6b). It remains stationary relative to the surrounding environment, regard-

Figure 5.7: Indication of spatial context: Arrow indicator

less of the user's head movements. This feature is useful when users need a consistent reference point or observe a specific area independent of the user's gaze.

Indication of Spatial Context

The direction and position of viewpoint thumbnails placed in world space are tracked. Since the user can position them anywhere, it leads to a disassociation between the avatar and their viewport. To address this issue, an arrow pointing toward the avatar of a remote user is utilized, allowing local users to identify which remote user the viewport belongs to. By following the arrow, local users can quickly establish a connection between the viewpoint thumbnails they are observing and the corresponding remote user. The placement of egocentric and exocentric viewpoint thumbnails remains unchanged. In addition to the arrow indicator (see Figure 5.7), nametags are also placed on top of each viewpoint thumbnail, further assisting in identifying and associating specific viewports with the corresponding users.

5.3.4 Networking

Real-time audio and video communication are achieved using WebRTC within Unity3D, which acts as a client connecting to a Node.js web server through WebSocket. The web server acts as a central hub, facilitating signaling and negotiation between peers. To ensure direct connections between peers, TURN and STUN servers are utilized for NAT traversal. Each peer has dedicated video and audio streams for transmitting media, while non-media data exchange such as viewpoint synchronization, position, orientation, and annotations is handled through WebRTC Data Channels.

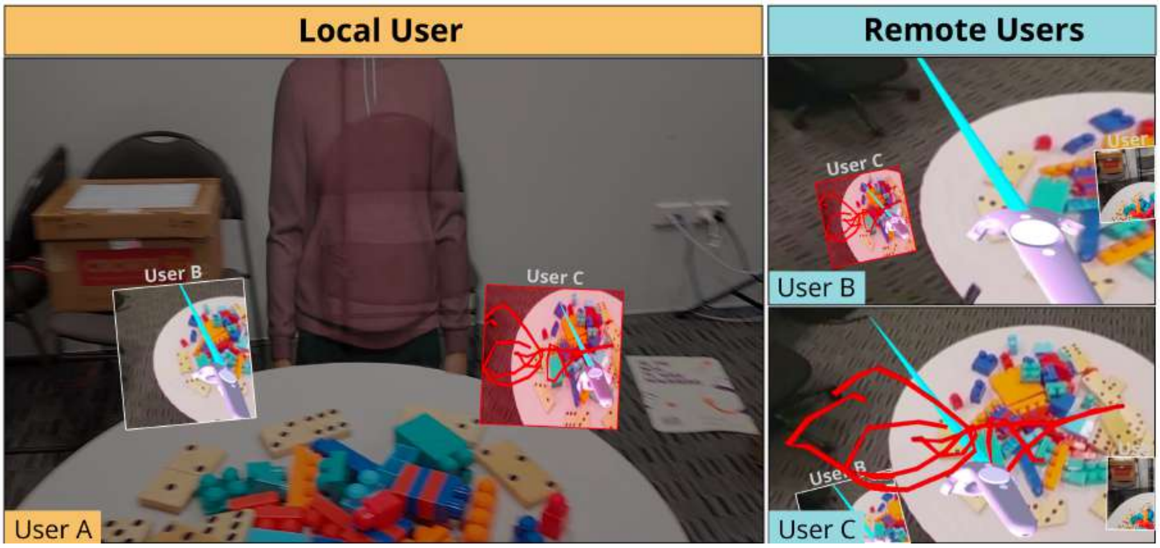


Figure 5.8: Viewpoint perspectives of local and remote users, highlighting remote user C's viewpoint in the GS condition.

5.4 User Study

A user study was conducted to investigate the effectiveness of context-aware viewpoint selection as described in **Section 5.3**. The study employed a 4×2 mixed factorial design. The within-subjects variable, *Viewpoint Condition*, consisted of four levels (No-view, Manual, Guided, and Auto), while the between-subjects variable, *Participants Role*, consisted of two levels (Local and Remote user) (Figure 5.8).

5.4.1 Hypotheses

Going back to the original research questions 4 and 5 outlined in Chapter 1, the study aimed to address various aspects of mixed-perspective collaboration by deriving several hypotheses from these research questions. The following breakdown outlines these two **RQ**s and the corresponding hypotheses:

To address **RQ4** regarding how users' viewpoints can be managed effectively, four different variations of viewpoint selection methods were presented: No-view Selection, Manual Selection, Guided Selection, and Automatic Selection to compare them and determine which is more effective to support collaborative tasks.

To address **RQ5**, which aimed to find out how multi-perspective collaboration can be effectively achieved for supporting collaborative tasks, a user study was conducted to assess the impact of the context-aware viewpoint selection method on collaboration performance and user experience. For this user study, to investigate **RQ5**, it was broken down into the following two sub-research questions:

RQ5.1 Would context-aware viewpoint selection increase the sense of presence in remote collaboration?

RQ5.2 Would context-aware viewpoint selection improve the task performance compared to using no-view selection?

Based on these research questions, **RQ5.1** and **RQ5.2**, the following hypotheses were formulated:

H1 Guided viewpoint selection will result in a higher level of spatial presence and social presence compared to other conditions.

H2 Both guided and automatic viewpoint selection will lead to reduced task completion time compared to the no-view and manual selection conditions.

H3 Usability will be significantly higher in the guided and automatic viewpoint selection conditions.

H4 Participants will prefer guided viewpoint selection more than other conditions.

5.4.2 Participants

Twenty-seven participants (21 identified as male, 6 as female) aged 18 to 55 years (M = 28.13, SD = 7.32) were recruited through advertisements and local university and community center flyers. All participants had normal or corrected normal vision using glasses or contact lenses. The sample was diverse in terms of ethnicity, with 12 participants identifying as European, 6 as Asian, 5 as mixed race, 3 as Latino/Hispanic, and 1 participant identifying as Pacific Islander. The majority of participants (n = 18) reported having no prior experience with AR/VR technology, while a few (n = 9) had used it for an average of 10 hours (SD = 5.35) in the past year.

5.4.3 Experimental Conditions

Altogether, four experimental conditions were utilized, as follows:

1. *No-view Selection (NS)*: is a stripped down version of *Vicarious* with all the view selection tools removed. Instead, users have full control over their viewing perspective without interference.

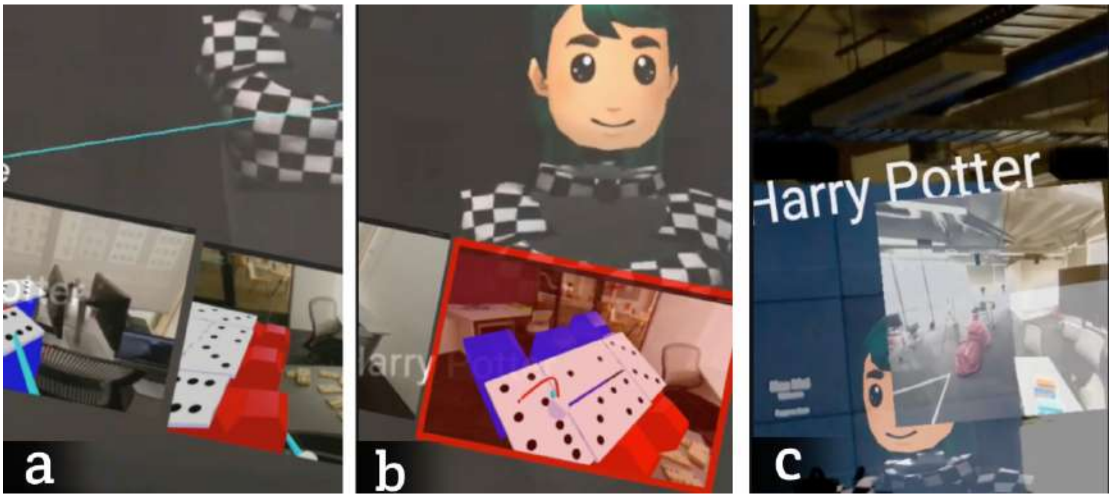
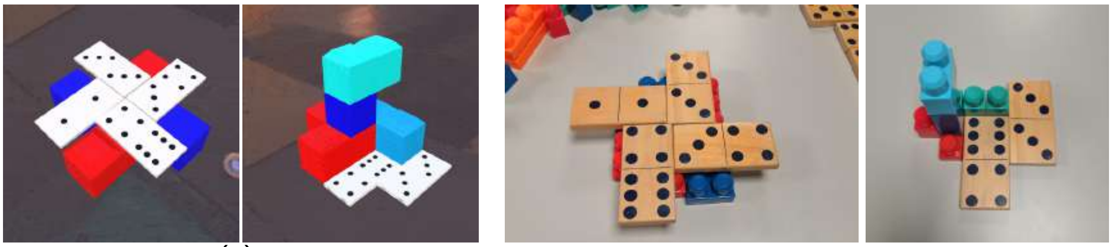


Figure 5.9: Experimental conditions: (a) Manual Selection (b) Guided Selection (c) Automatic Selection.

2. *Manual Selection (MS)*: allows users to manually choose between ego- or exocentric viewpoints by clicking or moving PiP windows within their field of view. The selected view gradually fills the user's field of view and can be reverted by clicking anywhere on the window (see Figure 5.9a).
3. *Guided Selection (GS)*: refers to the viewpoint selection technique that visually highlights the optimal view from the FoV list to prompt the local user to manually select the optimal viewpoint for the remote user's actions (see Figure 5.9b).
4. *Automatic Selection (AS)*: an ego- or exocentric view is automatically selected as the main view, which the system considers to be the optimal viewpoint representing user actions (see Figure 5.9c).

5.4.4 Experimental Tasks and Setup

A collaborative task was designed involving multiple remote users (using VR-HMD) guiding a local user (using AR-HMD) in building an assigned model using Legos and dominoes, (see Figure 5.10). The local user is in the task space where Legos and dominoes are randomly placed on a table. A 360° camera livestreams the task space to the remote users, who are located in a separate room. Both remote users have access to a visual representation of the desired final model and can provide instructions or guidance to the local user in locating specific Legos and dominoes in the task space. They can communicate steps to the local user through speech description, gesture pointing, and/or 3D annotation.



(a) (b)

Figure 5.10: Lego and domino blocks structures: (a) virtual replicas, (b) corresponding structures made using physical blocks.

The local user finds the objects and builds the model using the physical Legos or dominoes based on the guidance instructions received from the remote users. Different model structures were used for each condition, and the order and combinations of tasks and conditions were counterbalanced to reduce bias. Depending on the experiment condition, participants had access to either ego- and/or exocentric viewpoints.

5.4.5 Procedure

At the start of the study, participants signed a consent form and provided demographic information, including any prior experience with AR/VR. After a brief overview of the experiment, participants were divided into groups of three, with two participants acting as remote experts and one as the local user constructing the model. Then they were given about 5 minutes to acquaint themselves with the system.

Participants then proceeded to complete four conditions sequentially, with a 5-minute break in between. Each condition lasts approximately 10-12 minutes. After completing each condition, the participants were given 5 minutes to complete subjective questionnaires. The task concluded when the local user completed building the model. After completing the study, participants were given a post-study questionnaire and asked to rate their preferences across four conditions based on different criteria. On average, the study lasted slightly over an hour.

5.4.6 Measures

Users' activities, including voice chat and completion time, were logged throughout the experiment. To measure spatial presence, the *iGroup Presence Questionnaire* (IPQ) [204] was utilized, which features a 7-point Likert scale with four subscales: general presence (GP), realism (RL), involvement (INV), and spatial presence (SP). For social presence, a questionnaire was compiled with a 7-point Likert scale, including questions from various subscales: co-presence (CP) (from Bail [14] and Hauber [88]), as well as mutual attention (MA), mutual understanding (MU), and behavioral engagement (BE) (from the *NMM Social Presence Questionnaire* [86]). Workload was measured using the *NASA-TLX*[87], and system usability with the System Usability Scale (SUS)[35]. Motion sickness was evaluated using the *Simulator Sickness Questionnaire* (SSQ) [108]. The post-study questionnaire included measuring participants' preferences across various categories, which also included qualitative feedback through open-ended questions. Open-ended questions were asked to find out more details about participants' motivation and reasons behind their preference choice of conditions. The qualitative feedback provided by participants was analyzed using a thematic analysis approach. For example, the question "Which condition did you find easiest to use?" elicited responses explaining why participants found a particular condition the easiest to use. Analysis of responses involved coding based on specific features or aspects of the condition that participants found intuitive, user-friendly, or efficient. Themes included simplicity of interface, clarity of instructions, ease of navigation, etc. Each session was also recorded to observe interactions between users, noting verbal and non-verbal cues, turn-taking behavior, interruptions, and overall flow of communication.

5.5 Results

The Shapiro-Wilk test was conducted on the residuals to check for normality. Repeated measures ANOVA ($\alpha = 0.05$) with Tukey's HSD post hoc test was performed for normally distributed data, and repeated measures ANOVA with the Aligned Rank Transform (ART) [258] was used otherwise. Holm-Bonferroni correction was used for all post hoc tests.

5.5.1 Social Presence

Figure 5.11 shows the average Social Presence (SP) score of view selection conditions. Participants in the GS condition gave a significantly higher rating on the SP scale overall (Local: M = 5.78, SD = .21; Remote: M = 5.26, SD = .14) than those in the NS, MS, and AS. There was a significant main effect of the conditions (BE: $F_{3,183} = 3.11$, $p = .027$; CP: $F_{3,399} = 7.52$, $p < .001$; MA: $F_{3,291} = 14.06$, $p < .001$; MU: $F_{3,183} = 10.47$, $p < .001$), indicating that conditions had a statistically significant impact on the dependent variable. However, the main effect of the role was not statistically significant (BE: $F_{1,25} = 1.52$, $p = .228$; CP: $F_{1,25} = 2.16$, $p = .153$; MA: $F_{1,25} = 0.02$, $p = .885$; MU: $F_{1,25} = 0.17$, $p = .680$). Additionally, the interaction between conditions and role was not statistically significant (BE: $F_{3,183} = 0.46$, $p = .709$; MA:

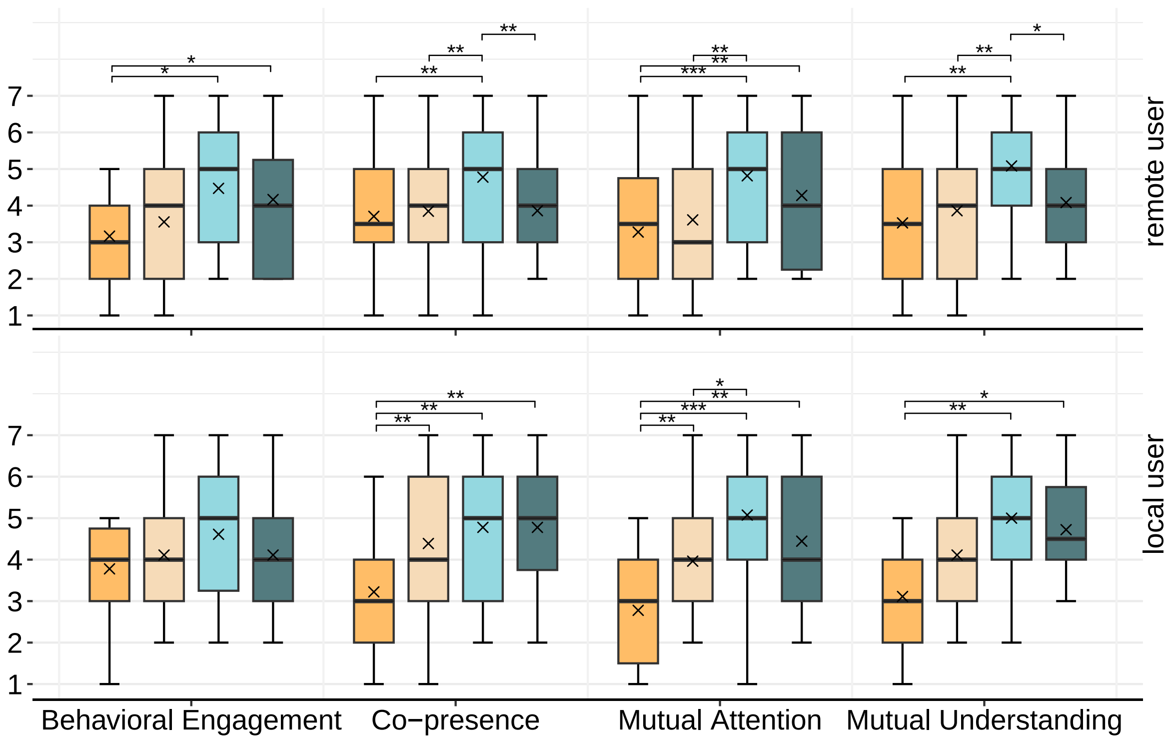


Figure 5.11: Social Presence results in each subscale (*=p<.05, **=p<.01, ***=p<.001). NS MS GS AS

$F_{3,291}$ = 1.08, p = .356; MU: $F_{3,183}$ = 0.86, p = .465) except for CP: $F_{3,399}$ = 3.62, p = .013. Post hoc pairwise comparisons suggest that participants in the GS condition felt significantly higher social presence compared to the NS (BE: p = .017; CP: p < .001; MA: p < .001; MU: p < .001) and MS (CP: p = .018; MA: p = .0002, MU: p = .001).

5.5.2 Spatial Presence

There was a significant difference in overall Spatial Presence (SP) scores between conditions, $F_{3,1076}$ = 9.53, p < .001. Post-hoc tests revealed significant differences between GS vs. AS (p = .001), GS vs. NS (p = .001), and GS vs. MS (p < .001). These findings suggest that the GS condition is associated with higher SP scores compared to the AS, NS, and MS conditions. Figure 5.12 shows the average Spatial Presence (SP) score of view selection conditions. There was a significant main effect of the conditions for GP: $F_{3,75}$ = 0.97, p = .414; INV: $F_{3,183}$ = 5.99, p = .001; RL: $F_{3,183}$ = 5.89, p < .001), with marginal main effect for SP: $F_{3,507}$ = 2.57, p = .053 indicating that conditions had a statistically significant impact on the dependent variable. However, the main effect of the role was not statisti-

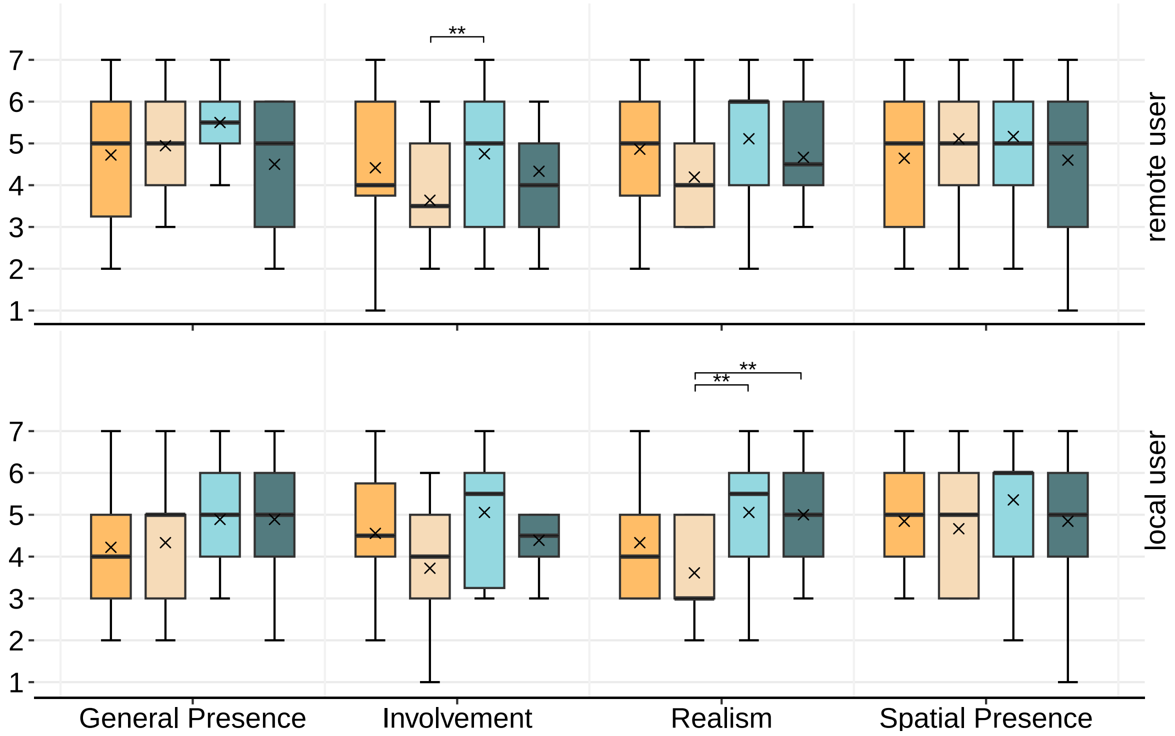


Figure 5.12: Spatial Presence results in each subscale (*=p<.05, **=p<.01, ***=p<.001). NS MS GS AS

cally significant (GP: $F_{1,25}$ = 0.98, p = .331; INV: $F_{1,25}$ = 0.83, p = .370; RL: $F_{1,25}$ = 1.45, p = .239; SP: $F_{1,25}$ = 0.0009, p = .999). Additionally, the interaction between conditions and role was not statistically significant (GP: $F_{3,75}$ = 0.58, p = .629; INV: $F_{3,183}$ = 0.18, p = .904; RL: $F_{3,183}$ = 1.28, p = .282; SP: $F_{3,507}$ = 1.56, p = .198). Participants in the GS condition gave a significantly higher rating on the SP scale overall (Local: M = 6.21, SD = 1.49; Remote: M = 6.08, SD = 1.38) than those in the NS, MS, and AS. Post hoc pairwise comparisons suggest that participants in the GS condition felt significantly higher spatial presence compared to the NS (SP: p = .053), MS (INV: $p < .001$, RL: $p < .001$), and AS (RL: p = .043).

5.5.3 System Usability

SUS scores for both VR and AR fell within an acceptable range (see Figure 5.13). The VR role received an average rating (M = 68.71, SD = 22.17), while the AR role received an "ok" rating (M = 60.13, SD = 23.63). These findings suggest that there were no significant differences in usability scores between VR and AR and that the different conditions did not significantly affect the overall SUS scores. A two-way ANOVA results indicated that there

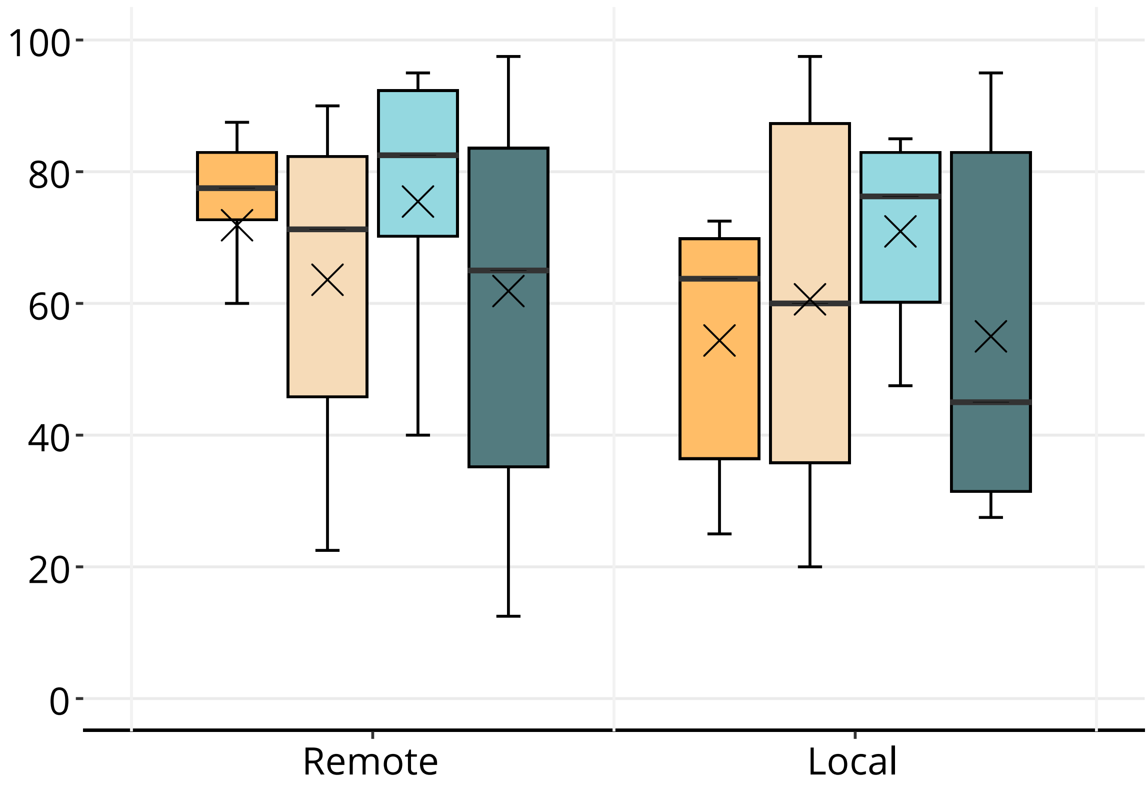


Figure 5.13: System usability results (80.3 or higher is considered good, 68 and above classified as average, and below 51 considered poor). NS MS GS AS

was no significant main effect of role ($F_{1,25}$ = 4.04, p = .053), suggesting that the SUS scores in VR (M = 68.71, SD = 22.17) did not differ significantly from AR (M = 60.14, SD = 23.63). Similarly, the main effect of conditions was not significant ($F_{3,75}$ = .98, p = .404), and thus failed to show a significant difference between conditions. The interaction between role and conditions was also not significant ($F_{3,75}$ = .79, p = .497), suggesting that the relationship between role and conditions did not significantly influence the SUS scores.

5.5.4 Completion Time and Task Load

The completion time for searching Lego blocks shown in Figure 5.14, did not differ significantly among the conditions ($F_{3,140}$ = .07, p = .973). Bartlett's test of sphericity was not statistically significant (χ^2 = 2.97, p = .395). For assembling Lego blocks, there was a significant difference in completion time among the conditions ($F_{3,140}$ = 8.96, p < .001). Pairwise comparisons showed significant differences between the *GS* and both *MS* (p = .001) and *NS* (p < .001) conditions. In total time for all Lego blocks, there was a significant difference

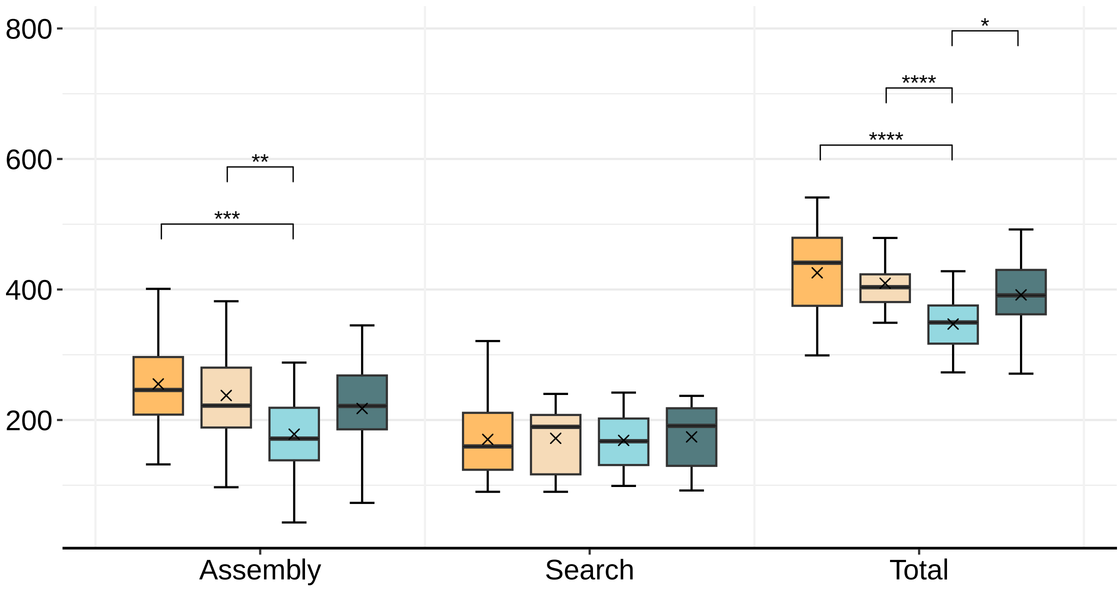


Figure 5.14: Average task completion time for each condition NS MS GS AS

among the conditions ($F_{3,140}$ = 14.63, $p < .001$). The assumption of equal variances was reasonable (Bartlett's χ^2 = 7.55, p = .056). Pairwise comparisons showed significant differences between the *GS* condition and the other conditions (*AS*, p = .002; *MS* and *NS*, $p < .001$).

Furthermore, the six subscales of the NASA-TLX were summed up with their weights to obtain the overall NASA-TLX score (see Figure 5.15). The overall NASA-TLX score met the assumption of homogeneity of variances and indicated a significant effect on the overall NASA-TLX score ($F_{2,18}$ = 5.70, p = .010). Results of the posthoc test indicate that the overall NASA-TLX score in the GS condition (M = 35.6, SD = 17.94) was significantly lower than that in the other conditions.

5.5.5 Simulator Sickness

Figure 5.16 shows the average score of SSQ questionnaire [108], with 16 items rated from 0: none - 3: severe, then calculated the three subscales (nausea, oculomotor, and disorientation) and the total score. The SSQ was administered pre-experiment and post-experiment for each task in each condition. The results indicate very low simulator sickness scores for MS conditions. SSQ values for the NS_{post} (M = 4.71, SD = 4.41) and AS_{post} (M = 3.76, SD = 3.84) do not show significant differences between SSQ values for the NS_{pre} (M = 4.57, SD = 6.62) and AS_{pre} (M = 4.22, SD = 5.72), indicating that the different viewpoint selection conditions may not have induced simulator sickness. A Wilcoxon signed-rank test was applied to examine differences between pre and post SSQ scores for participants exposed across vari-

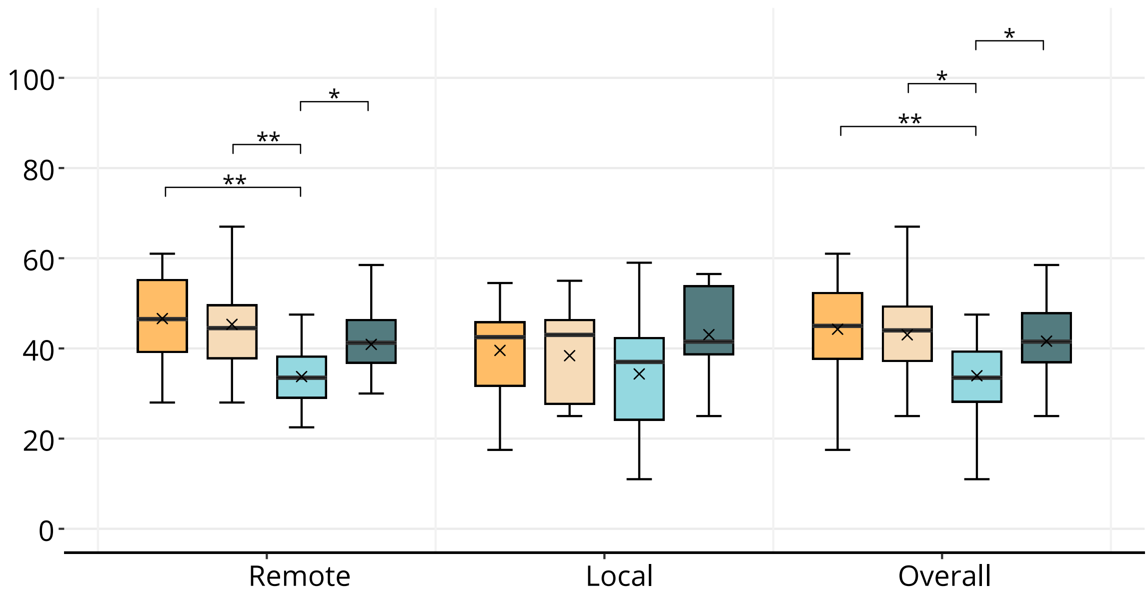


Figure 5.15: NASA-TLX score (0: very low to 100: very high), overall score is shown to the right. NS MS GS AS

ous conditions. Among AR users, no significant difference was found between pre and post scores for all four conditions. However, VR participants demonstrated a marginally significant difference in pre and post SSQ scores for both the MS condition (W = 8.5, p = 0.058, r = 2.00) and the AS condition (W = 58.5, p = 0.091, r = 13.79), suggesting potential effects of the interventions that merit additional investigation. The MS condition, where users manually select viewpoints, may involve greater cognitive and physical engagement compared to the AS condition, where the system automatically selects viewpoints. This difference may indicate varying levels of comfort or immersion between manual and automatic viewpoint selection methods. These findings hint at possible differences in user experience and adaptation to the VR environment between manual and automated selection methods. Further exploration of these trends through larger sample sizes or additional measures may provide insights into the effectiveness of interventions.

5.5.6 Preferences

Figure 5.17 shows the responses to the user preference questionnaire regarding six different aspects of collaboration for both local and remote users across all four conditions. In general, the majority of participants showed a preference for the GS condition (48%), followed by the MS condition (27%). A chi-square goodness of fit test was used to determine whether the four conditions (NS, MS, GS, AS) were equally preferred and the results show prefer-

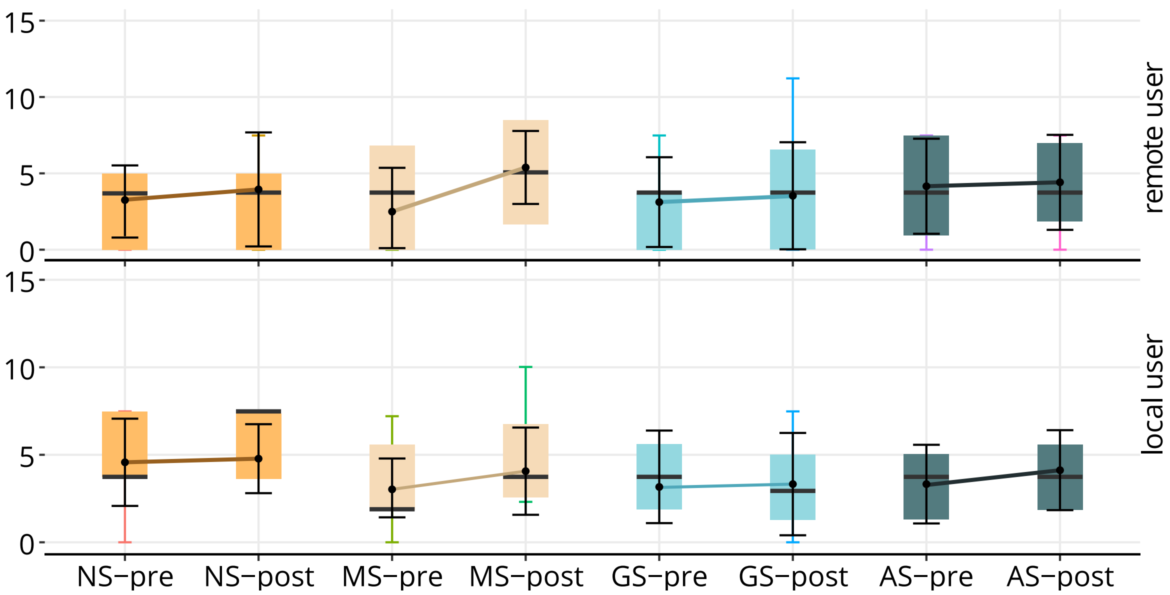


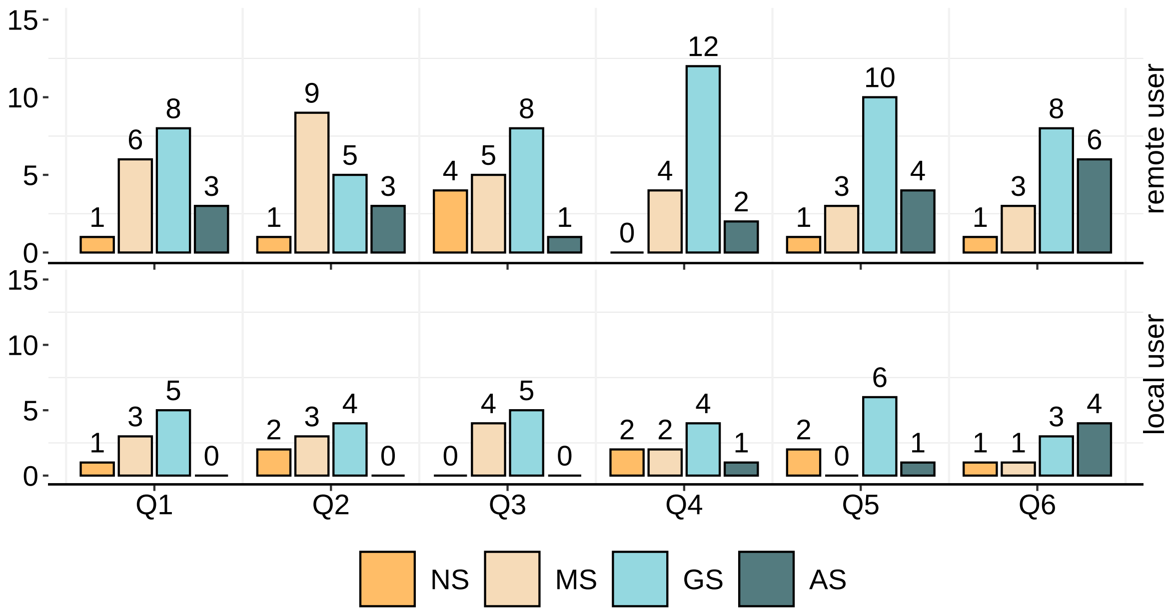
Figure 5.16: Increase in the simulator sickness score (pre- and post-exposure)

ence was not equally distributed ($\chi^2(3) = 55.63$, $p < .001$). For local users, the MS condition and for remote users, the GS condition was reported as the most user-friendly in terms of ease of use (Q1) $\chi^2(3) = 11.96$, $p = .007$. Regarding task environment understanding (Q2) $\chi^2(3) = 9$, $p = .029$ and task completion (Q5) $\chi^2(3) = 17.29$, $p < .001$, participants found both the GS and MS conditions to be most helpful. Additionally, the GS conditions were perceived as the most effective for understanding the partner's instructions (Q3) $\chi^2(3) = 12.55$, $p = .005$ and communicating effectively with partners (Q4) $\chi^2(3) = 18.18$, $p < .001$. Finally, both local and remote users reported GS as the most preferred condition (Q6) $\chi^2(3) = 8.70$, $p < .003$, while remote users specifically preferred both AS and GS. Chi-Square Test of Independence to determine any significant association between the participants' roles (VR vs. AR) and their preferences for specific conditions shows no significance.

5.6 Discussion

The user study results were overall positive for conditions associated with optimal viewpoint selection (GS and AS), suggesting that Vicarious had a positive effect on improving collaboration in MR remote collaboration compared to no-view selection.

The first research question (**RQ5.1**) aimed to determine whether context-aware viewpoint selection would increase the sense of presence in remote collaboration. The experimental results showed that participants in the GS condition had higher social presence



Q1	Which condition did you find easiest to use?
Q2	Which condition did you find most helpful for understanding the task environment?
Q3	Which condition did you find most helpful for understanding your partner's instructions?
Q4	Which condition did you find most effective for communicating with your partner?
Q5	Which condition did you find most helpful for completing the task?
Q6	Which condition did you prefer the most?

Figure 5.17: User preference among the four conditions.

scores in all subscales compared to the other conditions, thus supporting the first part of hypothesis (**H1**) that GS would lead to a higher level of social presence compared to other conditions. Participants found GS's highlighted features effective for understanding instructions, while NS lacked viewpoints and MS/AS obstructed the view, resulting in lower presence scores. One participant feedback highlighted the impact of GS on collaboration:

> *"It felt like we were in the same room. By looking at the viewport, I could easily see what they were drawing and looking at, which made it easier."* [G4/P12]

Results from the spatial presence questionnaire showed that the overall IPQ score was above the midpoint, although no statistical significance was found between role and interaction effects in General Presence and the Spatial Presence subscale. However, there was a significant difference in overall Spatial Presence (SP) scores between conditions, particularly with the GS condition associated with higher SP scores compared to the AS, NS, and MS conditions. Therefore, the second part of **H1** that GS would lead to a higher level of social presence compared to other conditions was also supported. Thus, **RQ5.1** was supported by the questionnaire results, revealing significant differences in presence in conditions associated with context-aware viewpoint selection.

The second research question (**RQ5.2**) aimed to assess whether Vicarious would improve task performance compared to using no-view selection. The findings suggest that although GS resulted in faster task completion times, AS took longer for assembly and searching tasks. This increase in time for AS may be attributed to the mode's potential distraction of the user's attention while automatically adjusting the viewpoint, resulting in a steeper learning curve and increased task load (TLX: M = 41.61, SD = 13.63). Therefore, the second hypothesis (**H2**) that GS and AS would reduce task completion time compared to NS and MS was not fully supported.

Similarly, the study partially supported **H3**, which proposed that usability would be highest in the guided and automatic viewpoint selection conditions. Although GS had a slightly higher usability score than AS, the difference was not statistically significant.

Furthermore, **H4**, which proposed that participants would exhibit a stronger preference for guided viewpoint selection, was supported. However, the study also revealed a similar preference for automatic selection. This suggests that both guided and automatic selection helped participants navigate the remote environment effectively and become aware of other users' activities. One participant feedback supports the findings:

> *"The tool (GS) was spot-on, and the mini viewport window helped me understand exactly what my partner was seeing."* [G3/P8]

These findings further support the research question **RQ5.2** that having a context-aware viewpoint selection option improves task performance compared to using no-view selection.

5.6.1 Open-ended Feedback

The analysis of participants' responses to the open-ended questions revealed several key themes regarding their preferences and experiences with the different conditions. In terms of ease of use, participants commonly cited the simplicity of interface and clarity of instructions as factors that contributed to their preference for a particular condition. For example, one participant stated,

> *"I found highlighted one (GS) the easiest to use because the interface was straightforward and intuitive. Using the little windows to switch views was easy. I could jump around different views without any trouble. "* [G2/P5]

Regarding understanding of the task environment, participants frequently mentioned features such as visual clarity and spatial orientation as important aspects of the condition that aided their comprehension. As one participant noted,

> *"That small screen (mini window) displayed a close-up view of the task area alongside my main view. So I was able to focus on specific details the other guy was saying, like finding LEGO pieces, without losing sight of the overall workspace."* [G6/P18]

> *"It was practical because it provided a detailed and magnified view. It made it easier to complete the build."* [G5/P14]

When it came to understanding their partner's instructions, participants emphasized the importance of clear audio communication and real-time feedback. One participant commented,

> *"I really liked how clear the audio was when we were working. I could understand what the person in the room was saying."* [G8/P22]

Participants expressed a strong preference for GS due to its intuitive nature. As one participant noted,

> *"I really liked the guided feature and how it highlighted. This made it much easier to understand where my attention should be."* [G1/P2]

In contrast, AS was met with mixed reviews. While participants acknowledged its convenience, concerns were raised regarding instances where AS interfered with their activities.

> *"Automatic selection was okay overall, but there were times when I felt it interfered with my activities or chose the wrong view. That was a bit distracting."* [G8/P24]

The absence of any view selection mechanism (no selection) resulted in frustration among participants. Without the ability to adjust viewpoints, participants experienced challenges in coordinating with collaborators. One participant expressed,

> *"Having no view selection at all was frustrating. I felt lost. It made the whole process more challenging."* [G7/P19]

Lastly, manual selection was perceived as labor-intensive and inefficient compared to guided selection. Participants found manual selection to be time-consuming. As noted by one participant,

> *"Manual selection needed too much work. I often found it difficult to locate a well-placed spot to position those floating windows. Being a tidy person, I felt like I was doing unnecessary work compared to the guided selection."* [G6/P16]

5.6.2 Communication Patterns

The remote users within the same group took turns speaking, alternating between the remote guide providing instructions and the physical builder seeking clarification or requesting assistance. The local user provided feedback to the remote users regarding the progress, challenges, and questions related to the Lego blocks, and the remote guide responded with appropriate instructions and guidance. The remote users extensively used spatial deictic references to indicate specific locations on the Lego blocks. For example, they employed terms like "over there," "to the left," or "next to the blue brick" to provide precise instructions or indicate a point of reference. During this time, the viewpoint window was particularly observed as being used most. Per session, the system switched to the optimal viewpoints an average of 11.89 times during the AS condition (SD = 2.55). Remote users used their 3D pointer extensively, while the local user focused on the avatar and the viewing window of the remote user to find a reference. Additionally, during those instances, local users adjusted the PiP window with their hands. On the other hand, the local user utilized demonstrative pronouns (e.g., "this," "that") to refer to specific Lego blocks, picking them up to show the remote user. By employing these pronouns, they established a shared understanding of the objects being discussed. For the remote user's viewpoint, it was most used during that time, which helped the remote user understand which piece the local user was referring to from an egocentric view rather than an exocentric view, although this wasn't uniformly the case. In situations requiring time-based instructions, partners used temporal deictic language. For instance, they might say, "Wait for a moment," "Place the brick after the blue one," or "Build this section first."

5.6.3 Design Implications

Several design implications emerged from the user study, which can be significant for future MR remote collaboration systems:

1. *Flexibility in Viewpoint Selection:* Providing users with the flexibility to choose between different viewpoint selection conditions can enhance user satisfaction and adaptability to their specific preferences and needs. By offering a range of options, such as manual selection, guided selection, automatic selection, and no-view selection, systems cater to different user preferences and task requirements.

2. *Guided Selection for Optimal Viewpoints:* The results indicate that the Guided Selection (GS) condition was preferred and performed well in various aspects. This suggests that incorporating audio and visual cues and highlighting the optimal viewpoints can assist users in understanding the task environment, communicating effectively with partners, and completing tasks. Designing intuitive and informative cues for guiding users toward optimal viewpoints can enhance the overall experience and task performance.

3. *Consideration of User Context:* The study involved participants who were both remotely located and physically present, representing different contexts of collaboration. These findings highlight the importance of considering user context when designing viewpoint selection mechanisms. Also, user-specified preferences for predefined actions and adding more dynamic user action recognition would be helpful.

4. *Usability and Workload Considerations:* The usability scores and NASA-TLX workload results provide insights into the user experience and cognitive load associated with each condition. Manually selecting the viewpoint adds additional workload and requires learning or getting used to, while automatic selection may require an initial understanding of its functionality.

5.6.4 Limitation and Future Work

In the study, a 360° camera was used, which is essentially a 2D representation. However, future research could explore integrating depth information or utilizing teleoperating robots to introduce additional viewpoints. Future work should also refine the user actions triggering guided and automatic selection. This can include incorporating additional cues such as gaze tracking and view frustum analysis, which has demonstrated benefits in the extensive literature on remote collaboration [231]. Leveraging these cues can provide users with more intuitive and accurate guidance, leading to improved selection of viewpoints and overall system performance. Additionally, advanced machine learning algorithms can be explored for scene feature extraction and gesture recognition. It is worth noting that the current implementation prioritized real-time operation, which was crucial for collaboration, further experiments are needed to evaluate the effectiveness of the pipeline and identify potential enhancements. Decoupling the head-tracked viewpoint could offer advantages for external users, allowing more natural navigation and aligning visual perspectives with camera angles [226]. Instructors can lead learners through different viewpoints to facilitate remote mentoring and ensure precise guidance [178, 133]. Instructors can guide learners through different viewpoints for effective remote mentoring [178, 133]. To enhance and optimize the Vicarious system for scenarios involving more than one-to-many remote collaboration, it is imperative to expand the user study, encompassing larger teams and diverse contexts. The integration of advanced cues and interaction techniques will play a pivotal role in improving the overall effectiveness and user experience in collaborative tasks. Future user studies can also aim for larger and more diverse participant samples. Despite the small group size, this study provided valuable initial insights into the usability and preferences of viewpoint selection conditions for remote collaboration, indicating how these conditions may manifest for larger teams.

5.7 Conclusion

This chapter presented *Vicarious*, a novel context-aware viewpoint-sharing method is presented that manage users' viewpoints effectively and support various collaborative tasks in the mixed reality collaboration space. This method addresses the challenge of dynamically selecting optimal viewpoints for sharing among multiple users. To answer **RQ4** about how users' viewpoints can be managed effectively, four different variations of viewpoint selection methods were presented: No-view Selection, Manual Selection, Guided Selection, and Automatic Selection. Among these methods, Guided Selection and Automatic Selection leverage contextual information, including user actions, object interactions, and scene context. Also to find out how multi-perspective collaboration can be effectively achieved for supporting collaborative tasks (**RQ5**), a user study (n = 27) was conducted to assess the impact of the context-aware viewpoint selection method on collaboration performance and user experience. Vicarious employed an asymmetric AR-VR setup to explore collaboration under four distinct conditions: No-view, Manual, Guided, and Automatic selection.

The outcomes from the user study indicated that *Vicarious* improved users' task space understanding and task performance while reducing cognitive load. Guided Selection (GS) was the most preferred, performing well across multiple metrics such as user preference, system usability, understanding of task space, and task completion. Automatic Selection (AS), was particularly favored in the AR context (local users) whereas VR users had a split preference between the GS and AS conditions. The lower scores associated with No-view Selection (NS) across all metrics underscore the necessity for the viewpoint selection methods.

Overall, the findings of this study underscore the effectiveness of the *Vicarious* method in simplifying and automating viewpoint selection for remote collaboration tasks. To advance the understanding of viewpoint selection mechanisms and their influence on collaborative efforts, future research should explore and compare various attributes. This could include investigating the impact of different environments, team sizes, and task complexities on the effectiveness of viewpoint selection methods.

In summary, the findings highlight that optimal viewpoint selection improved users' task space understanding, task performance, and reduced cognitive load. The results also showed that users preferred the intervention, and their feedback provided valuable insights and recommendations for design implications, guiding future research endeavors in this domain.

Why work remotely when you can work Vicariously?

Chapter 6

Conclusion

The goal of this thesis was to enable multiple remote users to collaborate in a shared physical space alongside local users, regardless of their physical location, with a focus on achieving a high sense of presence. To achieve this goal, the thesis designed, developed, and evaluated a novel multi-user asymmetric MR system presented in **Chapter 3** and **Chapter 4**. The results suggest that the system effectively bridged the physical gap between remote and local users, thus enabling seamless collaboration with an immersive and induced high sense of presence. In addition, the thesis presented a method to manage multiuser viewpoints and applied it to mixed perspectives collaboration scenarios. This was achieved through the implementation of Vicarious, as presented in **Chapter 5**, which simplified the process of selecting and sharing multiple viewpoints within the shared MR environment. Vicarious was found to improve task space understanding and performance, reduce cognitive load, and be preferred by users. Furthermore, the thesis explored several design implications and provided a comprehensive understanding of the collaborative potential within multi-user MR scenarios. Additionally, certain limitations of multiuser MR have been identified, leading to recommendations for future research.

This final chapter provides a summary of the research conducted in this thesis. It discusses the significant contributions made to the field of multi-user MR experiences, returns to the original research questions, highlights the findings through the research carried out, and evaluates the accomplishments. Additionally, it identifies potential areas for future research and development, and ultimately explores the thesis's significance within the broader research community.

6.1 Summary of the Contributions

Throughout the research, two systems have been designed, developed, and evaluated: MRMAC and Vicarious, each addressing different aspects of the main research question (RQ) of effectively connecting and collaborating multiple remote (VR) users with local (AR) users in the Mixed Reality Collaboration (MRC) space. In the following, the achievements of this thesis are correlated with the objectives set in the introduction.

- **Design and Development of MRMAC**: The design and development of the Mixed Reality Multi-user Asymmetric Collaboration system, MRMAC presented in Chapter 3. To address **RQ1** regarding how multiple remote users achieve a high sense of presence while maintaining spatial awareness and understanding in the MRC space, a design concept and protocol for a multi-user asymmetric remote collaboration was introduced. The novel client-server architecture live-streams the physical environment of local users using a 360° camera in high resolution and low latency, while seamlessly integrating 3D virtual assets into the mixed-reality collaboration space. This allows remote users to see the local user's space in real-time.

  To address the question of how multiple remote VR users can be effectively represented in the MRC space to achieve co-presence (**RQ2**), MRMAC incorporated features such as 3D Avatars with avatar positioning and personalized avatar generation options. These features allow users to represent themselves in virtual environments, avoid overlapping with other avatars, providing a sense of ownership and identity, and facilitate social interaction by enabling users to visually perceive and identify each other within the MRC space.

  MRMAC also addressed **RQ3** regarding how multiple VR users can interact effectively in the MRC space by incorporating features such as viewpoint controls and synchronized low-latency audio, video, and asset streaming. These functionalities enable users to receive real-time feedback, navigate the space, and interact with other users and the environment seamlessly.

- **System and User Evaluation of MRMAC**: To evaluate the MRMAC system and assess its performance in achieving the aspects outlined in **RQ1** (teleportation), **RQ2** (representation), and **RQ3** (interaction), a comprehensive system evaluation and two distinct user studies were conducted, as presented in Chapter 4.

  Findings from the system evaluation showcased the system's capability to provide low-latency synchronized communication for both AR and VR users. The scalability results indicated that network latency remained under 1 second (650ms) for up to 15 users.

In User Study 1, which examined how user roles affect presence, interactions, and task performance, it was found that user roles did not significantly affect the immersive experience, with both roles experiencing an equal level of presence and no notable changes in task performance.

In User Study 2, which compared MRMAC against existing collaborative tools, it was observed that MRMAC induces significantly higher spatial presence than 2D and 360° video alone.

Overall, the evaluation and user studies demonstrated that rich real-time multiuser telepresence experiences can feasibly be created. However, it was also noted that these experiences still pose a significant engineering challenge. Despite this challenge, the findings opened up a plethora of future research opportunities in the field.

- **Vicarious Viewpoint Sharing**: MRMAC's support for multiple users with their viewpoint sharing motivated the exploration of multi-perspective collaboration, laying the foundation for subsequent research endeavors like Vicarious. This progression illustrates how the investigation of interaction dynamics in MRMAC naturally evolved to address the nuanced aspects of **RQ4** (viewpoint management) and **RQ5** (multi-perspective collaboration).

  In Vicarious, a novel context-aware viewpoint-sharing method was presented to manage users' viewpoints effectively and support various collaborative tasks in the mixed reality collaboration space, as outlined in Chapter 5. This method addresses the challenge of dynamically selecting optimal viewpoints for sharing among multiple users.

  To answer **RQ4** about how users' viewpoints can be managed effectively, four different variations of viewpoint selection methods were presented: No-view Selection, Manual Selection, Guided Selection, and Automatic Selection. Among these methods, Guided Selection and Automatic Selection leverage contextual information, including user actions, object interactions, and scene context.

- **User Study on Context-Aware Viewpoint Selection**: As the final research question aimed to find out how multi-perspective collaboration can be effectively achieved for supporting collaborative tasks (**RQ5**), a user study was conducted to assess the impact of the context-aware viewpoint selection method on collaboration performance and user experience, also presented in Chapter 5. Vicarious employed an asymmetric AR-VR setup to explore collaboration under four distinct conditions: No-view, Manual, Guided, and Automatic selection.

The outcomes from the user study indicated that optimal viewpoint selection improved users' task space understanding, task performance, and reduced cognitive load. Additionally, the results showed that users preferred the intervention, and their feedback provided valuable insights and recommendations for design implications, guiding future research endeavors in this area.

These contributions collectively contribute to an enhanced understanding of multiuser MR collaboration systems, offering valuable insights for the advancement of multiuser telecollaboration and viewpoint sharing in various applications.

6.2 Limitations

While the 360° video offers an immersive experience for remote users, it comes with certain limitations. One such limitation is the fixed position of the 360° camera, providing only 3 DoF for rotation and lacking the full 6 DoF. Additionally, the 360° video is projected onto 3D geometry to offer a basic sense of depth and limited parallax. However, this setup restricts the remote user's ability to move freely within the local space, and straying too far from the center of the 360° video can lead to noticeable projection artifacts. The proposed systems lack the capability to capture depth information, resulting in constraints on users' freedom to move and interact with objects in the MR environment. Furthermore, 360° video streaming is susceptible to issues such as jittering and video stretching due to compression or streaming problems. These issues can detract from the immersive experience and disrupt collaborative efforts.

While MRMAC and Vicarious have shown promise in enhancing collaboration, they may not fully address all possible collaborative scenarios or user preferences in different settings. Further customization may be required for specific use cases. User adoption and acceptance of MR technologies can be influenced by factors such as hardware constraints, cost, and technological literacy. These external factors can impact the widespread adoption of such solutions.

Another limitation of this thesis is the small convenience sample size, which was unevenly distributed across conditions in all three studies. This imbalance resulted from the constraints imposed by the COVID-19 pandemic, which hindered our ability to recruit a larger number of participants. The Ph.D. program was initiated right on the eve of the imposition of COVID-19 lockdown restrictions and was completed after the lifting of these restrictions. The way groups interacted and the availability of participants were altered by the effects of COVID-19, even after the restrictions were lifted. Furthermore, the lockdown restrictions and safety concerns also limited the ability to conduct experiments and utilize essential research tools and resources in the traditional way. The research prototypes needed to be tested within controlled settings to ensure accurate measurements could be taken, particularly to avoid network restrictions and firewall issues. Since commercial cloud solutions were cost-prohibitive and out of the scope of the research, in-lab studies were necessary. Additionally, the requirement for specialized equipment like head-mounted displays (HMDs) added another layer of complexity to recruiting participants, especially during the pandemic. Due to the multi-user group nature of the studies, equipment accessibility was a challenge. Not all potential participants may have had access to headsets or a reliable networking connection, presenting a significant barrier to conducting the study remotely.

Despite these challenges, the studies were conducted by implementing various strategies to navigate the evolving landscape. Safety protocols and guidelines were maintained to ensure the well-being of participants, and sanitation protocols for these specialized devices were utilized throughout the studies. Personal outreach was conducted to recruit participants, and additional efforts were made to reach out and accommodate potential participants. The number of participants recruited for each study, although uneven between conditions, is believed to be substantial enough to derive meaningful discussions and insights.

In addition, the research was conducted within a specific timeframe, which might have influenced the depth and breadth of data collection and analysis. For future research, efforts to expand the participant pool should be considered to address this limitation and provide a more comprehensive evaluation of the MR application and its associated hypotheses.

6.3 Future Work

The findings of this thesis suggest that the systems presented in this thesis have the potential to be a valuable tool for improving communication and collaborators' sense of presence. While this thesis has made several contributions, there are still some opportunities for further investigation and development. Besides the future work given for each chapter, the following future research directions can be explored:

- In the current configuration, the 360° video is streamed at its maximum resolution.

There is an opportunity for optimization by transmitting only the data that the remote user can see. Previous research has already explored this concept [265], and further exploration could lead to reduced bandwidth and computational requirements for the system. This could potentially lead to a better user experience.

- The current process for calibrating the space between local and remote users involves manually positioning and rotating a virtual representation of the 360° camera. AR devices detect reference points in the physical environment for inside-out tracking. One potential solution for automating camera calibration is to detect corresponding tracking points in the 360° video. Alternatively, in related research, tracking devices have been attached to the 360° camera to determine its position and orientation.
- Future enhancements and refinements could be implemented for MRMAC and Vicarious based on user feedback and evolving technology.
- Investigation into the scalability of these technologies for larger than 50 users groups and complex collaborative scenarios.
- In the current implementation, remote users are represented by semi-realistic or cartoonish avatars. Feedback from user evaluations almost unanimously suggests that participants expressed a preference for semi-realistic or cartoonish avatars. However, there is an interest in exploring more realistic 3D reconstructed avatars, as research has shown that the user's sense of embodiment in a remote space can be strongly influenced by the realism of their avatar [122, 109]. A concern is that transitioning to these more realistic representations might give rise to the uncanny valley effect, where users may find the avatars uncomfortably close to realism but not quite there. This potential issue requires thorough exploration and evaluation in future research.

6.4 Conclusion

In conclusion, the contributions made in this thesis have effectively addressed critical challenges in multi-user asymmetric collaboration and viewpoint selection, thereby enabling more immersive, inclusive, and effective collaborative experiences. As MR continues to evolve, this work serves as a strong foundation upon which future innovations can be built. This research carries significant implications for both academia and industry. The system developed can find applications across various domains, including education, healthcare, entertainment, and remote work. Asymmetric collaboration tools like MRMAC have the potential to foster more inclusive and efficient teamwork, while context-aware viewpoint selection, such as Vicarious, has the capacity to enhance user engagement and comprehension. Through the design, development, and evaluation of these systems, this thesis has shown to effectively connect multiple remote (VR) users with local (AR) users in the MRC space to facilitate collaborative tasks.

Multi-user MR collaboration is a rapidly developing technology with the potential to revolutionize the way we work, learn, and collaborate. As the technology continues to mature, we can expect to see even more innovative and exciting applications for multiuser MR collaboration. The hope is that this research inspires further exploration and development in this exciting and rapidly evolving field. The journey does not end here, and the transformative impact of MR collaboration in the years to come is eagerly anticipated.

Great sci-fi movies like *Star Wars* featured holographic messaging and teleportation almost 50 years ago. MR collaboration is so close to transforming today's desktop video conferencing into a life-size virtual telepresence system that the concept of teleportation is no longer a distant fantasy.

Appendix A

Ethics

There were various user studies conducted in this thesis. Each user study involved image quality or experience evaluation. The following is the generic information sheet, generic consent form, and generic image quality questions used. The exact wording and setup are described in each chapter that has a user study.

Collaboration System User Study

INFORMATION SHEET FOR PARTICIPANTS

You are invited to take part in this research. Please read this information before deciding whether or not to take part. If you decide to participate, thank you. If you decide not to participate, thank you for considering this request.

Who am I?

My name is Faisal Zaman and I am a PhD candidate in the Computational Media Innovation Center at Victoria University of Wellington.

What is the aim of the project?

This research project aims to provide an understanding of the users' needs and practices that will inform the design of a teleportation/telecollaboration platform. The intention of the platform is to allow people to communicate and collaborate remotely in real-time using immersive technologies such as Virtual Reality (VR) and Augmented Reality (AR). Your participation will support this research by providing expert knowledge towards user requirements for the platform. This research has been approved by the Victoria University of Wellington Human Ethics Committee #0000028674.

How can you help?

If you agree to take part, you will be asked a series of questions regarding your communication and collaboration work practices, any challenges you are facing and your thoughts with respect to the digital platform that is to be developed. During the interview, please avoid disclosing any confidential work-related information. I will audio record this interview with your permission and write it up later. If you would like to receive a copy of the audio recording of your interview you can request them.

The interview will take approximately 30 minutes. This interview will take place in person. You can stop the interview at any time by letting me know, without giving a reason. You can also choose not to answer any questions. You can withdraw from the study by contacting me within four weeks following this interview. If you withdraw, the information you provided will be destroyed or returned to you.

What will happen to the information you give?

This research is confidential. This means the researchers named below will be aware of your identity, but the research data will be aggregated, and your identity will not be revealed in any reports, presentations, or public documentation. You will not be named in any dissemination of findings, and pseudonyms may be assigned to protect your identity. Only A/Prof. Taehyun Rhee, Dr Craig Anslow, and I will access the audio recordings and transcripts. The material I collect will be kept securely and destroyed on conclusion of the project, no later than 31/12/2025.

What will the project produce?

The information from this research will be used to inform future development of the platform, which may result in commercialisable intellectual property. The information gathered may also be the subject of academic dissemination in professional journals and conferences as well as form parts of postgraduate theses.

What are the conditions for my participation in this study?

In order to participate in this study, you must:

- Be over 18 years of age
- Able to view images on a screen
- Not have any serious visual abnormalities (amblyopia, spherical aberrations, etc)

If you have a visual condition and are not sure if it will be an issue, please mention this to the researcher. Note that requiring prescription focal correction (i.e. long or short sightedness) is not an issue. If you are colour blind, you are still welcome to participate in the study, but you need to let the researcher know of your colour blind condition.

If you accept this invitation, what are your rights as a research participant?

You do not have to accept this invitation if you don't want to. If you do decide to participate, you have the right to:

- choose not to answer any question;
- ask for the recorder to be turned off at any time during the interview;
- withdraw from the study within one month from when the interview took place;
- ask any questions about the study at any time;
- receive a copy of your interview recording;
- be able to read any reports of this research by emailing the researcher to request a copy.

If you have any questions or problems, who can you contact?

If you have any questions, either now or in the future, please feel free to contact me:

- Student researcher: Faisal Zaman
  Role: PhD Candidate
  Email: *faisal.zaman@vuw.ac.nz*
- Supervisor: Associate Professor Taehyun Rhee
  Role: Associate Professor
  School: Engineering and Computer Science
  Phone: 04 463 5233 x 7088 or Email: *taehyun.rhee@ecs.vuw.ac.nz*
- Supervisor: Dr Craig Anslow
  Role: Senior Lecturer
  School: Engineering and Computer Science
  Phone: 04 463 6449 or Email: *craig.anslow@vuw.ac.nz*

Human Ethics Committee information
If you have any concerns about the ethical conduct of the research you may contact the Victoria University of Wellington HEC Convenor: Associate Professor Judith Loveridge. Email hec@vuw.ac.nz or telephone +64-4-463 6028.

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA

Collaboration System User Study
PARTICIPANTS CONSENT FORM

This consent form will be held for five years.

Researcher: Faisal Zaman, Computational Media Innovation Centre, Victoria University of Wellington.

- I have read the Information Sheet, and the project has been explained to me. My questions have been answered to my satisfaction. I understand that I can ask further questions at any time.
- I agree to take part in this user test.

I understand that:

- I may withdraw from this study at any point, and any information that I have provided will be returned to me or destroyed.
- All identifiable information I have provided will be destroyed on 31/12/2025.
- I understand that any information I provide may be used for the development of the platform, commercializable intellectual property, and/or dissemination in academic publications and conferences.
- My name will not be used in reports, and utmost care will be taken not to disclose any information that would identify me.

I would like to receive a copy of the final report and have added my email address below.

☐ Yes

☐ No

Signature of participant: ______________________________
Name of participant: ______________________________
Date: ______________________________
Contact details: ______________________________
Email address: ______________________________

Appendix B

User Evaluation Questionnaire

The following questionnaires were administered to participants to evaluate the systems detailed in both Chapter 4 and Chapter 5. Each questionnaire was completed after each condition to assess the level of spatial awareness, social interaction, co-presence, task workload, system usability, and any simulator sickness experienced by the participants. Following the completion of all conditions, a preference questionnaire was distributed to participants to determine their preferences and gather feedback.

Participant Demographics I

Participant	Age	Sex	Ethnicity	VR/AR Usage	Hours	Vision
P1	23	Male	European	Never	-	Normal
P2	41	Male	European	Monthly	20	Normal
P3	25	Male	Mixed race	Never	-	Normal
P4	74	Male	European	Weekly	5	Normal
P5	23	Male	Asian	Weekly	10	Normal
P6	22	Male	European	Never	-	Normal
P7	51	Male	Mixed race	Never	-	Normal
P8	29	Male	Mixed race	Never	-	Normal
P9	27	Male	European	Never	-	Normal
P10	27	Male	Latino/Hispanic	Never	-	Corrected
P11	26	Male	Asian	Never	-	Normal
P12	36	Female	Asian	Never	-	Normal
P13	24	Male	Asian	Never	-	Normal
P14	25	Male	Mixed race	Yearly	-	Corrected
P15	26	Male	Mixed race	Yearly	-	Normal
P16	33	Male	Latino/Hispanic	Never	-	Normal
P17	35	Female	European	Yearly	-	Normal
P18	29	Male	European	Never	-	Corrected
P19	26	Male	European	Yearly	-	Corrected
P20	22	Male	European	Monthly	15	Corrected
P21	27	Female	European	Never	-	Normal
P22	28	Male	European	Never	-	Normal
P23	26	Male	Asian	Never	-	Corrected
P24	25	Male	Pacific Islander	Never	-	Normal

Participant Demographics II

Participant	Age	Sex	Ethnicity	Vision	VR/AR Usage
P1	24	Female	European	Corrected	Very frequently
P2	81	Male	Asian	Normal	Somewhat frequently
P3	60	Female	European	Normal	Occasionally
P4	65	Male	Asian	Corrected	Rarely
P5	26	Female	European	Normal	Never
P6	22	Male	European	Corrected	Very frequently
P7	18	Male	European	Normal	Occasionally
P8	23	Female	European	Normal	Occasionally
P9	20	Female	European	Normal	Occasionally
P10	24	Female	European	Normal	Occasionally
P11	32	Female	Asian	Normal	Occasionally
P12	32	Male	European	Normal	Occasionally
P13	24	Female	European	Corrected	Occasionally
P14	21	Male	European	Normal	Occasionally
P15	37	Female	European	Corrected	Occasionally
P16	23	Male	European	Corrected	Occasionally
P17	22	Female		Corrected	Occasionally
P18	33	Male	European	Corrected	Occasionally
P19	23	Female	Asian	Corrected	Occasionally
P20	23	Female	European	Corrected	Somewhat frequently
P21	19	Female	Latin American	Corrected	Somewhat frequently
P22	25	Female	Asian	Corrected	Somewhat frequently
P23	20	Female	Asian	Normal	Somewhat frequently
P24	19	Female	Polynesian	Normal	Rarely
P25	36	Female	Polynesian	Corrected	Very frequently
P26	49	Male	Asian	Normal	Somewhat frequently
P27	49	Male	Latin American	Normal	Occasionally
P28	32	Male	Asian	Normal	Very frequently
P29	18	Female	European	Corrected	Somewhat frequently
P30	27	Female	European	Corrected	Occasionally
P31	21	Male	European	Corrected	Occasionally
P32	31	Male	European	Corrected	Never
P33	33	Male	Asian	Corrected	Very frequently
P34	29	Male	European	Corrected	Very frequently
P35	36	Male	Latin American	Normal	Rarely
P36	33	Male	Asian	Corrected	Very frequently

Participant Demographics III

Participant	Age	Sex	Ethnicity	VR/AR Usage	Vision
P1	23	Male	European	Never	Normal
P2	41	Male	European	Monthly	Normal
P3	25	Male	Mixed race	Never	Normal
P4	55	Male	European	Weekly	Normal
P5	23	Male	Asian	Weekly	Normal
P6	31	Male	European	Never	Normal
P7	25	Male	Mixed race	Never	Normal
P8	29	Male	Mixed race	Never	Normal
P9	27	Male	European	Never	Normal
P10	21	Male	Latino/Hispanic	Never	Corrected
P11	26	Male	Asian	Never	Normal
P12	36	Female	Asian	Never	Normal
P13	28	Male	Asian	Never	Normal
P14	25	Male	Mixed race	Yearly	Corrected
P15	26	Male	Mixed race	Yearly	Normal
P16	33	Male	Latino/Hispanic	Never	Normal
P17	18	Female	European	Yearly	Normal
P18	29	Male	European	Never	Corrected
P19	26	Male	European	Yearly	Corrected
P20	22	Male	European	Monthly	Corrected
P21	27	Female	European	Never	Normal
P22	28	Male	European	Never	Normal
P23	26	Male	Asian	Never	Corrected
P24	25	Male	Pacific Islander	Never	Normal
P25	27	Female	European	Never	Normal
P26	31	Female	Asian	Weekly	Normal
P27	21	Female	Latino/Hispanic	Never	Normal

User Study Questionnaire I

1. In the computer-generated world I had a sense of being there.

   Not at all 1 2 3 4 5 6 7 Very much

2. Somehow I felt that the virtual world surrounded me.

   Very Difficult 1 2 3 4 5 6 7 Very Easy

3. I felt like I was just perceiving pictures.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. I did not feel present in the virtual space.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

5. I had a sense of acting in the virtual space, rather than operating something from outside.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

6. I felt present in the virtual space.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

7. How aware were you of the real world surrounding you while navigating in the virtual world?

   Extremely aware 1 2 3 4 5 6 7 Not aware at all

8. I was not aware of my real environment.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

9. I still paid attention to the real environment.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

10. I was completely captivated by the virtual world.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

11. How real did the virtual world seem to you?

Not real at all 1 2 3 4 5 6 7 Completely real

12. Any additional comments?

User Study Questionnaire II

1. I often felt as if I was all alone

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

2. I think the other individual often felt alone.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

3. I hardly noticed the other individual.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

4. The other person hardly noticed me.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

5. I was often aware of others in the environment.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

6. Others were often aware of me in the room.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

7. I think the other individual often felt alone.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

8. I often felt as if I was all alone.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

9. Any additional comments?

Simulator Sickness Questionnaire

Kennedy, Lane, Berbaum, & Lilienthal (1993)[1]

Instructions: Circle how much each symptom below is affecting you right now.

1. General discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Severe	Moderate	Severe
3. Headache	None	Severe	Moderate	Severe
4. Eye strain	None	Severe	Moderate	Severe
5. Difficulty focusing	None	Severe	Moderate	Severe
6. Increased salivation	None	Severe	Moderate	Severe
7. Sweating	None	Severe	Moderate	Severe
8. Nausea	None	Severe	Moderate	Severe
9. Difficulty concentrating	None	Severe	Moderate	Severe
10. Fullness of head*	None	Severe	Moderate	Severe
11. Blurred vision	None	Severe	Moderate	Severe
12. Dizziness (eyes open)	None	Severe	Moderate	Severe
13. Dizziness (eyes shut)	None	Severe	Moderate	Severe
14. Vertigo†	None	Severe	Moderate	Severe
15. Stomach awareness‡	None	Severe	Moderate	Severe
16. Burping	None	Severe	Moderate	Severe

* Fullness of the head refers to a sensation of pressure in the head without any pain like that experienced when upside down.

† Vertigo is a loss of orientation with respect to verticality like the sensation felt at great heights.

‡ Stomach awareness is usually used to indicate a feeling of discomfort that is just short of nausea.

[1] Original version: Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. International Journal of Aviation Psychology, 3(3), 203-220.

Bibliography

[1] ABTAHI, P., GONZALEZ-FRANCO, M., OFEK, E., AND STEED, A. I'm a giant: Walking in large virtual environments at high speed gains. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–13.

[2] ADCOCK, M., ANDERSON, S., AND THOMAS, B. Remotefusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry* (2013), pp. 235–242.

[3] ADCOCK, M., AND GUNN, C. Using projected light for mobile remote guidance. *Computer Supported Cooperative Work (CSCW) 24*, 6 (2015), 591–611.

[4] AL-ADHAMI, M., WU, S., AND MA, L. Extended reality approach for construction quality control. In - (2019), International Council for Research and Innovation in Building and ....

[5] AMORES, J., BENAVIDES, X., AND MAES, P. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (2015), pp. 1343–1348.

[6] ANSLOW, C. Collaborative software visualization in co-located environments. *Victoria University of Wellington* (2013).

[7] ANTON, D., KURILLO, G., AND BAJCSY, R. User experience and interaction performance in 2d/3d telecollaboration. *Future Generation Computer Systems 82* (2018), 77–88.

[8] APICHARTTRISORN, K., BALASUBRAMANIAN, B., CHEN, J., SIVARAJ, R., TSAI, Y.-Z., JANA, R., KRISHNAMURTHY, S., TRAN, T., AND ZHOU, Y. Characterization of

multi-user augmented reality over cellular networks. In *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (2020), IEEE, pp. 1–9.

[9] APPLEGATE, L. M. Technology support for cooperative work: A framework for studying introduction and assimilation in organizations. *Journal of Organizational Computing 1*, 1 (1991), 11–39.

[10] ARUNPRASAD, P., DEY, C., JEBLI, F., MANIMUTHU, A., AND EL HATHAT, Z. Exploring the remote work challenges in the era of covid-19 pandemic: review and application model. *Benchmarking: An International Journal 29*, 10 (2022), 3333–3355.

[11] AZUMA, R., BAILLOT, Y., BEHRINGER, R., FEINER, S., JULIER, S., AND MACINTYRE, B. Recent advances in augmented reality. *IEEE computer graphics and applications 21*, 6 (2001), 34–47.

[12] BAECKER, R. M. *Readings in groupware and computer-supported cooperative work: Assisting human-human collaboration*. Morgan Kaufmann, 1993.

[13] BAECKER, R. M., GRUDIN, J., BUXTON, W. A., AND GREENBERG, S. Chapter 11 - groupware and computer-supported cooperative work. In *Readings in Human–Computer Interaction*, Interactive Technologies. Morgan Kaufmann, 1995, pp. 741–753.

[14] BAILENSON, J. N., SWINTH, K., HOYT, C., PERSKY, S., DIMOV, A., AND BLASCOVICH, J. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence 14* (2005), 379–393.

[15] BAKER, L., MILLS, S., ZOLLMANN, S., AND VENTURA, J. Casualstereo: Casual capture of stereo panoramas with spherical structure-from-motion. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), IEEE, pp. 782–790.

[16] BANNISTER, H., SELWYN-SMITH, B., ANSLOW, C., ROBINSON, B., DOS ANJOS, R. K., PIRES, D., KANE, P., AND LEONG, A. Linacvr: Vr simulation for radiation therapy education. In *VRST* (2019), pp. 79–1.

[17] BEKELE, M. K., PIERDICCA, R., FRONTONI, E., MALINVERNI, E. S., AND GAIN, J. A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH) 11*, 2 (2018), 1–36.

[18] BEN-JOSEPH, E., ISHII, H., UNDERKOFFLER, J., PIPER, B., AND YEUNG, L. Urban simulation and the luminous planning table: Bridging the gap between the digital and the tangible. *Journal of planning Education and Research 21*, 2 (2001), 196–203.

[19] BENKO, H., ISHAK, E. W., AND FEINER, S. Collaborative mixed reality visualization of an archaeological excavation. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality* (2004), IEEE, pp. 132–140.

[20] BILLINGHURST, M., BEE, S., BOWSKILL, J., AND KATO, H. Asymmetries in collaborative wearable interfaces. In *Digest of Papers. Third International Symposium on Wearable Computers* (1999), IEEE, pp. 133–140.

[21] BILLINGHURST, M., BELCHER, D., GUPTA, A., AND KIYOKAWA, K. Communication behaviors in colocated collaborative ar interfaces. *International Journal of Human-Computer Interaction 16*, 3 (2003), 395–423.

[22] BILLINGHURST, M., CORDEIL, M., BEZERIANOS, A., AND MARGOLIS, T. Collaborative immersive analytics. In *Immersive Analytics*. Springer, 2018, pp. 221–257.

[23] BILLINGHURST, M., KATO, H., AND POUPYREV, I. The magicbook: a transitional ar interface. *Computers & Graphics 25*, 5 (2001), 745–753.

[24] BILLINGHURST, M., NASSANI, A., AND REICHHERZER, C. Social panoramas: using wearable computers to share experiences. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications* (2014), pp. 1–1.

[25] BILLINGHURST, M., AND THOMAS, B. H. Mobile collaborative augmented reality. In *Recent trends of mobile collaborative augmented reality systems*. Springer, 2011, pp. 1–19.

[26] BIMBER, O., AND RASKAR, R. *Spatial augmented reality: merging real and virtual worlds*. CRC press, 2005.

[27] BIOCCA, F. Virtual reality technology: A tutorial. *Journal of communication 42*, 4 (1992), 23–72.

[28] BIOCCA, F., HARMS, C., AND BURGOON, J. K. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments 12* (2003), 456–480.

[29] BIOCCA, F., KIM, J., AND CHOI, Y. Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence: Teleoperators & Virtual Environments 10*, 3 (2001), 247–265.

[30] BLEI, D. M., NG, A. Y., AND JORDAN, M. I. Latent dirichlet allocation. *Journal of machine Learning research 3*, Jan (2003), 993–1022.

[31] BOLETSIS, C. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction 1*, 4 (2017), 24.

[32] BORHANI, Z., SHARMA, P., AND ORTEGA, F. R. Survey of annotations in extended reality systems. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[33] BREITENEDER, C. J., GIBBS, S. J., AND ARAPIS, C. Teleport—an augmented reality teleconferencing environment. In *Virtual Environments and Scientific Visualization'96: Proceedings of the Eurographics Workshops in Monte Carlo, Monaco, February 19–20, 1996, and in Prague, Czech Republic, April 23–25, 1996* (1996), Springer, pp. 41–49.

[34] BROLL, W., LINDT, I., OHLENBURG, J., WITTKÄMPER, M., YUAN, C., NOVOTNY, T., MOTTRAM, C., STROTHMANN, A., ET AL. Arthur: A collaborative augmented environment for architectural design and urban planning. *JVRB-Journal of Virtual Reality and Broadcasting 1*, 1 (2004).

[35] BROOKE, J., ET AL. Sus-a quick and dirty usability scale. *Usability evaluation in industry 189* (1996), 4–7.

[36] BÜSCHEL, W., LEHMANN, A., AND DACHSELT, R. Miria: A mixed reality toolkit for the in-situ visualization and analysis of spatio-temporal interaction data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–15.

[37] BUTZ, A., HOLLERER, T., FEINER, S., MACINTYRE, B., AND BESHERS, C. Enveloping users and computers in a collaborative 3d augmented reality. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)* (1999), IEEE, pp. 35–44.

[38] BUXTON, W. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface* (1992), Canadian Information Processing Society Toronto, Canada, pp. 123–129.

[39] CAO, D., AND KIM, S. Immediate presence of handwritten annotation on moving picture in real-time collaborative system. *IEEE Access 8* (2020), 133258–133268.

[40] CARLSSON, C., AND HAGSAND, O. Dive a multi-user virtual reality system. In *Proceedings of IEEE virtual reality annual international symposium* (1993), IEEE, pp. 394–400.

[41] Castañer, X., and Oliveira, N. Collaboration, coordination, and cooperation among organizations: Establishing the distinctive meanings of these terms through a systematic literature review. *Journal of management 46*, 6 (2020), 965–1001.

[42] Chang, Y.-C., Wang, H.-C., Chu, H.-k., Lin, S.-Y., and Wang, S.-P. Alpharead: Support unambiguous referencing in remote collaboration with readable object annotation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (2017), pp. 2246–2259.

[43] Chellali, A., Milleville-Pennel, I., and Dumas, C. Influence of contextual objects on spatial interactions and viewpoints sharing in virtual environments. *Virtual Reality 17*, 1 (2013), 1–15.

[44] Chen, Y., Farley, T., and Ye, N. Qos requirements of network applications on the internet. *Information Knowledge Systems Management 4*, 1 (2004), 55–76.

[45] Choi, S. H., Kim, M., and Lee, J. Y. Situation-dependent remote ar collaborations: Image-based collaboration using a 3d perspective map and live video-based collaboration with a synchronized vr mode. *Computers in Industry 101* (2018), 51–66.

[46] Chow, K., Coyiuto, C., Nguyen, C., and Yoon, D. Challenges and design considerations for multimodal asynchronous collaboration in vr. *Proceedings of the ACM on Human-Computer Interaction 3*, CSCW (2019), 1–24.

[47] Cisco Systems, I. Cisco telepresence system t3 system assembly guide. https://www.cisco.com/c/dam/en/us/td/docs/telepresence/endpoint/t3/guides/t3_system_assembly_guide.pdf, December 2011.

[48] Clark, H. H., and Brennan, S. E. Grounding in communication. *In Perspectives on socially shared cognition.* (1991), 127–149.

[49] Clergeaud, D., Roo, J. S., Hachet, M., and Guitton, P. Towards seamless interaction between physical and virtual locations for asymmetric collaboration. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (2017), pp. 1–4.

[50] Clerke, T., and Hopwood, N. *Doing ethnography in teams: A case study of asymmetries in collaborative research.* Springer Science & Business Media, 2014.

[51] Daly-Jones, O., Monk, A., and Watts, L. Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *International Journal of Human-Computer Studies 49*, 1 (1998), 21–58.

[52] DE BELEN, R. A. J., NGUYEN, H., FILONIK, D., DEL FAVERO, D., AND BEDNARZ, T. A systematic review of the current state of collaborative mixed reality technologies: 2013–2018. *AIMS Electronics and Electrical Engineering 3*, 2 (2019), 181.

[53] DE GUZMAN, J. A., THILAKARATHNA, K., AND SENEVIRATNE, A. Security and privacy approaches in mixed reality: A literature survey. *ACM Computing Surveys (CSUR) 52*, 6 (2019), 1–37.

[54] DEY, A., BILLINGHURST, M., LINDEMAN, R. W., AND SWAN, J. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI 5* (2018), 37.

[55] DOOLY, M. Telecollaboration. *The handbook of technology and second language teaching and learning* (2017), 169–183.

[56] DOURISH, P., ADLER, A., BELLOTTI, V., AND HENDERSON, A. Your place or mine? learning from long-term use of audio-video communication. *Computer Supported Cooperative Work (CSCW) 5* (1996), 33–62.

[57] DU, J., SHI, Y., ZOU, Z., AND ZHAO, D. Covr: Cloud-based multiuser virtual reality headset system for project communication of remote users. *Journal of Construction Engineering and Management 144*, 2 (2018), 04017109.

[58] EIRIS, R., GHEISARI, M., AND ESMAEILI, B. Desktop-based safety training using 360-degree panorama and static virtual reality techniques: A comparative experimental study. *Automation in construction 109* (2020), 102969.

[59] ELKIN, L. A., KAY, M., HIGGINS, J. J., AND WOBBROCK, J. O. An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology* (2021), pp. 754–768.

[60] ELLIS, C., AND WAINER, J. Groupware and computer supported cooperative work. *Multiagent systems: a modern approach to distributed artificial intelligence* (1999), 425–457.

[61] ENS, B., LANIR, J., TANG, A., BATEMAN, S., LEE, G., PIUMSOMBOON, T., AND BILLINGHURST, M. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies 131* (2019), 81–98.

[62] FAGES, A., FLEURY, C., AND TSANDILAS, T. Understanding multi-view collaboration between augmented reality and remote desktop users. *Proceedings of the ACM on Human-Computer Interaction 6*, CSCW2 (2022), 1–27.

[63] FAIRCHILD, A. J., CAMPION, S. P., GARCÍA, A. S., WOLFF, R., FERNANDO, T., AND ROBERTS, D. J. A mixed reality telepresence system for collaborative space operation. *IEEE Transactions on Circuits and Systems for Video Technology 27*, 4 (2016), 814–827.

[64] FARIDAN, M., KUMARI, B., AND SUZUKI, R. Chameleoncontrol: Teleoperating real human surrogates through mixed reality gestural guidance for remote hands-on classrooms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023), pp. 1–13.

[65] FELD, N., AND WEYERS, B. Overview of collaborative virtual environments using augmented reality. *Mensch und Computer 2019-Workshopband* (2019).

[66] FIDALGO, C. G., YAN, Y., CHO, H., SOUSA, M., LINDLBAUER, D., AND JORGE, J. A survey on remote assistance and training in mixed reality environments. *IEEE Transactions on Visualization and Computer Graphics 29*, 5 (2023), 2291–2303.

[67] FRIBOURG, R., ARGELAGUET, F., LÉCUYER, A., AND HOYET, L. Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE transactions on visualization and computer graphics 26*, 5 (2020), 2062–2072.

[68] FUSSELL, S. R., KRAUT, R. E., AND SIEGEL, J. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (2000), pp. 21–30.

[69] FUSSELL, S. R., SETLOCK, L. D., PARKER, E. M., AND YANG, J. Assessing the value of a cursor pointing device for remote collaboration on physical tasks. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* (2003), pp. 788–789.

[70] FUSSELL, S. R., SETLOCK, L. D., YANG, J., OU, J., MAUER, E., AND KRAMER, A. D. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction 19*, 3 (2004), 273–309.

[71] GAO, L., BAI, H., LEE, G., AND BILLINGHURST, M. An oriented point-cloud view for mr remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*. ., 2016, pp. 1–4.

[72] GAO, L., BAI, H., LINDEMAN, R., AND BILLINGHURST, M. Static local environment capturing and sharing for mr remote collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications* (2017), pp. 1–6.

[73] GAUGLITZ, S., NUERNBERGER, B., TURK, M., AND HÖLLERER, T. In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (2014), pp. 197–205.

[74] GAUGLITZ, S., NUERNBERGER, B., TURK, M., AND HÖLLERER, T. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (2014), pp. 449–459.

[75] GOEBBELS, G., AND LALIOTI, V. Co-presence and co-working in distributed collaborative virtual environments. In *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation* (2001), pp. 109–114.

[76] GRANDI, J. G., BERNDT, I., DEBARBA, H. G., NEDEL, L., AND MACIEL, A. Collaborative manipulation of 3d virtual objects in augmented reality scenarios using mobile devices. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)* (2017), IEEE, pp. 264–265.

[77] GRANDI, J. G., DEBARBA, H. G., AND MACIEL, A. Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), IEEE, pp. 127–135.

[78] GRASSET, R., AND GASCUEL, J.-D. Mare: Multiuser augmented reality environment on table setup. In *ACM SIGGRAPH 2002 conference abstracts and applications* (2002), pp. 213–213.

[79] GRASSET, R., LAMB, P., AND BILLINGHURST, M. Evaluation of mixed-space collaboration. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)* (2005), IEEE, pp. 90–99.

[80] GREENBERG, S., AND MARWOOD, D. Real time groupware as a distributed system: concurrency control and its effect on the interface. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (1994), CSCW '94, ACM, p. 207–217.

[81] GREENWALD, S. W., CORNING, W., AND MAES, P. Multi-user framework for collaboration and co-creation in virtual reality. In . (2017), 12th International Conference on Computer Supported Collaborative Learning ....

[82] GUGENHEIMER, J., STEMASOV, E., FROMMEL, J., AND 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 (2017), pp. 4021–4033.

[83] GUPTA, K., LEE, G. A., AND BILLINGHURST, M. Do you see what i see? the effect of gaze tracking on task space remote collaboration. *IEEE transactions on visualization and computer graphics 22*, 11 (2016), 2413–2422.

[84] GUREVICH, P., LANIR, J., AND COHEN, B. Design and implementation of teleadvisor: a projection-based augmented reality system for remote collaboration. *Computer Supported Cooperative Work (CSCW) 24* (2015), 527–562.

[85] GUTWIN, C., AND GREENBERG, S. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW) 11* (2002), 411–446.

[86] HARMS, C., AND BIOCCA, F. Internal consistency and reliability of the networked minds measure of social presence. In *Seventh annual international workshop: Presence* (2004), vol. 2004, Universidad Politecnica de Valencia Valencia, Spain.

[87] HART, S. G., AND STAVELAND, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52. Elsevier, 1988, pp. 139–183.

[88] HAUBER, J., REGENBRECHT, H., BILLINGHURST, M., AND COCKBURN, A. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (2006), pp. 413–422.

[89] HEETER, C. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments 1* (1992), 262–271.

[90] HEWLETT-PACKARD. Hp unveils halo collaboration studios: Life-like communication leaps across geographic boundaries. https://www.itnews.com.au/gallery/inside-hp-halo-us350000-of-technology-128062/page5, December 2005.

[91] HIGUCH, K., YONETANI, R., AND SATO, Y. Can eye help you? effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 5180–5190.

[92] HITCHCOCK, A., AND SUNG, K. Multi-view augmented reality with a drone. In *Proceedings of the 24th ACM symposium on virtual reality software and technology* (2018), pp. 1–2.

[93] HUANG, W., AND ALEM, L. Handsinair: a wearable system for remote collaboration on physical tasks. In *Proceedings of the 2013 conference on Computer supported cooperative work companion* (2013), pp. 153–156.

[94] HUANG, W. P. Modelling the effects of typhoons on morphological changes in the estuary of beinan, taiwan. *Continental Shelf Research 135* (2017), 1–13.

[95] IBAYASHI, H., SUGIURA, Y., SAKAMOTO, D., MIYATA, N., TADA, M., OKUMA, T., KURATA, T., MOCHIMARU, M., AND IGARASHI, T. Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In *SIGGRAPH Asia 2015 Emerging Technologies*. ., 2015, pp. 1–2.

[96] IRLITTI, A., SMITH, R. T., VON ITZSTEIN, S., BILLINGHURST, M., AND THOMAS, B. H. Challenges for asynchronous collaboration in augmented reality. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)* (2016), IEEE, pp. 31–35.

[97] JACK, M. M., GATTOZZI, D. A., CAMARATA, P. J., AND SHAH, K. J. Live-streaming surgery for medical student education-educational solutions in neurosurgery during the covid-19 pandemic. *Journal of surgical education 78*, 1 (2021), 99–103.

[98] JANSEN, P., FISCHBACH, F., GUGENHEIMER, J., STEMASOV, E., FROMMEL, J., AND RUKZIO, E. Share: Enabling co-located asymmetric multi-user interaction for augmented reality head-mounted displays. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (2020), pp. 459–471.

[99] JEONG, K., KIM, J., KIM, M., LEE, J., AND KIM, C. Asymmetric interface: user interface of asymmetric virtual reality for new presence and experience. *Symmetry 12*, 1 (2020), 53.

[100] JO, H., AND HWANG, S. Chili: viewpoint control and on-video drawing for mobile video calls. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ., 2013, pp. 1425–1430.

[101] JOHANSEN, R. *Groupware: Computer support for business teams*. The Free Press, 1988.

[102] JONES, B., WITCRAFT, A., BATEMAN, S., NEUSTAEDTER, C., AND TANG, A. Mechanics of camera work in mobile video collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), pp. 957–966.

[103] JUAN, L., AND OUBONG, G. Surf applied in panorama image stitching. In *2010 2nd international conference on image processing theory, tools and applications* (2010), IEEE, pp. 495–499.

[104] KASAHARA, S., NAGAI, S., AND REKIMOTO, J. Jackin head: Immersive visual telepresence system with omnidirectional wearable camera. *IEEE transactions on visualization and computer graphics 23*, 3 (2016), 1222–1234.

[105] KASAHARA, S., AND REKIMOTO, J. Jackin: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th augmented human international conference* (2014), pp. 1–8.

[106] KASAHARA, S., AND REKIMOTO, J. Jackin head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM symposium on virtual reality software and technology* (2015), pp. 217–225.

[107] KATO, H., BILLINGHURST, M., POUPYREV, I., IMAMOTO, K., AND TACHIBANA, K. Virtual object manipulation on a table-top ar environment. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)* (2000), Ieee, pp. 111–119.

[108] KENNEDY, R. S., LANE, N. E., BERBAUM, K. S., AND LILIENTHAL, M. G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology 3* (1993), 203–220.

[109] KIM, D. Y., LEE, H. K., AND CHUNG, K. Avatar-mediated experience in the metaverse: The impact of avatar realism on user-avatar relationship. *Journal of Retailing and Consumer Services 73* (2023), 103382.

[110] KIM, M., AND WOHN, K. Holobox: Augmented visualization and presentation with spatially integrated presenter. *Interacting with Computers 30*, 3 (2018), 224–242.

[111] KIM, S., BILLINGHURST, M., LEE, C., AND LEE, G. Using freeze frame and visual notifications in an annotation drawing interface for remote collaboration. *KSII Transactions on Internet and Information Systems (TIIS) 12*, 12 (2018), 6034–6056.

[112] KIM, S., JING, A., PARK, H., KIM, S.-H., LEE, G., AND BILLINGHURST, M. Use of gaze and hand pointers in mixed reality remote collaboration. In *The 9th International Conference on Smart Media and Applications. SMA, Jeju, Republic of Korea* (2020), pp. 1–6.

[113] KIM, S., LEE, G., BILLINGHURST, M., AND HUANG, W. The combination of visual communication cues in mixed reality remote collaboration. *Journal on Multimodal User Interfaces 14* (2020), 321–335.

[114] KIM, S., LEE, G., HUANG, W., KIM, H., WOO, W., AND BILLINGHURST, M. Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (2019), pp. 1–13.

[115] KIM, S., LEE, G., SAKATA, N., AND BILLINGHURST, M. Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2014), IEEE, pp. 83–92.

[116] KIYOKAWA, K., TAKEMURA, H., AND YOKOYA, N. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. In *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028)* (1999), vol. 6, IEEE, pp. 48–53.

[117] KOMIYAMA, R., MIYAKI, T., AND REKIMOTO, J. Jackin space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference* (2017), pp. 1–9.

[118] KRATZ, S., KIMBER, D., SU, W., GORDON, G., AND SEVERNS, D. Polly: " being there" through the parrot and a guide. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (2014), pp. 625–630.

[119] KRAUT, R. E., FUSSELL, S. R., BRENNAN, S. E., AND SIEGEL, J. Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. *MIT press* (2002).

[120] KRAUT, R. E., FUSSELL, S. R., AND SIEGEL, J. Visual information as a conversational resource in collaborative physical tasks. *Human–computer interaction 18*, 1-2 (2003), 13–49.

[121] KURILLO, G., AND BAJCSY, R. 3d teleimmersion for collaboration and interaction of geographically distributed users. *Virtual Reality 17* (2013), 29–43.

[122] LATOSCHIK, M. E., ROTH, D., GALL, D., ACHENBACH, J., WALTEMATE, T., AND BOTSCH, M. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (2017), pp. 1–10.

[123] LE CHÉNÉCHAL, M., DUVAL, T., GOURANTON, V., ROYAN, J., AND ARNALDI, B. Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)* (2016), IEEE, pp. 9–12.

[124] LE CHÉNÉCHAL, M., DUVAL, T., GOURANTON, V., ROYAN, J., AND ARNALDI, B. Help! i need a remote guide in my mixed reality collaborative environment. *Frontiers in Robotics and AI 6* (2019), 106.

[125] LEE, G., KANG, H., LEE, J., AND HAN, J. A user study on view-sharing techniques for one-to-many mixed reality collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), IEEE, pp. 343–352.

[126] LEE, G. A., AHN, S., HOFF, W., AND BILLINGHURST, M. Enhancing first-person view task instruction videos with augmented reality cues. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2020), IEEE, pp. 498–508.

[127] LEE, G. A., KIM, S., LEE, Y., DEY, A., PIUMSOMBOON, T., NORMAN, M., AND BILLINGHURST, M. Improving collaboration in augmented video conference using mutually shared gaze. In *ICAT-EGVE* (2017), pp. 197–204.

[128] LEE, G. A., TEO, T., KIM, S., AND BILLINGHURST, M. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications* (2017), pp. 1–4.

[129] LEE, G. A., TEO, T., KIM, S., AND BILLINGHURST, M. Sharedsphere: Mr collaboration through shared live panorama. In *SIGGRAPH Asia 2017 Emerging Technologies*. ., 2017, pp. 1–2.

[130] LEE, G. A., TEO, T., KIM, S., AND BILLINGHURST, M. A user study on mr remote collaboration using live 360 video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2018), IEEE, pp. 153–164.

[131] LEE, J., KIM, M., AND KIM, J. Rolevr: Multi-experience in immersive virtual reality between co-located hmd and non-hmd users. *Multimedia Tools and Applications 79*, 1 (2020), 979–1005.

[132] LEVIN, A., ZOMET, A., PELEG, S., AND WEISS, Y. Seamless image stitching in the gradient domain. In *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV 8* (2004), Springer, pp. 377–389.

[133] LIN, C., ROJAS-MUNOZ, E., CABRERA, M. E., SANCHEZ-TAMAYO, N., ANDERSEN, D., POPESCU, V., NOGUERA, J. A. B., ZARZAUR, B., MURPHY, P., ANDERSON, K., ET AL. How about the mentor? effective workspace visualization in ar telementoring. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), IEEE, pp. 212–220.

[134] LINDLBAUER, D., AND WILSON, A. D. Remixed reality: manipulating space and time in augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–13.

[135] LIU, Z., LAN, G., STOJKOVIC, J., ZHANG, Y., JOE-WONG, C., AND GORLATOVA, M. Collabar: Edge-assisted collaborative image recognition for mobile augmented reality. In *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)* (2020), IEEE, pp. 301–312.

[136] LOMBARD, M., DITTON, T. B., AND WEINSTEIN, L. Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence* (2009), pp. 1–15.

[137] LORETO, S., AND ROMANO, S. P. *Real-time communication with WebRTC: peer-to-peer in the browser*. " O'Reilly Media, Inc.", 2014.

[138] LUKOSCH, S., BILLINGHURST, M., ALEM, L., AND KIYOKAWA, K. Collaboration in augmented reality. *Computer Supported Cooperative Work (CSCW) 24* (2015), 515–525.

[139] LUKOSCH, S., LUKOSCH, H., DATCU, D., AND CIDOTA, M. Providing information on the spot: Using augmented reality for situational awareness in the security domain. *Computer Supported Cooperative Work (CSCW) 24* (2015), 613–664.

[140] MAHMOOD, T., FULMER, W., MUNGOLI, N., HUANG, J., AND LU, A. Improving information sharing and collaborative analysis for remote geospatial visualization

using mixed reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2019), IEEE, pp. 236–247.

[141] MANN, S., FURNESS, T., YUAN, Y., IORIO, J., AND WANG, Z. All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. *arXiv preprint arXiv:1804.08386* (2018).

[142] MARQUES, B., SILVA, S., ALVES, J., ARAUJO, T., DIAS, P., AND SANTOS, B. S. A conceptual model and taxonomy for collaborative augmented reality. *IEEE transactions on visualization and computer graphics 28*, 12 (2021), 5113–5133.

[143] MCCARTHY, A., HARLEY, P., AND SMALLWOOD, R. Virtual arthroscopy training-do the" virtual skills" developed match the real skills required? In *Medicine meets virtual reality* (1999), IOS Press, pp. 221–227.

[144] MERINO, L., SCHWARZL, M., KRAUS, M., SEDLMAIR, M., SCHMALSTIEG, D., AND WEISKOPF, D. Evaluating mixed and augmented reality: A systematic literature review (2009-2019). *arXiv preprint arXiv:2010.05988* (2020).

[145] MEROLA, N., AND PE, J. The effects of avatar appearance in virtual worlds. *Journal For Virtual Worlds Research 2*, 5 (2009).

[146] MESH, M. Introducing microsoft mesh — connect like never before. https://www.microsoft.com/en-us/microsoft-teams/microsoft-mesh, 2024.

[147] MILGRAM, P., AND KISHINO, F. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems 77*, 12 (1994), 1321–1329.

[148] MINATANI, S., KITAHARA, I., KAMEDA, Y., AND OHTA, Y. Face-to-face tabletop remote collaboration in mixed reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (2007), IEEE, pp. 43–46.

[149] MOHR, P., MORI, S., LANGLOTZ, T., THOMAS, B. H., SCHMALSTIEG, D., AND KALKOFEN, D. Mixed reality light fields for interactive remote assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–12.

[150] MONAHAN, T. mclev-r: Design and evaluation of an interactive and collaborative m-learning application. *International Journal of Emerging Technologies in Learning (iJET) 2*, 2 (2007).

[151] MÜLLER, J., LANGLOTZ, T., AND REGENBRECHT, H. Panovc: Pervasive telepresence using mobile phones. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (2016), IEEE, pp. 1–10.

[152] MÜLLER, J., RÄDLE, R., AND REITERER, H. Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 6481–6486.

[153] NAGAI, S., KASAHARA, S., AND REKIMOTO, J. Livesphere: Sharing the surrounding visual environment for immersive experience in remote collaboration. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (2015), pp. 113–116.

[154] NARCISO, D., BESSA, M., MELO, M., COELHO, A., AND VASCONCELOS-RAPOSO, J. Immersive 360 video user experience: impact of different variables in the sense of presence and cybersickness. *Universal Access in the Information Society 18* (2019), 77–87.

[155] NGUYEN, C., DIVERDI, S., HERTZMANN, A., AND LIU, F. Collavr: collaborative in-headset review for vr video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (2017), pp. 267–277.

[156] NGUYEN, H., AND BEDNARZ, T. User experience in collaborative extended reality: Overview study. In *International Conference on Virtual Reality and Augmented Reality* (2020), Springer, pp. 41–70.

[157] NIEDERHAUSER, M., ALLSBROOK, W., ZANANIRI, E., AND FITZGERALD, J. Metamorphic: A social vr experience. In *ACM SIGGRAPH 2020 Immersive Pavilion*. ., 2020, pp. 1–2.

[158] ODA, O., ELVEZIO, C., SUKAN, M., FEINER, S., AND TVERSKY, B. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (2015), pp. 405–415.

[159] OHL, S. Tele-immersion concepts. *IEEE transactions on visualization and computer graphics 24*, 10 (2017), 2827–2842.

[160] OHSHIMA, T., SATOH, K., YAMAMOTO, H., AND TAMURA, H. Ar2 hockey. In *ACM SIGGRAPH 98 Conference abstracts and applications* (1998), p. 110.

[161] ORTS-ESCOLANO, S., RHEMANN, C., FANELLO, S., CHANG, W., KOWDLE, A., DEGTYAREV, Y., KIM, D., DAVIDSON, P. L., KHAMIS, S., DOU, M., ET AL. Holopor-

tation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology* (2016), pp. 741–754.

[162] OZTEMEL, E., AND GURSEV, S. Literature review of industry 4.0 and related technologies. *Journal of intelligent manufacturing 31*, 1 (2020), 127–182.

[163] PARADISO, J. A., AND LANDAY, J. A. Guest editors' introduction: Cross-reality environments. *IEEE Pervasive Computing 8*, 3 (2009), 14–15.

[164] PATEL, S., PANCHOTIYA, B., PATEL, A., BUDHARANI, A., AND RIBADIYA, S. A survey: virtual, augmented and mixed reality in education. *International Journal of Engineering Research & Technology (IJERT) 9*, 05 (2020), 2020.

[165] PEJOSKA-LAAJOLA, J., REPONEN, S., VIRNES, M., AND LEINONEN, T. Mobile augmented communication for remote collaboration in a physical work context. *Australasian Journal of Educational Technology 33*, 6 (2017).

[166] PEJSA, T., KANTOR, J., BENKO, H., OFEK, E., AND WILSON, A. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* (2016), pp. 1716–1725.

[167] PHILLIPS, K., AND PIEKARSKI, W. Possession techniques for interaction in real-time strategy augmented reality games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology* (2005), pp. 2–es.

[168] PIERCE, J. S., STEARNS, B. C., AND PAUSCH, R. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics* (1999), pp. 141–145.

[169] PIUMSOMBOON, T., DAY, A., ENS, B., LEE, Y., LEE, G., AND BILLINGHURST, M. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. ., 2017, pp. 1–5.

[170] PIUMSOMBOON, T., DEY, A., ENS, B., LEE, G., AND BILLINGHURST, M. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI 6* (2019), 5.

[171] PIUMSOMBOON, T., LEE, G. A., AND BILLINGHURST, M. Snow dome: A multi-scale interaction in mixed reality remote collaboration. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–4.

[172] PIUMSOMBOON, T., LEE, G. A., ENS, B., THOMAS, B. H., AND BILLINGHURST, M. Superman vs giant: A study on spatial perception for a multi-scale mixed reality flying telepresence interface. *IEEE transactions on visualization and computer graphics 24*, 11 (2018), 2974–2982.

[173] PIUMSOMBOON, T., LEE, G. A., HART, J. D., ENS, B., LINDEMAN, R. W., THOMAS, B. H., AND BILLINGHURST, M. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (2018), pp. 1–13.

[174] PIUMSOMBOON, T., LEE, G. A., IRLITTI, A., ENS, B., THOMAS, B. H., AND BILLINGHURST, M. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (2019), pp. 1–17.

[175] PIUMSOMBOON, T., LEE, Y., LEE, G. A., DEY, A., AND BILLINGHURST, M. Empathic mixed reality: Sharing what you feel and interacting with what you see. In *2017 international symposium on ubiquitous virtual reality (ISUVR)* (2017), IEEE, pp. 38–41.

[176] PODKOSOVA, I., VASYLEVSKA, K., SCHOENAUER, C., VONACH, E., FIKAR, P., BRONEDERK, E., AND KAUFMANN, H. Immersivedeck: A large-scale wireless vr system for multiple users. In *2016 IEEE 9th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)* (2016), IEEE, pp. 1–7.

[177] POELMAN, R., AKMAN, O., LUKOSCH, S., AND JONKER, P. As if being there: mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (2012), pp. 1267–1276.

[178] PONTO, K., SHIN, H. J., KOHLMANN, J., AND GLEICHER, M. Online real-time presentation of virtual experiences forexternal viewers. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology* (2012), pp. 45–52.

[179] POULIQUEN-LARDY, L., MILLEVILLE-PENNEL, I., GUILLAUME, F., AND MARS, F. Remote collaboration in virtual reality: asymmetrical effects of task distribution on spatial processing and mental workload. *Virtual Reality 20* (2016), 213–220.

[180] PRAVEENA, P., WANG, Y., SENFT, E., GLEICHER, M., AND MUTLU, B. Periscope: A robotic camera system to support remote physical collaboration. *Proceedings of the ACM on Human-Computer Interaction 7*, CSCW2 (2023), 1–39.

[181] RADIANTI, J., MAJCHRZAK, T. A., FROMM, J., AND WOHLGENANNT, I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education 147* (2020), 103778.

[182] RAMBACH, J., LILLIGREEN, G., SCHÄFER, A., BANKANAL, R., WIEBEL, A., AND STRICKER, D. A survey on applications of augmented, mixed and virtual reality for nature and environment. In *Virtual, Augmented and Mixed Reality: 13th International Conference, VAMR 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings* (2021), Springer, pp. 653–675.

[183] RAN, X., SLOCUM, C., GORLATOVA, M., AND CHEN, J. Sharear: Communication-efficient multi-user mobile augmented reality. In *Proceedings of the 18th ACM Workshop on Hot Topics in Networks* (2019), pp. 109–116.

[184] RASKAR, R., WELCH, G., CUTTS, M., LAKE, A., STESIN, L., AND FUCHS, H. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), pp. 179–188.

[185] RAUSCHNABEL, P. A., FELIX, R., HINSCH, C., SHAHAB, H., AND ALT, F. What is xr? towards a framework for augmented and virtual reality. *Computers in human behavior 133* (2022), 107289.

[186] REILLY, D., SALIMIAN, M., MACKAY, B., MATHIASEN, N., EDWARDS, W. K., AND FRANZ, J. Secspace: prototyping usable privacy and security for mixed reality collaborative environments. In *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems* (2014), pp. 273–282.

[187] REINHARDT, J., AND WOLF, K. Go-through: Disabling collision to access obstructed paths and open occluded views in social vr. In *Proceedings of the Augmented Humans International Conference* (2020), pp. 1–10.

[188] REITMAYR, G., AND SCHMALSTIEG, D. Mobile collaborative augmented reality. In *Proceedings IEEE and ACM International Symposium on Augmented Reality* (2001), IEEE, pp. 114–123.

[189] REN, P., QIAO, X., HUANG, Y., LIU, L., PU, C., DUSTDAR, S., AND CHEN, J.-L. Edge ar x5: An edge-assisted multi-user collaborative framework for mobile web augmented reality in 5g and beyond. *IEEE Transactions on Cloud Computing* (2020).

[190] RESEARCH, P. Extended reality market: Global industry analysis, size, share, growth, trends, regional outlook, and forecast (2022-2030). Precedence Research, 2024.

[191] RHEE, T., CHALMERS, A., HICKS, M., KUMAGAI, K., ALLEN, B., LOH, I., PETIKAM, L., AND ANJYO, K. Mr360 interactive: playing with digital creatures in 360° videos. In *SIGGRAPH Asia 2018 Virtual & Augmented Reality*. ACM, 2018, pp. 1–2.

[192] RHEE, T., PETIKAM, L., ALLEN, B., AND CHALMERS, A. Mr360: Mixed reality rendering for 360 panoramic videos. *IEEE transactions on visualization and computer graphics 23*, 4 (2017), 1379–1388.

[193] RHEE, T., THOMPSON, S., MEDEIROS, D., DOS ANJOS, R., AND CHALMERS, A. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE Transactions on Visualization and Computer Graphics 26*, 5 (2020), 1923–1933.

[194] ROKHSARITALEMI, S., SADEGHI-NIARAKI, A., AND CHOI, S.-M. A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences 10*, 2 (2020), 636.

[195] ROO, J. S., AND HACHET, M. Towards a hybrid space combining spatial augmented reality and virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)* (2017), IEEE, pp. 195–198.

[196] ROTH, D., YU, K., PANKRATZ, F., GORBACHEV, G., KELLER, A., LAZAROVICI, M., WILHELM, D., WEIDERT, S., NAVAB, N., AND ECK, U. Real-time mixed reality teleconsultation for intensive care units in pandemic situations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2021), IEEE, pp. 693–694.

[197] RUBIO-TAMAYO, J. L., GERTRUDIX BARRIO, M., AND GARCÍA GARCÍA, F. Immersive environments and virtual reality: Systematic review and advances in communication, interaction and simulation. *Multimodal technologies and interaction 1*, 4 (2017), 21.

[198] RYSKELDIEV, B., COHEN, M., AND HERDER, J. Streamspace: Pervasive mixed reality telepresence for remote collaboration on mobile devices. *Journal of Information Processing 26* (2018), 177–185.

[199] SABET, M., ORAND, M., AND W. MCDONALD, D. Designing telepresence drones to support synchronous, mid-air remote collaboration: An exploratory study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–17.

[200] SASIKUMAR, P., GAO, L., BAI, H., AND BILLINGHURST, M. Wearable remotefusion: A mixed reality remote collaboration system with local eye gaze and remote hand gesture sharing. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (2019), IEEE, pp. 393–394.

[201] SCHNACK, A., WRIGHT, M. J., AND HOLDERSHAW, J. L. Immersive virtual reality technology in a three-dimensional virtual simulated store: Investigating telepresence and usability. *Food Research International 117* (2019), 40–49.

[202] SCHOTT, D., SAALFELD, P., SCHMIDT, G., JOERES, F., BOEDECKER, C., HUETTL, F., LANG, H., HUBER, T., PREIM, B., AND HANSEN, C. A vr/ar environment for multi-user liver anatomy education. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (2021), IEEE, pp. 296–305.

[203] SCHROEDER, R. Being there together and the future of connected presence. *Presence 15*, 4 (2006), 438–454.

[204] SCHUBERT, T., FRIEDMANN, F., AND REGENBRECHT, H. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments 10* (2001), 266–281.

[205] SCHUBERT, T. W. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Z. für Medienpsychologie 15*, 2 (2003), 69–71.

[206] SERENO, M., WANG, X., BESANÇON, L., MCGUFFIN, M. J., AND ISENBERG, T. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics 28*, 6 (2020), 2530–2549.

[207] SHAKERI, H., AND NEUSTAEDTER, C. Teledrone: Shared outdoor exploration using telepresence drones. In *Conference companion publication of the 2019 on computer supported cooperative work and social computing* (2019), pp. 367–371.

[208] SIMPSON, B. D., BOLIA, R. S., AND DRAPER, M. H. Spatial audio display concepts supporting situation awareness for operators of unmanned aerial vehicles. *Human Performance, Situation Awareness, and Automation: Current Research and Trends HPSAA II, Volumes I and II 2* (2013), 61.

[209] SKARBEZ, R., BROOKS, JR, F. P., AND WHITTON, M. C. A survey of presence and related concepts. *ACM computing surveys (CSUR) 50*, 6 (2017), 1–39.

[210] Škola, F., Rizvić, S., Cozza, M., Barbieri, L., Bruno, F., Skarlatos, D., and Liarokapis, F. Virtual reality with 360-video storytelling in cultural heritage: Study of presence, engagement, and immersion. *Sensors 20*, 20 (2020), 5851.

[211] Skourup, C., Graven, T. G., and Husôy, K. Can visualisation improve remote collaboration? *IFAC Proceedings Volumes 40*, 16 (2007), 251–256.

[212] Sodhi, R. S., Jones, B. R., Forsyth, D., Bailey, B. P., and Maciocci, G. Bethere: 3d mobile collaboration with spatial input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), pp. 179–188.

[213] Sony. Sony 3d telepresence. https://www.tzmc.us/sony/3d_telepresence/index.htm, 2008.

[214] Spatial.io. Host your next event in spatial. https://support.spatial.io/hc/en-us/articles/4419855279892-Host-Your-Next-Event-In-Spatial, 2024. [Accessed: April 21, 2024].

[215] Speicher, M., Cao, J., Yu, A., Zhang, H., and Nebeling, M. 360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proceedings of the ACM on Human-Computer Interaction 2* (2018), 1–20.

[216] Sra, M., Garrido-Jurado, S., Schmandt, C., and Maes, P. Procedurally generated virtual reality from 3d reconstructed physical space. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (2016), pp. 191–200.

[217] Stafford, A., Piekarski, W., and Thomas, B. H. Implementation of god-like interaction techniques for supporting collaboration between outdoor ar and indoor tabletop users. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality* (2006), IEEE, pp. 165–172.

[218] Stafford, A., Thomas, B. H., and Piekarski, W. Efficiency of techniques for mixed-space collaborative navigation. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (2008), IEEE, pp. 181–182.

[219] Stanton, N. A., Chambers, P. R., and Piggott, J. Situational awareness and safety. *Safety science 39*, 3 (2001), 189–204.

[220] Stefik, M., Bobrow, D. G., Foster, G., Lanning, S., and Tatar, D. Wysiwis revised: early experiences with multiuser interfaces. *ACM Transactions on Information Systems (TOIS) 5*, 2 (1987), 147–167.

[221] STOAKLEY, R., CONWAY, M. J., AND PAUSCH, R. Virtual reality on a wim: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1995), pp. 265–272.

[222] SZELISKI, R., ET AL. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision 2*, 1 (2007), 1–104.

[223] TAIT, M., AND BILLINGHURST, M. The effect of view independence in a collaborative ar system. *Computer Supported Cooperative Work (CSCW) 24* (2015), 563–589.

[224] TAMURA, H., YAMAMOTO, H., AND KATAYAMA, A. Mixed reality: Future dreams seen at the border between real and virtual worlds. *IEEE Computer Graphics and Applications 21*, 6 (2001), 64–70.

[225] TAN, Z., HU, Y., AND XU, K. Virtual reality based immersive telepresence system for remote conversation and collaboration. In *Next Generation Computer Animation Techniques: Third International Workshop, AniNex 2017, Bournemouth, UK, June 22-23, 2017, Revised Selected Papers 3* (2017), Springer, pp. 234–247.

[226] TANASE, M., AND YANAGIDA, Y. Video stabilization for hmd-based telexistence-concept and prototype configuration. In *2015 IEEE/SICE International Symposium on System Integration (SII)* (2015), IEEE, pp. 106–111.

[227] TECCHIA, F., ALEM, L., AND HUANG, W. 3d helping hands: a gesture based mr system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry* (2012), pp. 323–328.

[228] TEO, T., A. LEE, G., BILLINGHURST, M., AND ADCOCK, M. 360drops: Mixed reality remote collaboration using 360 panoramas within the 3d scene. In *SIGGRAPH Asia 2019 Emerging Technologies*. ., 2019, pp. 1–2.

[229] TEO, T., F. HAYATI, A., A. LEE, G., BILLINGHURST, M., AND ADCOCK, M. A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes. In *25th ACM Symposium on Virtual Reality Software and Technology* (2019), pp. 1–11.

[230] TEO, T., LAWRENCE, L., LEE, G. A., BILLINGHURST, M., AND ADCOCK, M. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (2019), pp. 1–14.

[231] TEO, T., LEE, G. A., BILLINGHURST, M., AND ADCOCK, M. Investigating the use of different visual cues to improve social presence within a 360 mixed reality remote collaboration. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry* (2019), pp. 1–9.

[232] TEO, T., LEE, G. A., BILLINGHURST, M., AND ADCOCK, M. Merging live and static 360 panoramas inside a 3d scene for mixed reality remote collaboration. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (2019), IEEE, pp. 22–25.

[233] TEO, T., NORMAN, M., LEE, G. A., BILLINGHURST, M., AND ADCOCK, M. Exploring interaction techniques for 360 panoramas inside a 3d reconstructed scene for mixed reality remote collaboration. *Journal on Multimodal User Interfaces 14* (2020), 373–385.

[234] THANYADIT, S., PUNPONGSANON, P., AND PONG, T.-C. Observar: Visualization system for observing virtual reality users using augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2019), IEEE, pp. 258–268.

[235] THOMASON, J., RATSAMEE, P., KIYOKAWA, K., KRIANGKOMOL, P., ORLOSKY, J., MASHITA, T., URANISHI, Y., AND TAKEMURA, H. Adaptive view management for drone teleoperation in complex 3d structures. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (2017), pp. 419–426.

[236] THOMSEN, L. A., NILSSON, N. C., NORDAHL, R., AND LOHMANN, B. Asymmetric collaboration in virtual reality: A taxonomy of asymmetric interfaces for collaborative immersive learning. *Tidsskriftet Læring Og Medier (LOM) 12*, 20 (2019).

[237] THORAVI KUMARAVEL, B., ANDERSON, F., FITZMAURICE, G., HARTMANN, B., AND GROSSMAN, T. Loki: Facilitating remote instruction of physical tasks using bidirectional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (2019), pp. 161–174.

[238] THORAVI KUMARAVEL, B., NGUYEN, C., DIVERDI, S., AND HARTMANN, B. Transceivr: Bridging asymmetrical communication between vr users and external collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (2020), pp. 182–195.

[239] TIAN, H., LEE, G. A., BAI, H., AND BILLINGHURST, M. Using virtual replicas to improve mixed reality remote collaboration. *IEEE Transactions on Visualization and Computer Graphics 29*, 5 (2023), 2785–2795.

[240] TONG, W., XIA, M., WONG, K. K., BOWMAN, D. A., PONG, T.-C., QU, H., AND YANG, Y. Towards an understanding of distributed asymmetric collaborative visualization on problem-solving. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)* (2023), IEEE, pp. 387–397.

[241] VEAS, E., MULLONI, A., KRUIJFF, E., REGENBRECHT, H., AND SCHMALSTIEG, D. Techniques for view transition in multi-camera outdoor environments. In *Proceedings of Graphics Interface 2010* (2010), Citeseer, pp. 193–200.

[242] VENERELLA, J., SHERPA, L., TANG, H., AND ZHUI, Z. A lightweight mobile remote collaboration using mixed reality. In *Proceedings of Computer Vision and Pattern Recognition* (2019), pp. 1–4.

[243] WALLACE, J. R., OJI, S., AND ANSLOW, C. Technologies, methods, and values: Changes in empirical research at cscw 1990-2015. *Proceedings of the ACM on Human-Computer Interaction 1*, CSCW (2017), 1–18.

[244] WALLIS, K., AND ROSS, M. Fourth vr: Indigenous virtual reality practice. *Convergence 27*, 2 (2021), 313–329.

[245] WANG, H., AND SHI, B. E. Gaze awareness improves collaboration efficiency in a collaborative assembly task. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (2019), pp. 1–5.

[246] WANG, J., AND QI, Y. A multi-user collaborative ar system for industrial applications. *Sensors 22*, 4 (2022), 1319.

[247] WANG, L., WU, W., ZHOU, Z., AND POPESCU, V. View splicing for effective vr collaboration. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2020), IEEE, pp. 509–519.

[248] WANG, M. Social vr: A new form of social communication in the future or a beautiful illusion? In *Journal of Physics: Conference Series* (2020), vol. 1518, IOP Publishing, p. 012032.

[249] WANG, P., BAI, X., BILLINGHURST, M., ZHANG, S., WEI, S., XU, G., HE, W., ZHANG, X., AND ZHANG, J. 3dgam: using 3d gesture and cad models for train-

ing on mixed reality remote collaboration. *Multimedia Tools and Applications* (2020), 1–26.

[250] WANG, P., BAI, X., BILLINGHURST, M., ZHANG, S., WEI, S., XU, G., HE, W., ZHANG, X., AND ZHANG, J. 3dgam: using 3d gesture and cad models for training on mixed reality remote collaboration. *Multimedia Tools and Applications 80* (2021), 31059–31084.

[251] WANG, P., BAI, X., BILLINGHURST, M., ZHANG, S., ZHANG, X., WANG, S., HE, W., YAN, Y., AND JI, H. Ar/mr remote collaboration on physical tasks: a review. *Robotics and Computer-Integrated Manufacturing 72* (2021), 102071.

[252] WANG, P., ZHANG, S., BAI, X., BILLINGHURST, M., HE, W., SUN, M., CHEN, Y., LV, H., AND JI, H. 2.5 dhands: a gesture-based mr remote collaborative platform. *The International Journal of Advanced Manufacturing Technology 102*, 5 (2019), 1339–1353.

[253] WANG, P., ZHANG, S., BILLINGHURST, M., BAI, X., HE, W., WANG, S., SUN, M., AND ZHANG, X. A comprehensive survey of ar/mr-based co-design in manufacturing. *Engineering with Computers 36* (2020), 1715–1738.

[254] WANG, X., LOVE, P. E., KIM, M. J., AND WANG, W. Mutual awareness in collaborative design: An augmented reality integrated telepresence system. *Computers in Industry 65*, 2 (2014), 314–324.

[255] WANG, Z., ZHANG, S., AND BAI, X. A mixed reality platform for assembly assistance based on gaze interaction in industry. *The International Journal of Advanced Manufacturing Technology 116*, 9 (2021), 3193–3205.

[256] WELLS, T., AND HOUBEN, S. Collabar–investigating the mediating role of mobile ar interfaces on co-located group collaboration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–13.

[257] WELSFORD-ACKROYD, F., CHALMERS, A., KUFFNER DOS ANJOS, R., MEDEIROS, D., KIM, H., AND RHEE, T. Spectator view: Enabling asymmetric interaction between hmd wearers and spectators with a large display. *Proceedings of the ACM on Human-Computer Interaction 5*, ISS (2021), 1–17.

[258] WOBBROCK, J. O., FINDLATER, L., GERGLE, D., AND HIGGINS, J. J. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2011), pp. 143–146.

[259] WOLFF, R., ROBERTS, D. J., STEED, A., AND OTTO, O. A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology 29*, 1 (2007), 11–26.

[260] XIA, H., HERSCHER, S., PERLIN, K., AND WIGDOR, D. Spacetime: Enabling fluid individual and collaborative editing in virtual reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (2018), pp. 853–866.

[261] XIONG, Y., AND PULLI, K. Fast panorama stitching for high-quality panoramic images on mobile phones. *IEEE Transactions on Consumer Electronics 56*, 2 (2010), 298–306.

[262] XUE, T., LI, J., CHEN, G., AND CESAR, P. A social vr clinic for knee arthritis patients with haptics. In *Proceedings of the ACM International Conference on Interactive Media Experience* (2020).

[263] YANG, J., SASIKUMAR, P., BAI, H., BARDE, A., SÖRÖS, G., AND BILLINGHURST, M. The effects of spatial auditory and visual cues on mixed reality remote collaboration. *Journal on Multimodal User Interfaces 14* (2020), 337–352.

[264] YANG, L., HOLTZ, D., JAFFE, S., SURI, S., SINHA, S., WESTON, J., JOYCE, C., SHAH, N., SHERMAN, K., HECHT, B., ET AL. The effects of remote work on collaboration among information workers. *Nature human behaviour 6*, 1 (2022), 43–54.

[265] YAQOOB, A., BI, T., AND MUNTEAN, G.-M. A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. *IEEE Communications Surveys & Tutorials 22*, 4 (2020), 2801–2838.

[266] YOUNG, J., LANGLOTZ, T., COOK, M., MILLS, S., AND REGENBRECHT, H. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE transactions on visualization and computer graphics 25* (2019), 1908–1918.

[267] YOUNG, J., LANGLOTZ, T., MILLS, S., AND REGENBRECHT, H. Mobileportation: Nomadic telepresence for mobile devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4*, 2 (2020), 1–16.

[268] YOUNG, J., THOMPSON, S., DOWNER, H., ALLEN, B., PANTIDI, N., STOECKLEIN, L., AND RHEE, T. Telefest: Augmented virtual teleportation for live concerts. In *ACM International Conference on Interactive Media Experiences* (2022), pp. 69–78.

[269] ZAMAN, F., ANSLOW, C., CHALMERS, A., AND RHEE, T. Mrmac: Mixed reality multi-user asymmetric collaboration. In *22nd IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2023), IEEE, pp. 498–508.

[270] ZAMAN, F., ANSLOW, C., AND RHEE, T. Vicarious: Context-aware viewpoints selection for mixed reality collaboration. In *29th ACM Symposium on Virtual Reality Software and Technology (VRST)* (2023), pp. 1–11.

[271] ZHANG, C., CAI, Q., CHOU, P. A., ZHANG, Z., AND MARTIN-BRUALLA, R. Viewport: A distributed, immersive teleconferencing system with infrared dot pattern. *IEEE MultiMedia 20*, 1 (2013), 17–27.

[272] ZHANG, F., AND LIU, F. Casual stereoscopic panorama stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 2002–2010.

[273] ZHANG, W., HAN, B., HUI, P., GOPALAKRISHNAN, V., ZAVESKY, E., AND QIAN, F. Cars: Collaborative augmented reality for socialization. In *Proceedings of the 19th International Workshop on Mobile computing Systems & Applications* (2018), pp. 25–30.

[274] ZHANG, X., BAI, X., ZHANG, S., HE, W., WANG, P., WANG, Z., YAN, Y., AND YU, Q. Real-time 3d video-based mr remote collaboration using gesture cues and virtual replicas. *The International Journal of Advanced Manufacturing Technology 121*, 11-12 (2022), 7697–7719.

[275] ZHOU, Q., BHATTACHARYA, A., SURAWORACHET, W., NAGAHARA, H., AND CUKUROVA, M. Automated detection of students' gaze interactions in collaborative learning videos: A novel approach. In *European Conference on Technology Enhanced Learning* (2023), Springer, pp. 504–517.

[276] 'ASHER' ROSPIGLIOSI, P. Metaverse or simulacra? roblox, minecraft, meta and the turn to virtual reality for education, socialisation and work. *Interactive Learning Environments 30*, 1 (2022), 1–3.

Glossary

egocentric An egocentric viewpoint is a perspective that is centered around the observer or the self. In this viewpoint, the observer sees the world from their own point of view or position. It's like looking at the world from your own eyes, so everything is oriented relative to your own position and orientation. Egocentric perception is closely tied to a first-person perspective.. 120

exocentric An exocentric viewpoint is a perspective that is not centered around the observer but is based on an external reference point. In this viewpoint, the observer sees the world as if looking at it from an external, third-person perspective. It's like viewing the world from a fixed point outside of oneself, which may not be your actual physical location.. 120

situational awareness The ability to perceive, comprehend, and adapt to changing elements within a shared environment to make informed decisions in real-time.. 66

social presence The degree to which conversation feels unmediated; the sense of "being there with another".. 15

spatial presence The degree to which a user feels physically located within a real or virtual environment; the sense of "being there" rather than viewing the space through some surrogate.. 15

telepresence A technology that allows a person to feel present, as if they were physically there, in a remote location through the use of audio, video, and other communication means.. 2