MRMAC: Mixed Reality Multi-user Asymmetric Collaboration

Faisal Zaman [*]
CMIC, Victoria University of Wellington

Craig Anslow [†]
ECS, Victoria University of Wellington

Andrew Chalmers [‡]
CMIC, Victoria University of Wellington

Taehyun Rhee [§]
CMIC, Victoria University of Wellington

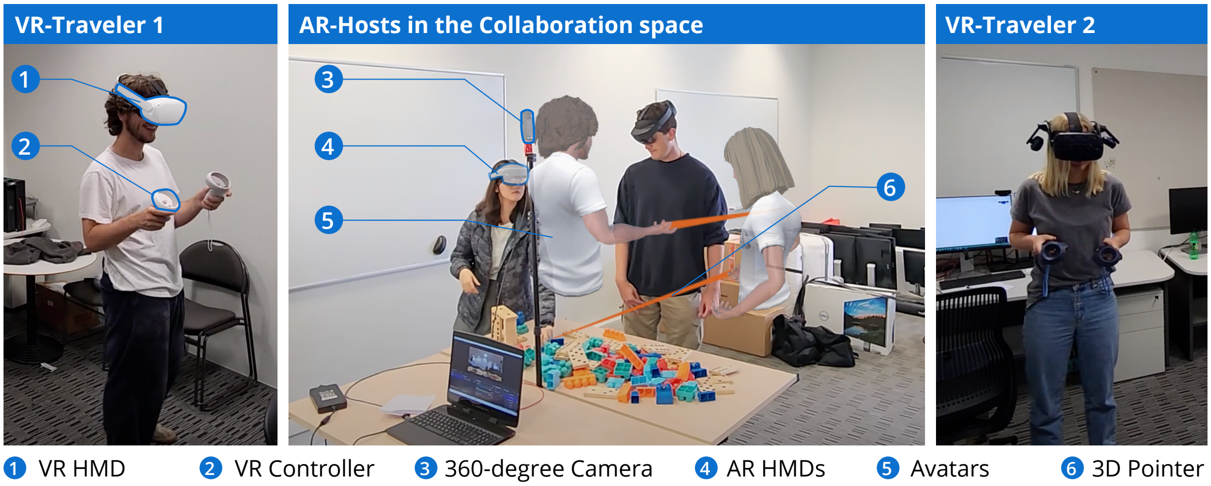

Figure 1: The MRMAC system, including the local AR hosts and remote VR travelers, 360° camera, controls, and display setup.

Abstract

We present 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. Our system enables telepresence for remote users by live-streaming the physical environment of local users using a 360° camera while blending 3D virtual assets into the mixed-reality collaboration space. Our novel client-server architecture enables asymmetric collaboration for multiple AR and VR users and incorporates avatars, view controls, as well as synchronized low-latency audio, video, and asset streaming. We evaluated our implementation with two baseline conditions: conventional 2D and standard 360° videoconferencing. Results show that MRMAC outperformed both baselines in inducing a sense of presence, improving task performance, usability, and overall user preference, demonstrating its potential for immersive multi-user telecollaboration.

Keywords: Mixed Reality, Telecollaboration, Telepresence

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 [45]. While prior work focused on virtual environments created using computer-generated 3D assets [58, 63], recent research has transitioned towards integrating users' real physical environments, thus enabling telecollaboration within real-world task spaces [54, 67, 75].

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 [55]. 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 [35, 55, 56]. To create such a system, it's crucial to capture the local users' physical 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 [1, 57] or using 360° cameras [55, 64, 66]. 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 [27, 76]. 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.

We present MRMAC (Figure 1), a Mixed Reality Multi-user Asymmetric Collaboration system that enables multiple remote users to virtually teleport into a real-world task space to collaborate with local users. Our work includes a design framework, key features supporting multi-user protocols, and a novel client-server architecture that facilitates asymmetrical telecollaboration for multiple AR and VR users. We evaluated MRMAC's effectiveness through a user study (N = 36), comparing it with two baseline conditions: conventional 2D and standard 360° videoconferencing.

The contributions of this paper are 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 among remote and local users.
- The **MRMAC implementation**, including a novel client-server architecture enables asymmetric telecollaboration for multiple AR and VR users with avatars, view controls, and low-latency video and asset streaming with synchronization.
- A **system evaluation and user study** of MRMAC in multi-user telecollaboration scenarios demonstrates its low-latency synchronized communication for multiple AR and VR users.

[*]e-mail: faisal.zaman@vuw.ac.nz
[†]e-mail: craig.anslow@vuw.ac.nz
[‡]e-mail: andrew.chalmers@vuw.ac.nz
[§]e-mail: taehyun.rhee@vuw.ac.nz

Our user study also shows that MRMAC significantly outperforms two baseline conditions in terms of spatial and social presence, usability, and remote collaboration performance.

2 Related work

2.1 Multi-user Collaboration in VR

Researchers have explored multi-user VR collaboration in various settings, encompassing co-located, remote, and hybrid environments that involve both co-located and remote users. For instance, DIVE [11] enabled users to engage in a shared 3D virtual world while maintaining a high level of situational awareness. Systems like MMVR [41] and mCLEV-R [42] have been applied across diverse domains, such as conferences, presentations, prototyping, and design. Furthermore, collaboration in VR has also shown useful in applications including communication and collaboration [71], education and training [51], and therapeutic interventions [73].

Social VR platforms like Facebook Horizon [3], Metamorphic [43], and CocoVerse [23] have emerged as popular tools for multi-user VR collaboration, allowing users to build, explore, socialize, and learn in virtual worlds. Systems such as TransceiVR [68] have demonstrated the potential for synchronous interaction for both co-located and remote users, allowing them to engage seamlessly. Similarly, ShareVR [24] and CoVR [14] enable external observers to independently view and interact with virtual environments through projection or touch displays. However, multi-user VR collaboration systems do exhibit certain limitations, including confinement to the virtual task space, lack of awareness of users' physical surroundings, and low situational and contextual awareness [51].

2.2 Multi-user Collaboration in AR

AR collaboration provides a natural and immersive way for users within a physical task space to engage with remote users [9, 77]. Multi-user AR collaboration, facilitating interaction between local and remote users, has demonstrated its effectiveness in enhancing remote telepresence through video conferencing [34], training [70], and remote assistance [22, 28]. In addition, addressing cross-platform multi-user synchronization issues [52] has helped reduce spatial inconsistency and latency in remote multi-user AR scenarios [53]. These advancements underline the potential of remote AR collaboration for natural interactions and foster unique collaborative opportunities.

Examples like AR2 Hockey [46] and AquaGauntlet [62] have focused on aspects of social communication, enabling co-located users to interact with virtual objects and characters. Lukosch et al. [39] emphasize the significance of designing AR systems that support natural interactions and enhance the collaborative experience. Collaborative AR has shown to be useful across diverse application domains, such as architectural and urban planning [6], education [4], and online meetings [7]. CollabAR [37] focuses on handheld group collaboration within a co-located environment.

However, while collaboration in VR provides remote awareness, AR technology offers limited understanding from one side of the collaboration [40]. Asymmetric setups that merge VR and AR technologies aim to overcome these limitations by offering accurate representations of remote users and the environment, resulting in a higher sense of presence [55, 65].

2.3 Asymmetric Collaboration in MR

Asymmetric MR collaboration involves users taking on different roles and locations within a collaborative task, utilizing different device settings in the collaborative environment (host/guest) [18]. The asymmetric MR collaboration has been shown to be helpful in remote collaboration in a physical task space, where a remote VR user interacts with a local AR user using hand gestures [20, 48], annotations [21], gaze [49, 69], virtual replicas of task objects [44], haptics [50], and more [16, 29]. Kiyokawa et al. [30] and Billinghurst et al. [8] presented mixed-space asymmetric collaboration systems allowing users to transition between AR and VR seamlessly. These mixed perspectives have been helpful in various applications such as engineering prototyping [13] and architectural design [26].

Live 360° video is becoming popular for remote collaboration [31, 64]. Poly [32] mounted a 360° panoramic camera on a backpack monopod to show the wearer's surroundings, allowing for real-time remote collaboration on a 360° panorama. Lee et al. [36] enable hand gestures and real-time viewing of awareness cues in a 360° panorama captured by a head-mounted camera.

Recent research has presented collaborative systems designed for real-world environments [35, 55]. Various techniques have been developed, including the procedural generation of virtual worlds from 3D-reconstructed physical spaces [61]. Real-time, high-quality 3D reconstructions of physical spaces developed by Orts-Escolano et al. [47], enable low-latency communication between remote users as if they were co-present in the same physical space.

3 MRMAC Design

We present a novel multi-user AR-VR asymmetric collaboration system that supports multiple local and remote users, each with distinct communication protocols and collaborative activities.

3.1 Asymmetric Telecollaboration System

The asymmetric setup is often employed in remote collaboration scenarios due to variations in user roles, physical locations, and the task environment. An AR-VR asymmetric telecollaboration system facilitates distinct setups that combine AR and VR to enable telecollaboration between local users (AR-Hosts) and remote users (VR-Travelers). This system allows them to interact and collaborate within a mixed reality collaboration (MRC) space [55], serving as a remote meeting point where AR-Hosts and VR-Travelers interact in real-time using verbal and visual cues. In a recent implementation of the asymmetric telecollaboration system [55], the VR-Traveler gains a sense of presence in the physical location of the AR-Host through live streaming of 360° video, while the AR-Host perceives the remote users as 3D avatars through AR-HMD. While VR-Travelers cannot physically interact with objects in the task space, they have the ability to observe the remote physical environment in real-time and use augmented visual cues to guide AR-Hosts. This guidance enables AR-Hosts to make physical changes [13, 28]. Thus, the AR-VR asymmetric system enhances communication and telepresence for remote users, enabling more effective collaboration in physical spaces compared to traditional video conferencing tools.

3.2 Multi-User Asymmetric Collaboration (MAC)

Multi-User Asymmetric Collaboration involves multiple individuals working together on a task but with varying access, permissions, and responsibilities. This paper extends AR-VR asymmetric telecollaboration to support multiple local and remote users collaborating in real-world tasks. We provide synchronized verbal and visual communication, enabling users to experience seamless collaboration as if they are physically co-present in the same task space. To achieve this, we focus on three key components.

Awareness: In AR-VR asymmetric telecollaboration, ensuring accurate awareness cues for remote users is essential, as these users are not physically present within the task space [33]. Examples of such cues include panoramic views of the task space [55], multiple 3D avatar representations [74], and view sharing [17, 31].

Communication: Effective communication in remote collaboration requires both verbal and visual cues [17]. Visual cues, such as annotation, pointers, gaze, and hand movements, play a crucial role in enabling remote users to compensate for their limited physical access to the task space [2, 19].

Streaming and Synchronization: Further, we need to ensure synchronized communication among multiple users with low-latency network streaming and synchronization. This involves seamless audio and video streaming across the network and synchronizing devices and data to ensure a consistent user experience.

4 MRMAC IMPLEMENTATION

4.1 Architecture Overview

We present a novel client-server architecture to implement MRMAC that addresses the three key design features outlined in section 3.2. Our system provides communication and awareness cues, allowing for 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.

VR-Traveler site: VR-Travelers receive the live-streamed 360° video from the local user, allowing them to immerse themselves in the local user's site through the live-streamed 360° video displayed in their VR HMDs. Their perspective views of the 360° video and visual cues are then shared with AR-Hosts and other VR-Travelers. These VR-Travelers are represented as 3D avatars, communicating and interacting with other users through visual cues.

The system architecture and data flow are illustrated in Figure 2. Our solution is fully integrated into the networking environment and implemented using the Unity 3D game engine with WebXR Exporter. We used Google's open-source Web Real-Time Communication protocol (WebRTC) [38] for networking on a dedicated NodeJS server that receives application requests over HTTP requests.

4.2 Awareness

MRMAC provides remote situational awareness by live streaming 360° video of the local space, with augmented visual cues for remote collaboration. It enhances multi-user presence and awareness by supporting multiple 3D avatars and view sharing.

4.2.1 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 1). However, controlling the movements of remote avatars, positioning them within the MRC space, and preserving their distinct identities present challenges.

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). 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 other user displays.

Avatar positioning: Placing multiple remote users' avatars in the center of the 360° camera can cause issues where their avatars can overlap, causing difficulties in identifying them during collaboration. To address that, we position multiple avatars in a circle around the camera's center, placing each avatar at a different position along this circular perimeter. The radius of the circle determines the distance from the camera, and the arrangement of the avatars follows a clockwise order with a specified angular offset between each subsequent avatar. Note that the first VR-Traveler to join the session is a special case, placed at the center camera position with no offset. We experimentally define the radius as 2m and the angular offset as 30 degrees. If the perimeter is at capacity with many avatars, we generate a second perimeter with 1.5 times the radius distance and continue the clockwise offset pattern. 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 our experiments, we carefully selected an offset sufficient to separate the avatars from each other while keeping it small enough to ensure that this difference in positioning was not noticeable to the participants.

Personalized avatar generation: The avatars were generated in real-time using Ready Player Me [1] and the Headshot plugin [2], with a Python wrapper to facilitate the process. Remote users took profile pictures that were sent to our API, which generated personalized avatar configurations based on parameters such as LOD, texture size, mesh, and facial features.

4.2.2 View Sharing

The view-sharing feature allows collaborators to see each other's perspectives, which is helpful for discussing spatial distribution. Both remote and local users can share their views through a picture-in-picture (PiP) window displayed on a 2D plane. The PiP window can be moved to any relative position and orientation to avoid head movement interference.

4.3 Communication

4.3.1 Verbal cues

Voice chat features were implemented 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 for Unity is employed to implement audio spatialization.

4.3.2 Visual cues

We integrated visual cues to enhance user communication and collaboration. We share three visual cues from the remote to the local side (see Figure 3):

Annotation: We used hand tracking through VR controllers to enable free-form 3D drawing annotations. To annotate, users create a pointer that extends from the controller's position to a plane surface, generating a mesh object at the point of collision. This mesh object serves as a canvas for drawing, and its position and orientation are continuously updated in real time to match the movements of the controller. Spatial alignment is not utilized for annotations, as they are rendered at the VR user's hand controller position when being drawn. Annotations were not spatially aligned since they are rendered at the position of the VR user's hand controller when drawn. As a result, annotations appear at VR users' arm's length.

3D pointer: The 3D pointer is added to indicate spatial targets or points of interest within the 360° video. This is implemented by projecting a raycast from the remote user's VR controller in the direction it's pointing. The direction of the pointer is determined by the VR controller's position and orientation. As the user moves their VR controller, the pointer updates accordingly. Since each avatar has offset compensation, we apply a corresponding offset to the base of the pointer. This ensures that the 3D pointer remains anchored to the avatar's hand, even if the avatar's position has been shifted. Then, a raycast is projected from the adjusted position of the avatar's hand to the original location in the video that was pointed to by VR-Travelers

[1] https://readyplayer.me/

[2] https://www.reallusion.com/character-creator/headshot/

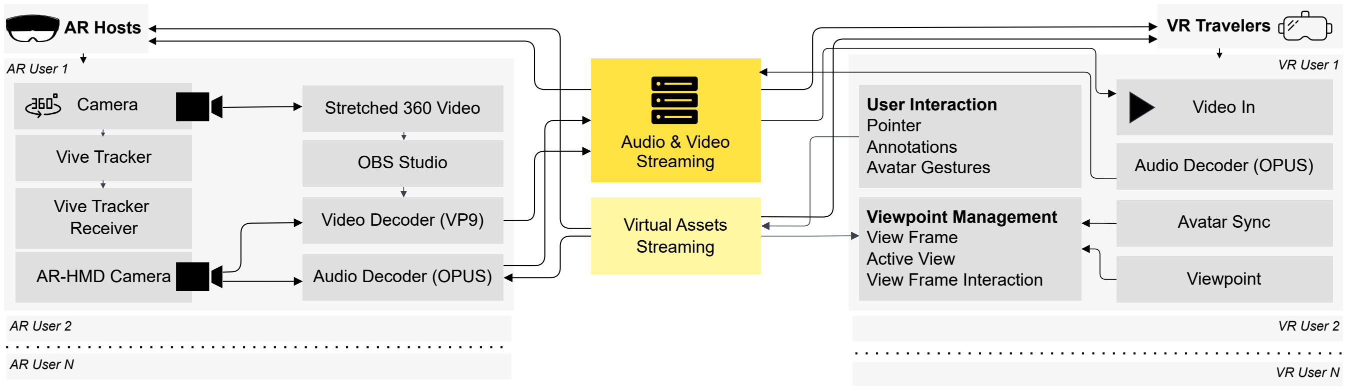

Figure 2: MRMAC: Overview of the system architecture and the network data flow, demonstrating the mapping between each component.

from the center of the camera. This makes the tip of the 3D pointer retain its initial intended position, regardless of any shifts in the base due to avatar adjustments. To maintain the precision of placing 3D annotation into the 360° video we approximate the real-world environment with a ground plane (either placed on the floor or a flat surface such as a table). This allows the pointer to intersect with the plane. Users can point at real objects by pointing at the base of the object - where the pointer will intersect the plane directly underneath the object in the video. The surrounding walls are also geometrically generated by manually setting a room size distance. We found this simple geometric approximation was sufficient for pointing since depth data were not available.

Hand gestures: VR-Travelers' hand gestures are also tracked using VR controllers, and their transformation is used to control avatars' 3D hands. The system also includes predefined gestures, each corresponding to specific actions or poses that users can initiate using their VR controllers.

To calibrate visual cues, we track the position and rotation of AR users' HMD in relation to the fixed 360° camera. This provides distance information between AR users and the camera, acting as a reference point in the shared space. To maintain spatial consistency during interactions such as pointing and annotating, we leverage this tracking data along with avatar position and rotation offset to render visual cues that align with the avatars. These aligned visual cues are then displayed to VR-Travelers and AR-Hosts.

4.4 Streaming and Synchronization

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 tracking data to each user while keeping latency to a minimum. Figure 2 provides an overview of the MRMAC system, including video streaming, data exchange between the server and client, and managing and distributing direct object manipulation.

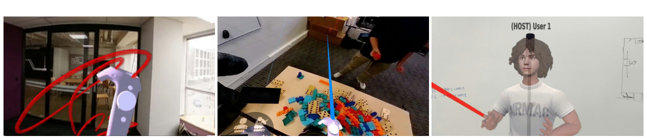

Figure 3: Communication and awareness cues in MRMAC: Annotation, 3D pointer, and avatar in collaborative environments.

4.4.1 Video and Audio Streaming

A 360° camera mounted at eye level in the physical space captures 4K video and encodes it using OBS Studio. To stream the video, we implemented the WebRTC SFU architecture using the Janus WebRTC plugin for Unity that supports VP9 video and OPUS audio codecs and is deployed on a local server. STUN and TURN servers are also implemented to facilitate NAT and firewall traversal. While direct peer-to-peer connections are the preferred option for establishing communication between the server and Unity, some network configurations may block such connections. STUN and TURN servers allow for fallback options to establish connectivity. We also implemented a relay service 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 allows for seamless communication between clients and the server, it can introduce additional latency and reduce the quality of the communication. Therefore, we prioritize establishing direct peer-to-peer connections whenever possible and use the relay service as a fallback option.

4.4.2 Virtual Assets Streaming

Non-media data is exchanged between clients using a Node.js [3] server with Socket.io [4], 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.

4.4.3 Synchronization

Our synchronization techniques maintain a consistent state across all connected devices to ensure seamless collaboration among multiple clients. This involves creating a new pair of threads for each received client of a successful connection. A main server thread handles all interactions between the client and the server during the collaboration process. Whenever a user launches a drawing annotation or pointers to any object, the system automatically connects to all clients running on the same panel and synchronizes each change in the position of objects and avatars with all connected clients. This

[3] https://nodejs.org/
[4] https://socket.io/

allows participants to see the avatar's movement and annotation, not just its final position. Absolute 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.

5 System Evaluation

We evaluated the performance of MRMAC by breaking down each component. We then assess the scalability of the system by ensuring low-latency synchronized data streaming to support multi-user telecollaboration. We measured the time it takes for changes to appear in the VR/AR window after a user action. Data is recorded when the 360° video is first viewable on the local computer and when viewed on the remote computer in Unity. We also used Unity profiler tools to measure latency at different stages of the rendering pipeline and within the AR/VR components.

5.1 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 (Figure 1).

5.2 System Evaluation Results

The performance metrics include the time taken in milliseconds for 360° video capturing, 360° video processing, encoding/decoding, transmission, network latency, and Interaction Latency. We measured the latency in two parts: 1) we simulated an increasing number of VR Travellers (from 1 to 15) with a fixed AR Host, and 2) we fixed a VR Traveller and simulated an increasing number of AR Hosts (from 1 to 15). The result is shown in Figure 4. As more local and remote users were added, the latency rate remained relatively stable and linear in terms of data and interaction synchronization. Even for 15 remote users, the network latency remains under 1 second (650ms). Additionally, we conducted simulations with a scenario involving four users, two local and two remote. Each measurement was averaged over ten samples. Table 1 presents the average end-to-end latency for each process. These results were consistent across multiple measurements, with negligible standard deviations.

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 audio and visual synchronization. The setup was tested over a LAN wired with a 1 Gbps Cat6 Ethernet cable. Both VR and AR users connected to the same network, although AR users connected via the Archer WiFi 6 Router. The average data transfer rate was measured at 16.45KBps (σ = .58). Of this, 1.54KBps (σ = .31) is sent by the local user, and 3.76KBps (σ = .16) by a remote user. The remaining 11.20 KBps (σ = .47)

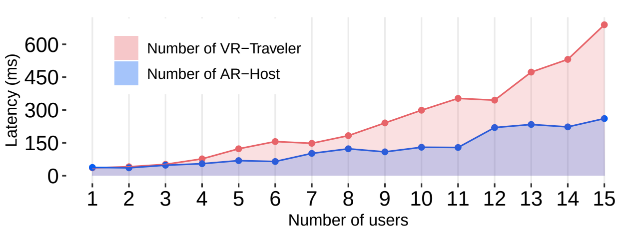

Figure 4: Scalability of MRMAC's network latency as the number of users (VR Travelers, AR Hosts) increases.

Table 1: System Performance Measurements

	VR-Traveler	AR-Host
360° Video Capture	75ms	×
Encoding/Decoding	40ms	49ms
Network Transmission	110ms	78ms
Network Latency	120ms	117ms
Interaction Latency*	42ms	69ms
Total Latency	387ms	313ms

*Data related to annotation, 3D pointer, and avatars.

is for transmitting the low-res video stream (480p at 15fps using the VP9 codec) from the remote user to the local user. Streaming 4K video average bandwidth was ~ 4Mbps. To avoid complications related to NAT, firewalls, and authentication, we tested the system on a LAN connection. User authentication was outside the scope of the current prototype. During the limited pilot test on a 5G hotspot, we deployed the signaling server and WebRTC SFU on a remote server. We found negligible performance differences, suggesting that our implementation can also function effectively in more challenging environments. Packet loss for 4K streaming was minimal ($< 1\%$). In WebRTC, 3 channels were established: OPUS-encoded audio, VP9-encoded video, and raw data streams. An identification tag is embedded in two 8×8 pixel regions in the top left corner of each image for synchronization. The displaced pixels and matching identification tags were stored in the data packet.

6 User Study

We conducted a user study (N = 36) to evaluate the key design features: communication and awareness. We compared MRMAC to two other experimental conditions: conventional 2D and standard 360-degree videoconferencing (Figure 5).

6.1 Design and Methodology

We employed a 3×2 mixed factorial design, 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 consists of four participants, two physically located together (AR-Hosts) and two remotely located (VR-Travelers).

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.

Experimental conditions and setup: Our study compares three experimental conditions (Figure 5):

- *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.
- *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) of the AR-Hosts space. On the other hand, AR-Hosts use a desktop computer to view the perspective view of the VR-Travelers.

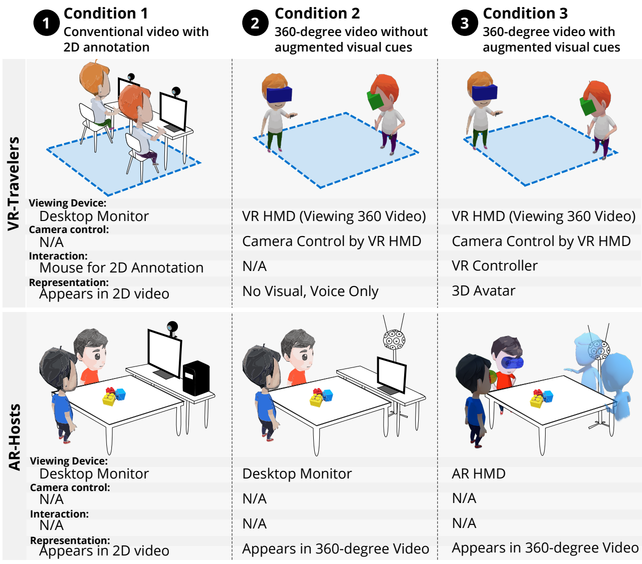

Figure 5: Overview of the User Study.

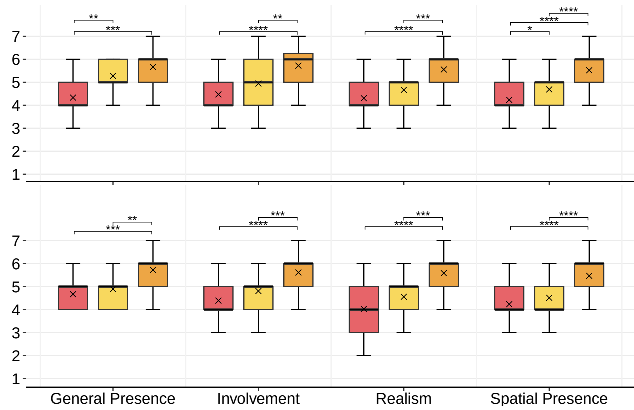

Figure 6: Spatial Presence results (C1 C2 C3).

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 [15, 60].

- *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.

Measures and Hypotheses: We evaluate the effectiveness of participants' communication and awareness of remote collaboration by 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 [59]) with four subscales: General Presence (GP), Realism (RL), Involvement (INV), and Spatial Presence (SP). For social presence, we compiled a questionnaire with four subscales: Behavioral Engagement (BE), Co-Presence(CP), Mutual Attention (MA), and Mutual Understanding (MU); based on Biocca et al. [10], Bale [5], and Hauber [25]. 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 were asked to rank their preferred condition under various categories and provide 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 C3 than in C1 and C2.

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

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

H4 Participants would prefer condition C3 over C1 and C2.

Participants: We recruited 36 participants aged 18-81 years (μ = 30.83, σ = 14.10), with 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. Most participants, 27 (75%) reported English as their primary language. Of the participants, 22 (61.1%) reported no prior VR/AR experience. All participants reported normal or corrected vision and hearing.

6.2 User Study Results

We assessed data normality using the Shapiro-Wilk test (α = 0.05). For normally distributed data, a two-way mixed ANOVA was conducted, followed by Tukey's HSD post hoc test (α = 0.05). Non-parametric analysis underwent a two-way mixed ANOVA with the Aligned Rank Transformation (ART), including the ART-C procedure [72]. 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 6 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 C3 resulted in significantly higher levels of immersion and connectedness to the remote environment than C1 and C2.

Social Presence: The average social presence rating is shown in Figure 7. We observed higher overall social presence ratings for condition C3 (μ = 5.80, σ = 0.89) than C1(μ = 4.20, σ = 0.87) and C2(μ = 4.84, σ = 0.84); also remote users rated slightly higher than local users (Remote: μ = 5.81, σ = 0.89; Local: μ = 5.78, σ = 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 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.

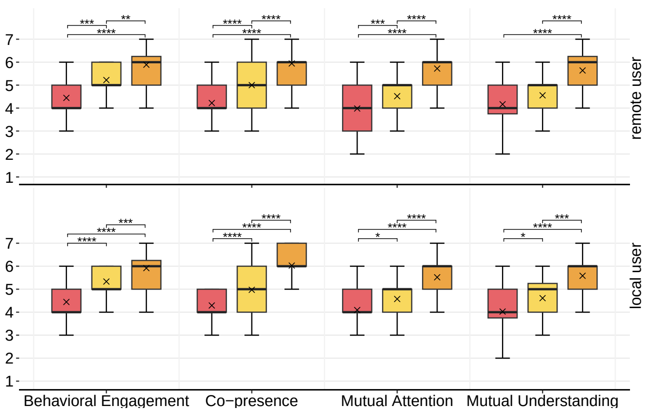

Figure 7: Social Presence results (C1 C2 C3).

Task Performance: The average task completion time for each of the three conditions shown in Figure 9 (b) 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, we randomly assigned participants to different conditions to randomly distribute any learning effect and help reduce its impact. Figure 8 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). 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 the lowest perceived overall workload than both C1 ($p < .001$) and C2 ($p = .0004$).

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) ($p < .001$, $p < .001$, $p < .001$). 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 9(a) shows the average SUS score for each condition and role. Participants in the C3 provided significantly higher ratings (Local: μ = 84.44, σ = 7.50; Remote: μ = 85.83, σ = 7.02) compared to the C2 (Local: μ = 75.27, σ = 9.62; Remote: μ = 72.36, σ = 8.33) and 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 10. Most of the participants ranked the C3 condition as their preferred choice. 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

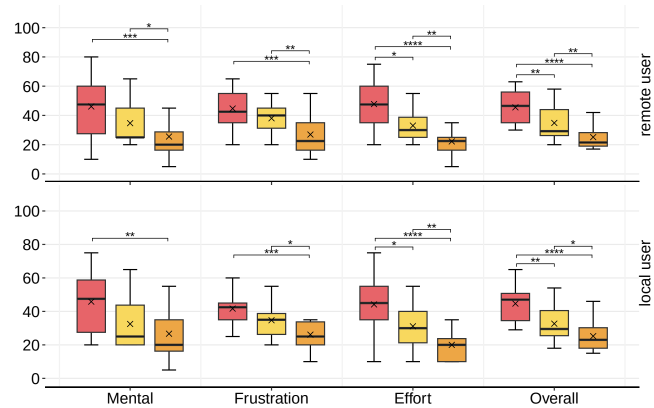

Figure 8: NASA-TLX score results (0: low workload - 100: high workload; lower scores are better). (C1 C2 C3).

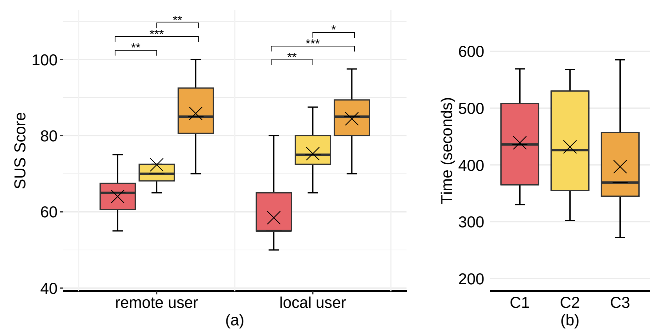

Figure 9: SUS results and task completion time. C1 C2 C3

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

7 Discussion

This paper showcases the first working system that supports multi-user mixed reality telecollaboration in an asymmetric setup with low latency data synchronization. Therefore, we were unable to conduct side-by-side system performance comparisons with prior works that only supported one-to-one collaboration scenarios [48,49,55,64].

System Performance: MRMAC 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

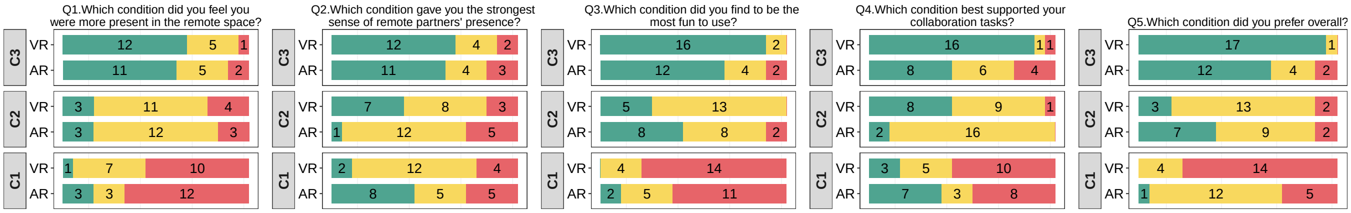

Figure 10: Users ranked the three conditions based on preference for presence, partner presence, fun, task support, and overall preference (Q1-Q5). Rank1 Rank2 Rank3 (Rank1 = most preferred rank, and C3 = MRMAC).

performance can be further improved through software optimization and advanced network infrastructure with higher bandwidth. However, our prototype shows promising results given that human participants typically perceive conversations as synchronized when the latency is below 250ms [12]. 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.

User Study: Our study shows that MRMAC (C3) improves collaboration in a multiuser MR remote collaboration compared to C2 and C1. Our first hypothesis was that users would experience significantly higher spatial and social presence in condition C3 compared to the other two baseline conditions. The experimental results support this hypothesis, as C3 had a higher score on the overall IPQ and spatial presence subscales compared to C1 and C2. This indicates that differences in communication and awareness cues in each condition likely contributed to these variations, thus supporting our hypothesis **H1**. Participants in C3 reported increased social presence compared to C1 and C2, reinforcing our hypothesis **H1**. The use of shared workspaces in C3 allowed users to interact with each other's avatars, fostering a stronger sense of social presence. View sharing also enabled participants to see where other participants were looking, creating a more natural and engaging collaborative environment, thereby enhancing social presence. Participants in C3 also completed tasks more quickly than those in C1 and C2, implying that C3's additional features facilitated smoother interactions and more efficient task execution, supporting **H2**. Our findings also support **H3**, indicating that system usability was significantly higher in C3 than in C1 and C2. The mean SUS scores indicate that users preferred the comprehensive feature set in MRMAC, including 360° video, annotations, and visual cues. 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.

Feedback and Observations: 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"*. 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 even more realistic"*. Voice communication was equally crucial for VR pairs to discuss problems and coordinate actions in real-time.

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 us from testing with larger groups. Although we have made efforts to minimize latency, this can be improved for larger collaboration scenarios. Furthermore, while compression standards offer extensions that can reduce the overall bandwidth, these extensions are primarily designed for camera arrays and do not have 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. Finally, while our system provides high audio-visual fidelity for remote collaboration using 360° video, the lack of depth perception may hinder certain collaborative tasks that require precise spatial understanding or mixed-reality collision handling.

Although we positioned multiple avatars using circular offset, it has limitations to accommodate many users in the large space. Multiple 360° cameras will cover the wide area while providing space for larger groups. Exploring optimal positioning between avatars and cameras can also be explored. We also aim to enable depth streaming for better spatial understanding and 6-DoF movement.

8 Conclusion

We presented MRMAC, a novel Mixed Reality Multi-user Asymmetric Collaboration system combining live-streaming 360° video with mixed-reality displays. We assessed our system by conducting performance evaluations and user studies. The performance results showcased the technical capabilities of the system, including high frame rates, low latency, and scalability. In contrast, the user study results demonstrated MRMAC's effectiveness in enhancing communication and awareness, while also inducing a higher level of spatial and social presence among participants in mixed-reality environments. Furthermore, our findings revealed that the system reduced the difficulty of collaborative work, leading to a lower workload and enabling participants to perform more efficiently.

MRMAC showed strong potential in addressing the needs of multi-user asymmetric remote collaboration by enabling an interactive, cross-device communication platform. We believe the design implications of our work will help improve video conferencing software by incorporating telepresence and virtual teleportation elements, producing a better user experience for mixed-reality multi-user asymmetric collaboration.

Acknowledgments

This work was supported by the Entrepreneurial University Programme from the Tertiary Education Commission of New Zealand.

REFERENCES

[1] M. Adcock, S. Anderson, and B. Thomas. Remotefusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH international conference on VR continuum and its applications in industry*, pages 235–242, 2013. 1

[2] D. Akkil, J. M. James, P. Isokoski, and J. Kangas. Gazetorch: enabling gaze awareness in collaborative physical tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016. 2

[3] P. 'asher' Rospigliosi. Metaverse or simulacra? roblox, minecraft, meta and the turn to virtual reality for education, socialisation and work. *Interactive Learning Environments*, 30(1):1–3, 2022. 2

[4] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001. 2

[5] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich. 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*, 14:379–393, 2005. 6

[6] E. Ben-Joseph, H. Ishii, J. Underkoffler, B. Piper, and L. Yeung. Urban simulation and the luminous planning table: Bridging the gap between the digital and the tangible. *Journal of planning Education and Research*, 21(2):196–203, 2001. 2

[7] M. Billinghurst, D. Belcher, A. Gupta, and K. Kiyokawa. Communication behaviors in colocated collaborative ar interfaces. *International Journal of Human-Computer Interaction*, 16(3):395–423, 2003. 2

[8] M. Billinghurst, H. Kato, and I. Poupyrev. The magicbook: a transitional ar interface. *Computers & Graphics*, 2001. 2

[9] M. Billinghurst and B. H. Thomas. Mobile collaborative augmented reality. In *Recent trends of mobile collaborative augmented reality systems*, pages 1–19. Springer, 2011. 2

[10] F. Biocca, C. Harms, and J. K. Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments*, 12:456–480, 2003. 6

[11] C. Carlsson and O. Hagsand. Dive a multi-user virtual reality system. In *Proceedings of IEEE virtual reality annual international symposium*, pages 394–400. IEEE, 1993. 2

[12] Y. Chen, T. Farley, and N. Ye. Qos requirements of network applications on the internet. *Information Knowledge Systems Management*, 4(1):55–76, 2004. 8

[13] D. Clergeaud, J. S. Roo, M. Hachet, and P. Guitton. Towards seamless interaction between physical and virtual locations for asymmetric collaboration. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pages 1–4, 2017. 2

[14] J. Du, Y. Shi, Z. Zou, and D. Zhao. Covr: Cloud-based multiuser virtual reality headset system for project communication of remote users. *Journal of Construction Engineering and Management*, 2018. 2

[15] R. Eiris, M. Gheisari, and B. Esmaeili. Desktop-based safety training using 360-degree panorama and static virtual reality techniques: A comparative experimental study. *Automation in construction*, 2020. 6

[16] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98, 2019. 2

[17] M. Faridan, B. Kumari, and R. Suzuki. 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. 2

[18] N. Feld and B. Weyers. Overview of collaborative virtual environments using augmented reality. *Mensch und Computer*, 2019. 2

[19] S. R. Fussell, L. D. Setlock, E. M. Parker, and J. Yang. 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*, pages 788–789, 2003. 2

[20] L. Gao, H. Bai, G. Lee, and M. Billinghurst. An oriented point-cloud view for mr remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, pages 1–4, 2016. 2

[21] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 449–459, 2014. 2

[22] R. Grasset and J.-D. Gascuel. Mare: Multiuser augmented reality environment on table setup. In *ACM SIGGRAPH 2002 conference abstracts and applications*, pages 213–213, 2002. 2

[23] S. W. Greenwald, W. Corning, and P. Maes. Multi-user framework for collaboration and co-creation in virtual reality. In *Multi-user framework for collaboration and co-creation in virtual reality*. 12th International Conference on Computer Supported Collaborative Learning, 2017. 2

[24] J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio. 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*, pages 4021–4033, 2017. 2

[25] J. Hauber, H. Regenbrecht, M. Billinghurst, and A. Cockburn. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 413–422, 2006. 6

[26] H. Ibayashi, Y. Sugiura, D. Sakamoto, N. Miyata, M. Tada, T. Okuma, T. Kurata, M. Mochimaru, and T. Igarashi. Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In *SIGGRAPH Asia 2015 Emerging Technologies*, pages 1–2, 2015. 2

[27] M. M. Jack, D. A. Gattozzi, P. J. Camarata, and K. J. Shah. Live-streaming surgery for medical student education-educational solutions in neurosurgery during the covid-19 pandemic. *Journal of surgical education*, 78(1):99–103, 2021. 1

[28] P. Jansen, F. Fischbach, J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio. 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*, pages 459–471, 2020. 2

[29] S. Kasahara and J. Rekimoto. Jackin: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th augmented human international conference*, pages 1–8, 2014. 2

[30] K. Kiyokawa, H. Takemura, and N. Yokoya. 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)*, volume 6, pages 48–53. IEEE, 1999. 2

[31] R. Komiyama, T. Miyaki, and J. Rekimoto. 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*, pages 1–9, 2017. 2

[32] S. Kratz, D. Kimber, W. Su, G. Gordon, and D. Severns. Polly: " being there" through the parrot and a guide. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 625–630, 2014. 2

[33] M. Le Chénéchal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi. 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)*, pages 9–12. IEEE, 2016. 2

[34] G. A. Lee, S. Kim, Y. Lee, A. Dey, T. Piumsomboon, M. Norman, and M. Billinghurst. Improving collaboration in augmented video conference using mutually shared gaze. In *ICAT-EGVE*, 2017. 2

[35] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pages 1–4, 2017. 1, 2

[36] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. Sharedsphere: Mr collaboration through shared live panorama. In *SIGGRAPH Asia 2017 Emerging Technologies*, pages 1–2. SIGGRAPH, 2017. 2

[37] Z. Liu, G. Lan, J. Stojkovic, Y. Zhang, C. Joe-Wong, and M. Gorlatova. Collabar: Edge-assisted collaborative image recognition for mobile augmented reality. In *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2020. 2

[38] S. Loreto and S. P. Romano. *Real-time communication with WebRTC: peer-to-peer in the browser*. " O'Reilly Media, Inc.", 2014. 3

[39] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa. Collaboration in augmented reality. *Computer Supported Cooperative Work*, 2015. 2

[40] B. Marques, S. Silva, J. Alves, T. Araujo, P. Dias, and B. S. Santos. A conceptual model and taxonomy for collaborative augmented reality. *IEEE transactions on visualization and computer graphics*, 28(12):5113–5133, 2021. 2

[41] A. McCarthy, P. Harley, and R. Smallwood. Virtual arthroscopy training-do the" virtual skills" developed match the real skills required? In *Medicine meets virtual reality*, pages 221–227. IOS Press, 1999. 2

[42] T. Monahan. mclev-r: Design and evaluation of an interactive and collaborative m-learning application. *International Journal of Emerging Technologies in Learning (iJET)*, 2(2), 2007. 2

[43] M. Niederhauser, W. Allsbrook, E. Zananiri, and J. Fitzgerald. Metamorphic: A social vr experience. In *ACM SIGGRAPH 2020 Immersive Pavilion*, pages 1–2. ACM SIGGRAPH, 2020. 2

[44] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 405–415, 2015. 2

[45] S. Ohl. Tele-immersion concepts. *IEEE transactions on visualization and computer graphics*, 24(10):2827–2842, 2017. 1

[46] T. Ohshima, K. Satoh, H. Yamamoto, and H. Tamura. Ar2 hockey. In *ACM SIGGRAPH 98 Conference abstracts and applications*, 1998. 2

[47] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*, pages 741–754, 2016. 2

[48] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G. Lee, and M. Billinghurst. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pages 1–5. SIGGRAPH, 2017. 2, 7

[49] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 6:5, 2019. 2, 7

[50] I. Podkosova, K. Vasylevska, C. Schoenauer, E. Vonach, P. Fikar, E. Bronederk, and H. Kaufmann. 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)*, pages 1–7. IEEE, 2016. 2

[51] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147:103778, 2020. 2

[52] X. Ran, C. Slocum, M. Gorlatova, and J. Chen. Sharear: Communication-efficient multi-user mobile augmented reality. In *ACM Workshop on Hot Topics in Networks*, 2019. 2

[53] P. Ren, X. Qiao, Y. Huang, L. Liu, C. Pu, S. Dustdar, and J.-L. Chen. 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. 2

[54] T. Rhee, L. Petikam, B. Allen, and A. Chalmers. Mr360: Mixed reality rendering for 360 panoramic videos. *IEEE transactions on visualization and computer graphics*, 23(4):1379–1388, 2017. 1

[55] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE transactions on visualization and computer graphics*, 26:1923–1933, 2020. 1, 2, 7

[56] S. Rokhsaritalemi, A. Sadeghi-Niaraki, and S.-M. Choi. A review on mixed reality: Current trends, challenges and prospects. *Applied Sciences*, 10(2):636, 2020. 1

[57] D. Roth, K. Yu, F. Pankratz, G. Gorbachev, A. Keller, M. Lazarovici, D. Wilhelm, S. Weidert, N. Navab, and U. Eck. 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)*, pages 693–694. IEEE, 2021. 1

[58] A. Schnack, M. J. Wright, and J. L. Holdershaw. Immersive virtual reality technology in a three-dimensional virtual simulated store: Investigating telepresence and usability. *Food Research International*, 117:40–49, 2019. 1

[59] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10:266–281, 2001. 6

[60] F. Škola, S. Rizvić, M. Cozza, L. Barbieri, F. Bruno, D. Skarlatos, and F. Liarokapis. Virtual reality with 360-video storytelling in cultural heritage: Study of presence, engagement, and immersion. *Sensors*, 20(20):5851, 2020. 6

[61] M. Sra, S. Garrido-Jurado, C. Schmandt, and P. Maes. Procedurally generated virtual reality from 3d reconstructed physical space. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 191–200, 2016. 2

[62] H. Tamura, H. Yamamoto, and A. Katayama. Mixed reality: Future dreams seen at the border between real and virtual worlds. *IEEE Computer Graphics and Applications*, 21(6):64–70, 2001. 2

[63] Z. Tan, Y. Hu, and K. Xu. 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*, pages 234–247. Springer, 2017. 1

[64] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. 360drops: Mixed reality remote collaboration using 360 panoramas within the 3d scene. In *SIGGRAPH Asia*. SIGGRAPH, 2019. 1, 2, 7

[65] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock. A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes. In *25th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, 2019. 2

[66] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. Investigating the use of different visual cues to improve social presence within a 360 mixed reality remote collaboration. In *International Conference on Virtual-Reality Continuum and its Applications in Industry*, 2019. 1

[67] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 161–174, 2019. 1

[68] B. Thoravi Kumaravel, C. Nguyen, S. DiVerdi, and B. Hartmann. 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. 2

[69] H. Wang and B. E. Shi. Gaze awareness improves collaboration efficiency in a collaborative assembly task. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019. 2

[70] J. Wang and Y. Qi. A multi-user collaborative ar system for industrial applications. *Sensors*, 22(4):1319, 2022. 2

[71] M. Wang. Social vr: A new form of social communication in the future or a beautiful illusion? In *Journal of Physics: Conference Series*, volume 1518, page 012032. IOP Publishing, 2020. 2

[72] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 143–146, 2011. 6

[73] T. Xue, J. Li, G. Chen, and P. Cesar. A social vr clinic for knee arthritis patients with haptics. In *Proceedings of the ACM International Conference on Interactive Media Experience*, 2020. 2

[74] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 547–556. IEEE, 2019. 2

[75] J. Young, T. Langlotz, S. Mills, and H. Regenbrecht. Mobileportation: Nomadic telepresence for mobile devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020. 1

[76] J. Young, S. Thompson, H. Downer, B. Allen, N. Pantidi, L. Stoecklein, and T. Rhee. Telefest: Augmented virtual teleportation for live concerts. In *ACM International Conference on Interactive Media Experiences*, pages 69–78, 2022. 1

[77] W. Zhang, B. Han, P. Hui, V. Gopalakrishnan, E. Zavesky, and F. Qian. Cars: Collaborative augmented reality for socialization. In *Proceedings of the 19th International Workshop on Mobile computing Systems & Applications*, pages 25–30, 2018. 2