A Comparison Study Understanding the Impact of Mixed Reality Collaboration on Sense of Co-Presence

Jianing Yin*
University of Rochester
Tsinghua University

Weicheng Zheng†
Duke Kunshan University

Yuntao Wang‡
Tsinghua University

Xin Tong§
Hong Kong University of Science and Technology (Guangzhou)

Yukang Yan¶
University of Rochester

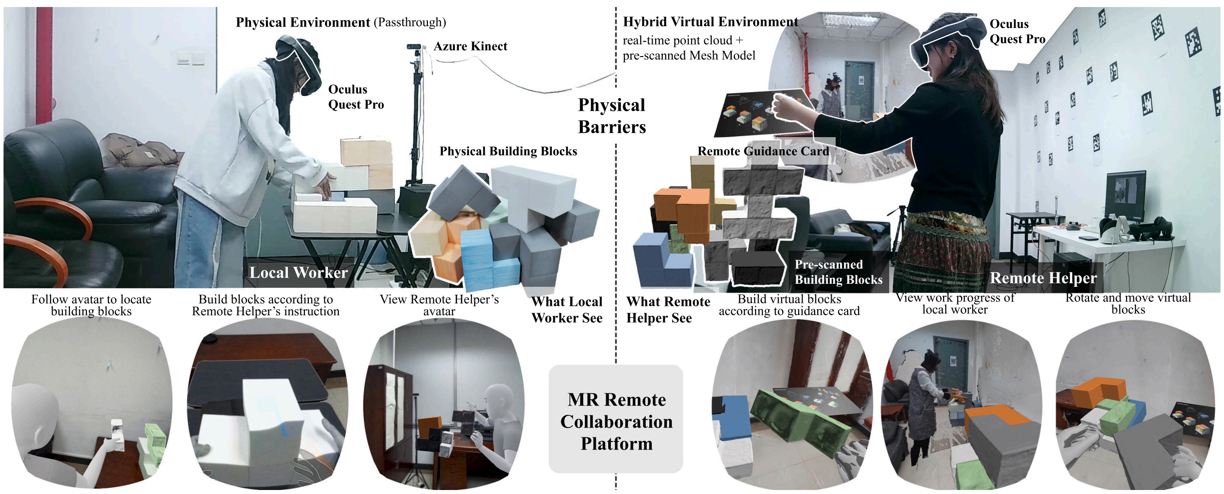

Figure 1: Experimental settings for the Mixed Reality condition in the study. Two roles are assigned to a dyad of participants: one collects and assembles the physical blocks as the **Local Worker** and the other gives instructions as the **Remote Helper**. Local Worker wears an Augmented Reality headset where they see the virtual avatar of Remote Helper in the local environment. Remote Helper wears a Virtual Reality headset where they see a scanned replica of the local environment and point cloud of Local Worker.

Abstract

Sense of co-presence refers to the perceived closeness and interaction between participants in a collaborative context, which critically impacts the collaboration experience and task performance. With the emergence of Mixed Reality (MR) technologies, we would like to investigate the effect of MR immersive collaboration environment on promoting co-presence in a remote setting by comparing it with non-MR methods, such as video conferencing. We conduct a comparison study, where we invited 14 dyads of participants to collaborate on block assembly tasks with video conferencing, MR system, and in a physically co-located scenario. Each participant of a dyad was assigned either a local worker to assemble the blocks or a remote helper to give the instructions. Results show that MR system can create comparable sense of co-presence with co-located situation, and allow users to interact more naturally with both the environment and each other. The adoption of mixed reality enhances collaboration and task performance by reducing reliance on verbal communication and favoring action-based interactions through gestures and direct manipulation of virtual objects.

Index Terms: Remote Collaboration, Co-presence, Mixed Reality.

*e-mail: catherineyin13@gmail.com. This work was done while Jianing Yin was a remote intern at University of Rochester.
†e-mail: jerrryzheng@gmail.com
‡e-mail: yuntaowang@tsinghua.edu.cn, corresponding author
§e-mail: xint@hkust-gz.edu.cn
¶e-mail: yukang.yan@rochester.edu

1 Introduction

Globalization has diversified work environments, spreading team members across different cultures and time zones [60], while technology advancements have shifted collaboration from physical co-location to remote, offering significant convenience. Popular remote collaboration tools, such as Zoom[1] and Skype[2], rely on 2D window-based interfaces, which limit spatial and depth perception [7], restrict non-verbal communication like gestures [41], gaze [75], and body language [38], crucial for face-to-face interactions. This limitation separates the physical task space from communication space [13], which significantly reduce the perceived co-presence [39] between the remote collaborators.

Co-presence refers to the sense of individuals feeling together in a virtual environment, becoming "accessible, available, and subject to one another" [19]. It is a key goal for various remote collaboration systems to reduce psychological distance, enhance connection, enthusiasm, and satisfaction in remote collaboration [68]. To achieve a high level of co-presence, researchers have developed systems leveraging immersive technologies, e.g., Mixed Reality (MR), as collaborators can be immersed in the same virtual environment and are able to communicate with each through full-body interactions [8, 81].

Our research investigates how different collaboration settings (co-location, video conferencing, MR collaboration) impact the sense of co-presence between two collaborators in a remote assistance scenario. We address three research questions: Is MR col-

[1]https://zoom.us/
[2]https://www.skype.com/en/

laboration more similar to co-located collaboration in terms of co-presence compared to video conferencing? **(RQ1)** Does MR enable behaviors more similar to co-located scenarios? **(RQ2)**What behaviors indicate the sense of co-presence?**(RQ3)**

We chose block assembly as the collaboration task. One collaborator will serve as the *local worker*, who is tasked to search for block pieces and assemble them to form a target pattern. The other collaborator will be the *remote helper* who has the instructions on the target pattern and the required pieces at hand. In the task, it requires *remote helper* to make sense of *local worker*'s physical environment to guide the search and the status of the blocks to instruct the assembly steps. The task also requires *local worker* to understand *remote helper*'s instructions and descriptions. In such a setting, the collaboration can be effective probe to test co-presence, as with optimal co-presence level, the collaborators should work as immersively as if they are in the same physical space.

To measure co-presence, we collected task completion time, subjective co-presence ratings, and also recorded communication behaviors, including speech and actions. The consideration is that behavioral features like deictic actions, word count, and confirmation methods can reflect how well the collaborators felt being in the same space. For example, when they were immersed in the task and forgot the remote nature of collaboration, they unintentionally pointed to a physical object in the environment, which may or may not be observed by the remote collaborator.

Our results show that metrics such as physical presence and psychological social presence in MR are closer to co-located collaboration than video conferencing, suggesting MR effectively enhances co-presence (**RQ1**). MR also facilitated communication behaviors (e.g., word count, demonstrative pronouns) that more closely resemble co-located scenarios (**RQ2**). Additionally, correlations between demonstrative pronouns, confirmation methods, task-related actions, and performance suggest these behaviors serve as indicators of co-presence (**RQ3**).

The main contributions of our work include:

- We provide empirical evidence that MR offers a sense of co-presence comparable to co-located scenarios, important for organizations transitioning to remote work.

- We identify behavioral indicators of co-presence, such as body language visibility and interactive speaking patterns, which can serve as objective metrics for future studies.

2 Related Work

2.1 Sense of Co-presence in MR Remote Collaboration

Sense of co-presence, originating from Goffman, refers to the feeling of being together in a virtual environment where individuals are "accessible, available, and subject to one another" [20]. It is further conceptualized as a psychological connection between minds [45] and plays a key role in enhancing the effectiveness of MR remote collaboration [12]. Enhancing co-presence improves spatial perception [67], communication efficiency [4], and emotional connection [57]. Co-presence includes two parts: physical presence, the sense of "being there," and social presence, the sense of "being together with another" [54, 24]. This study adopts a definition that combines physical and social presence, following Hein's conceptualization [24].

Previous research has identified several factors influencing co-presence in remote collaboration. Early studies focused on sharing video and verbal cues via phone and video conferencing, which proved inefficient and less immersive [49, 43] due to the lack of non-verbal cues [74, 1]. MR enhances remote collaboration by providing better spatial clarity and larger interaction spaces [15].

Recent research has shifted towards understanding and improving co-presence, focusing on: (1) the effects of individual and multi-modal communication cues, such as eye gaze [23] and facial expressions [37]; (2) sharing task space views, like 360-degree panoramas and real-time point clouds [70, 72]; and (3) human factors, including perception, cognition, and user profiles [34].

Co-presence is influenced by factors like input/output modalities and perspective sharing [15], which have been studied to improve collaboration. To verify improvements in MR remote collaboration, studies often use an unimproved MR system or video conferencing as a baseline [3, 66].

Based on related work, we formulated **RQ1** to compare MR remote collaboration with co-located collaboration in terms of co-presence, especially against traditional video conferencing. We aim to explore whether MR's immersive features simulate co-located interactions more effectively. To address **RQ1**, we conducted a comparative study of three collaboration settings and their impact on co-presence.

2.2 MR remote collaboration system setup

Remote collaboration systems are designed to bridge the gap between geographically distributed users, enabling seamless interaction through advanced technologies. A key focus has been on integrating physical and virtual elements to enhance collaboration. Frameworks like Partially Blended Realities align dissimilar physical and virtual spaces through adaptive environment mapping and shared spatial references, enabling distributed teams to collaborate effectively [21]. Similarly, systems such as the Immersive Interactive Virtual Cabin (IIVC) incorporate physical features of users' environments into virtual spaces, addressing tracking issues and improving situational awareness during collaborative tasks [14]. Building on this, SurfShare facilitates lightweight sharing of physical surfaces and spatially consistent virtual replicas, ensuring precise alignment between physical and virtual elements for activities like design reviews [26]. Multimodal interfaces, combining visual, auditory, and haptic feedback, have become essential for enabling intuitive communication and task execution [34].

To enhance co-presence, systems now focus on fostering natural interactions through features like dynamic avatars, spatial annotations, and shared virtual landmarks. These tools bridge the communication gap, align collaborators' focus, and reduce ambiguity, making them indispensable for spatial referencing and effective task coordination [40, 42, 52]. Telepresence technologies, such as life-size projections and 360-degree video sharing, further enhance co-presence by enabling natural, face-to-face-like interactions [50, 53].

Innovations like virtual replicas and tangible interaction tools bridge the gap between physical and virtual environments, supporting collaboration across different scales and complexities [46]. Awareness cues, such as gaze direction and gesture tracking, provide additional support by helping collaborators understand each other's focus and actions in real time [51].

Many MR remote collaboration systems leverage bi-directional collaboration to enable real-time, interactive participation across co-located and remote users. For instance, the Blended Whiteboard integrates physical affordances with MR flexibility to support dynamic workflows [22], while Loki employs bi-directional telepresence to enhance remote instruction through mutual communication, spatial alignment, and immersive interaction [71]. These systems highlight the widespread adoption of bi-directional approaches in fostering immersive, efficient, and equitable distributed teamwork.

Common MR remote collaboration setups integrate environmental information, shared virtual landmarks, multimodal interfaces, and virtual avatars to enhance co-presence, communication efficiency, and task coordination by blending physical and virtual environments.

2.3 Measuring sense of co-presence through behavioral observation

Current approaches to measuring co-presence fall into three categories: (1) physiological measurements, (2) behavioral observations, and (3) subjective measures [24]. Most research relies on subjective measures, such as questionnaires [55], accounting for 96.6% [65]. While subjective measures are easy to apply and low cost, they are prone to biases and may not capture real-time feelings [17]. Factors like memory, current emotional states [33], and questionnaire design [59] can also affect accuracy.

Behavioral observation provides an objective alternative by measuring postural and adaptive behaviors, offering more realistic responses to virtual environments [16, 48]. Virtual environments and the presence of others can influence user behaviors [78]. According to Clarke and Brennan's Grounding model, people continually exchange information to establish a common ground or shared understanding [10]. One source of common ground is physical co-presence, particularly having a shared visual space [11]. In task-space collaboration, sharing remote pointing or gesture cues is crucial for conversational grounding [36]. Studies have shown that behaviors, such as postural adjustments [62] and attention-based measures [44], are useful for assessing co-presence.

Recent investigations into MR co-presence predominantly rely on subjective measures, such as questionnaires, supplemented by completion time as an objective measure [40, 50]. However, a growing body of research is beginning to explore unique behavioral phenomena within MR remote collaboration. These include spatial expressions [42], indicative gestures [64], conversational efficiency [6], referencing behaviors of helpers [32, 42], gaze synchronization rates, and travel distances [51]. Despite these insights, there remains a notable gap in understanding how these behaviors directly relate to co-presence in MR environments.

Based on this, we formulated **RQ2** and **RQ3**, aiming to compare behavioral and speaking patterns across co-located, MR, and video conferencing scenarios, and identify behaviors and language features as indicators of co-presence.

3 Implementation

To enable the comparison of collaboration behaviors in different settings, we adopt existing software (Zoom) for video conferencing and build up our own MR collaboration system for Mixed Reality condition. In the remaining of this session, we describe the design and implementation of this system.

3.1 Design

The system (Figure 1) supports remote collaboration between a *local worker* and a *remote helper*. Based on Seungwon Kim et al.'s 2018 study, remote expert collaboration improves communication, focus, and enjoyment compared to mutual collaboration [35, 69], which guided our system and experiment design. The *local worker* performs physical tasks (Section 4), while the *remote helper* provides instructions. Given the asymmetry in tasks, we developed separate applications: an Augmented Reality (AR) app for the *local worker* to interact with their environment and receive AR guidance, and a Virtual Reality (VR) app for the *remote helper* to immerse in a virtual replica of the worker's environment. The two apps synchronize via WiFi and record behavioral data during collaboration.

3.2 AR application for local worker

As shown in Figure 1 (left), the *local worker* can see their physical environment with a virtual avatar representing the *remote helper*. Using the passthrough feature of Quest Pro, we achieve AR effects for the local worker while capturing the remote helper's position and body movements. Meta Avatars SDK [3] was used to create a neutral avatar showing the helper's head, upper body, facial expressions, and gestures, transmitted via Photon PUN2 [4].

In the task environment, the local worker manipulates nine differently shaped physical blocks. In the MR scenario, they also see virtual blocks operated by the remote helper, with positions and directions synced via Photon PUN2.

3.3 VR application for remote helper

To help the remote helper understand the task environment, we created a hybrid virtual environment by combining a pre-scanned environment mesh with a real-time point cloud model, providing real-time updates on the local worker's location (Figure 1, right). The scanning process used Agisoft Metashape[5] to generate a 3D model by capturing digital images of the task environment and creating a mesh using operations like Add Photos, Align Photos, and Build Mesh.

Individual blocks were scanned separately using a similar method. Real-time point clouds were captured with two Azure Kinect cameras using Azure Kinect SDK[6]. We used the Azure Kinect Examples for Unity package[7] to convert color and depth frames into point-cloud data. After calibrating and aligning spatial data from both cameras, we merged the point cloud model with the pre-scanned environment in Unity through manual adjustment.

3.4 Apparatus

The prototype system used two Oculus Quest Pro headsets, with the local worker using AR via passthrough and the remote helper using VR tethered to a desktop. Both AR and VR systems were developed in Unity 3D (2021.3.16c1), and Photon PUN2 was used to transmit spatial data such as the avatar's head, hands, and digital blocks. Furthermore, audio communication within the mixed reality (MR) collaboration system was facilitated using Photon Voice[8]. Two Microsoft Azure Kinect cameras were positioned at two corners of the workspace to maximize coverage. The cameras were configured to operate at a color resolution of 720p and the Wide Field-Of-View Unbinned mode for depth information, achieving a frame rate of 30 FPS..

4 Study Method

This study explores how MR remote collaboration compares to co-located collaboration and video conferencing in terms of perceived co-presence (**RQ1**), the resemblance of speaking and actions to co-located scenarios (**RQ2**), and identifying behaviors that indicate co-presence (**RQ3**).

In *Video Collaboration*, participants used a video meeting app, with the local worker holding a mobile device to navigate, while the remote helper joined via phone. The local worker could independently adjust the camera based on the task scenario and the remote helper's needs, similar to the way video conferencing software is used in daily life. For instance, the local worker moved the camera to locate blocks and focused on the workspace to confirm assembly accuracy with the remote helper. In *MR Collaboration*, the local worker wore an AR headset, and the remote helper used a VR headset. *Co-located Collaboration* served as a baseline, where both

[3]https://developer.oculus.com/documentation/unity/meta-avatars-overview/

[4]https://www.photonengine.com/zh-cn/pun

[5]https://www.agisoft.com/

[6]https://learn.microsoft.com/en-us/azure/kinect-dk/sensor-sdk-download

[7]https://assetstore.unity.com/packages/tools/integration/azure-kinect-examples-for-unity-149700

[8]https://doc.photonengine.com/zh-cn/voice/current/getting-started/voice-intro

participants were in the same room, communicating naturally. During the experiment, the remote helper assumed the role of a guide, while the local worker was responsible for executing the assembly tasks. To maintain consistency across all three experimental conditions and ensure that the local worker independently performed the assembly tasks, the remote helper was explicitly instructed to refrain from directly interacting with the blocks, even in the co-located setting.

4.1 Participants

We recruited 28 participants (18 female, 10 male) aged 19 to 29 (M = 23.14, SD = 2.45) to form 14 worker-helper dyads, with seven dyads knowing each other beforehand. Most participants were students from various backgrounds, including Architecture, Data Science, and Electronic Information.

26 participants had prior VR experience, and 13 had used AR devices. All participants provided informed consent, and audio/video recordings were made with consent. Participants received $14 per hour for completing the experiment.

4.2 Experiment Setup

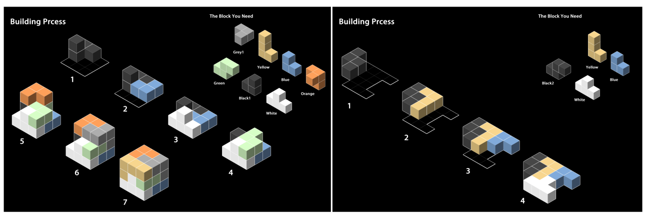

Figure 2: Guidance cards showing the specific blocks required for each task and step-by-step instructions for the assembly.

The experiment took place in two rooms, each approximately 2.5 by 5 meters. The local worker's room contained a bookcase, two sofas, a chair, and four desks, while the remote helper's room had a slightly different layout (Figure 3). In each scenario, collaborators completed two tasks with 9 blocks scattered around the room. The local worker had to find and assemble either 7 or 4 blocks while ignoring 2 or 5 unwanted blocks. The grey desk in the middle served as the assembly platform. In the remote helper's view, guidance cards appeared to instruct where to find and assemble the blocks. As illustrated in Figure 2, the guidance cards detailed the specific blocks required for each task and provided step-by-step instructions for their assembly. During the experiment, each dyad was tasked with completing seven block assembly activities guided by the instructions on the cards. These tasks included three assemblies consisting of seven blocks each and three assemblies consisting of four blocks each. To mitigate potential familiarity bias, the block assembly tasks were randomly distributed across co-located, MR, and video conferencing conditions. In the MR condition, the guidance cards were presented as virtual objects within the environment, while in the other two conditions, they were presented as physical cards. Additionally, the local worker was instructed not to directly view the contents of the guidance cards. One set of blocks was used during a preliminary training phase to familiarize participants with the task procedure and the equipment utilized in the experiment.

4.3 Collaboration Tasks

We selected block finding and building as the collaborative task, commonly used in remote collaboration studies [5, 23]. The task requires the local worker to seek guidance from the remote helper due to a lack of essential information, encouraging communication about progress and next steps: **Sub-task 1 - Block Finding:** The local worker collects scattered blocks, relying on the helper's guidance to identify the correct blocks while navigating interference from irrelevant blocks. **Sub-task 2 - Block Building:** The worker assembles the blocks into a predetermined shape on the table with guidance from the remote helper.

4.4 Procedure

We conducted within-group experiments across three scenarios. Each dyad completed two sub-tasks—Blocks Finding and Blocks Building—in video, MR, and co-located collaboration settings. The sub-task order was fixed, but the order of collaboration scenarios was randomized to reduce familiarity bias. Task object locations and final shapes were adjusted to ensure similar difficulty across scenarios.

Upon arrival, participants signed consent forms and completed a demographic questionnaire. Both remote helpers and local workers were trained on the Quest Pro. After training, workers were brought to the task environment and briefed, while helpers received instructions in a separate room. In the co-located scenario, both completed the task in the same room, with the worker performing operations and the helper providing guidance. After each scenario, participants completed questionnaires and brief interviews.

4.5 Measurements

We used questionnaires, audio and video recordings, and interviews for data analysis. The questionnaires assessed participants' perceived co-presence (physical and social) and workload during tasks, using a 7-point Likert scale (e.g., Strongly Disagree - 1; Strongly Agree - 7). Interviews and recordings were used to confirm questionnaire results and explore underlying reasons. Audio and video data were disassembled into behaviors to analyze their correlation with co-presence.

4.5.1 Physical Presence Measures

To measure the helper's immersion in the virtual environment, we combined the Slater-Usoh-Steed Questionnaire (SUS) [63, 73] and Presence Questionnaire (PQ) [79, 80] to assess their perception and control over the MR environment. SUS captures the subjective experience of presence, while PQ focuses on control and sensory fidelity, both essential for evaluating how well the MR environment mimics real-world interactions.

4.5.2 Social Presence Measures

Effective collaboration required communication between the worker and helper. We divided social presence into psychological and operational components. The Inclusion of Other in the Self (IOS) Scale [2] measured psychological closeness (psychological social presence), while the Social Presence Module of the Game Experience Questionnaire (GEQ) [28] assessed how much attention participants paid to their counterpart during tasks (operational social presence).

4.5.3 Interview

Semi-structured interviews were conducted to gain insights into participants' experiences, focusing on their perceptions of the environment, collaboration, task complexity, and speaking/action patterns across different scenarios and roles. We also asked what aspects of the MR collaboration scenario would make them feel more like collaborating in person.

4.5.4 Users' Behaviors

To assess participants' behaviors (**RQ2** and **RQ3**), we recorded audio and video during the experiments with participants' consent. Audio was captured using a recording pen on each subject's collar to analyze speech content and communication frequency. Video was recorded via a webcam in each room. The data were segmented

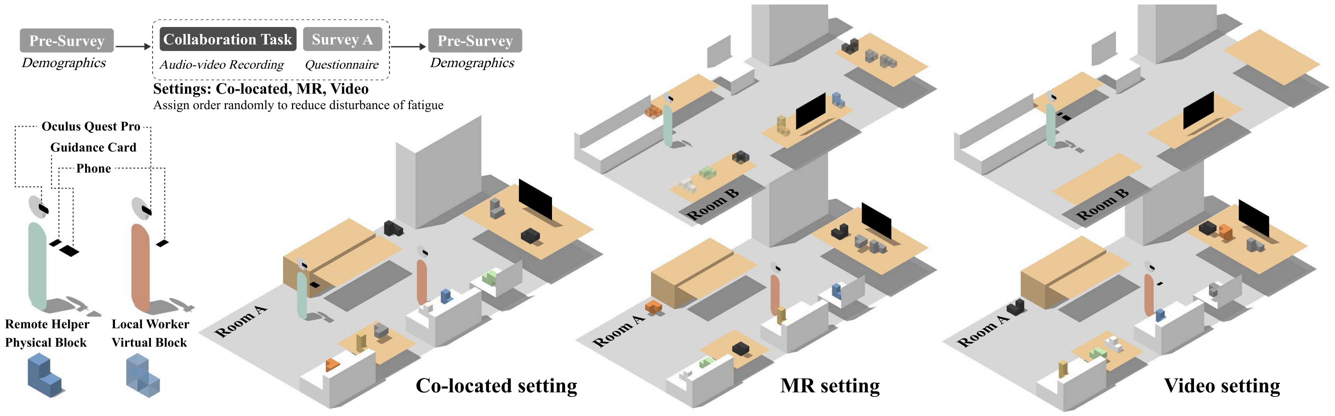

Figure 3: Experiment procedure: after filling pre-study surveys, the participants go through three collaboration conditions in a randomized order, followed by an exit interview.

into behaviors, such as communication frequency, use of demonstrative pronouns (audio), and the proportion of task-related and deictic actions (video).

4.5.5 Task Performance

Before the experiment, participants were informed that the task would end only upon correct completion. Thus, task completion time was used as the primary measure of performance, indicating that the blocks were assembled correctly.

4.6 Data Analysis

Before data analysis, we identified outliers based on task completion time for each group in each scenario. Outliers were defined as times exceeding three standard deviations from the scenario's average. Based on this, we excluded the first task of Group 2 and 11 in the co-located scenario, and the first task of Group 14 in the MR and video scenarios.

4.6.1 Questionnaire Data Analysis

For questionnaire analysis, we assessed physical and social presence. Physical presence was measured using average scores from the SUS (Slater-Usoh-Steed) and PQ (Presence Questionnaire) scales. Social presence was evaluated using the IOS (Inclusion of Other in the Self) and GEQ (Social Presence Module of the Game Experience Questionnaire) scales. IOS measures cognitive perception of psychological closeness, while GEQ evaluates operational social presence. IOS is a single-item scale, and GEQ scores were summed and averaged according to its guidelines.

4.6.2 Video Data Analysis

This study used both video and audio data to analyze the behaviors of remote helpers and local workers, aiming to comprehensively observe speaking and actions during remote collaboration. Below is the coding process for transforming the recordings into quantitative data.

Two researchers independently analyzed two sets of data from 14 groups, reviewing 12 videos across three scenarios. Weekly meetings helped resolve discrepancies and refine the initial codebook, categorizing actions into **task-related actions** and **deictic actions**. Task-related actions include searching, picking up, rotating, and moving blocks, essential for understanding object manipulation across scenarios. For remote helpers, deictic actions, like gesturing towards block positions and illustrating rotations, were key for effective non-verbal communication. This codebook was then applied to code the remaining 10 groups, further refining the analysis.

4.6.3 Audio Data Analysis

Audio recordings were transcribed into Chinese and proofread for accuracy. We analyzed the total number of words spoken during tasks and identified confirmation strategies, including **pre-question confirmations** (feedback before a question) and **after-question confirmations** (feedback after a question). We also examined the use of demonstrative pronouns "this" and "that," excluding non-referential instances. This approach enabled a precise assessment of speaking strategies and their role in task completion.

For parametric data, we first used the Shapiro-Wilk test to assess normality. If normality was violated, the Friedman test was applied; otherwise, Repeated Measures ANOVA was used. Mauchly's test examined sphericity, and if violated, the Greenhouse-Geisser correction was applied.

5 Results

5.1 Quantitative Data Analysis Result

5.1.1 Sense of Co-presence

For local workers, the Friedman test showed no significant differences in *physical presence* or *operational social presence*, but a significant difference in *psychological social presence* ($F = 3.376,\ p < 0.05$). LSD post-hoc revealed no significant difference between co-located and MR scenarios, but video conferencing had significantly lower *psychological presence* ($I - J = -1.286,\ p < 0.05$).

For remote helpers, significant differences were found in *physical*, *psychological*, and *operational social presence*. Video conferencing had significantly lower *physical presence* compared to co-located ($I - J = -1.551,\ p < 0.05$) and MR ($I - J = -1.327,\ p < 0.05$). *Operational presence* was lower in MR compared to co-located ($I - J = -0.976,\ p < 0.05$) and video conferencing ($I - J = -1.083,\ p < 0.05$). *Psychological presence* was also lower in video conferencing than in co-located ($I - J = -1.571,\ p < 0.05$).

These results indicate that MR aligns more closely with co-located scenarios in terms of perceived physical and social presence compared to video conferencing. However, MR shows a contrasting trend in operational and psychological social presence, as discussed in Section 6.

5.1.2 Action

In the co-located scenario, the Friedman test showed significant differences in task-related actions across the three scenarios ($F = 34.234,\ p < 0.05$). LSD post-hoc tests revealed that task-related actions in MR were significantly higher than in co-located ($I - J =$

18.750, $p < 0.05$) and video conferencing ($I-J = 19.000$, $p < 0.05$), likely due to the remote helper manipulating virtual blocks.

For the local worker, the Friedman test also showed significant differences ($F = 3.928$, $p < 0.05$). LSD post-hoc revealed fewer task-related actions in MR compared to co-located ($I-J = 14.643$, $p < 0.05$) and video conferencing ($I-J = 15.714$, $p < 0.05$), likely due to more effective guidance in MR, reducing trial and error.

We analyzed deictic actions by the remote helper across the three scenarios. The Friedman test revealed significant differences ($F = 11.031$, $p < 0.05$). LSD post-hoc tests showed deictic actions were significantly higher in the co-located scenario compared to MR ($I-J = 7.000$, $p < 0.05$) and video conferencing ($I-J = 7.571$, $p < 0.05$), likely because the remote helper prefers gestures when in the same environment.

5.1.3 Speaking

Total Number of Words Spoken.

For local workers, the Friedman test showed a significant difference across scenarios ($F = 6.164$, $p < 0.05$). LSD post-hoc results (Figure 4 III.1) showed no difference between co-located and MR, but significantly more words were spoken in video conferencing compared to co-located ($I-J = 144.640$, $p < 0.05$) and MR ($I-J = 107.360$, $p < 0.05$).

For remote helpers, the Friedman test also showed a significant difference ($F = 7.518$, $p < 0.05$). LSD post-hoc revealed no difference between co-located and MR, but significantly more words were spoken in video conferencing compared to co-located ($I-J = 478.040$, $p < 0.05$) and MR ($I-J = 493.200$, $p < 0.05$).

These results suggest that speaking performance in co-located and MR scenarios is more similar in terms of total words spoken.

In the co-located scenario, the Friedman test showed a significant difference in speaking frequency across scenarios ($F = 3.561$, $p < 0.05$). LSD post-hoc results revealed no difference between co-located and MR for local workers, but significantly higher speaking frequency in video conferencing compared to co-located ($I-J = 0.318$, $p = 0.011$).

For remote helpers, the Friedman test also showed significant differences ($F = 3.778$, $p < 0.05$). LSD post-hoc showed no difference between co-located and MR, but higher speaking frequency in video conferencing than MR ($I-J = 1.484$, $p < 0.05$).

These results suggest that speaking frequency is more similar between co-located and MR scenarios.

Confirmation Methods.

We quantified the number of confirmations made by local workers and remote helpers, including pre-question confirmations (helpers provide feedback before questions) and after-question confirmations (feedback after questions). The confirmation rate was calculated by dividing the total confirmations by the total words spoken.

The Friedman test showed significant differences in overall confirmation counts ($F = 5.731$, $p < 0.05$), pre-question confirmations ($F = 5.422$, $p < 0.05$), and after-question confirmations ($F = 4.797$, $p < 0.05$). LSD post-hoc tests found no difference between co-located and MR for overall counts or after-question confirmations. However, video conferencing had significantly higher confirmations than MR ($I-J = 5.240$, $p < 0.05$). Pre-question confirmations were lower in MR than in co-located ($I-J = -1.120$, $p < 0.05$) and video conferencing ($I-J = -1.640$, $p < 0.05$), while after-question confirmations were lower in MR than video conferencing ($I-J = -4.040$, $p < 0.05$).

The Friedman test showed significant differences in overall confirmation rates ($F = 3.574$, $p < 0.05$) and after-question confirmations ($F = 4.279$, $p < 0.05$). LSD post-hoc results showed no difference in overall confirmation rates between video conferencing and co-located scenarios, but MR had a significantly lower rate than co-located ($I-J = -0.006$, $p < 0.05$). Pre-question rates showed no differences, but after-question rates were lower in MR than in co-located ($I-J = -0.004$, $p < 0.05$) and video conferencing ($I-J = -0.003$, $p < 0.05$).

These results suggest that confirmation behaviors are more similar in co-located and video conferencing scenarios, both showing higher counts and rates than MR. Further discussion follows in subsequent sections.

Demonstrative pronouns.

We analyzed the frequency and proportion of demonstrative pronouns "this" and "that" used by local workers and remote helpers across the three scenarios. The proportion was calculated by dividing the frequency of demonstrative pronouns by the total number of words spoken in each scenario.

The Friedman test showed a significant difference in total demonstrative pronouns for local workers across the scenarios ($F = 7.238$, $p < 0.05$). LSD post-hoc tests found no difference between co-located and MR, but significantly higher frequency in video conferencing compared to co-located ($I-J = 9.240$, $p < 0.05$) and MR ($I-J = 9.760$, $p < 0.05$). For remote helpers, no significant difference was found across scenarios ($F = 2.128$, $p > 0.05$).

For the average rate of demonstrative pronouns, local workers showed significant differences ($F = 3.373$, $p < 0.05$), with the rate in co-located significantly higher than MR ($F = 0.047$, $p < 0.05$). Remote helpers also showed a significant difference ($F = 7.196$, $p < 0.05$), with video conferencing rates significantly lower than in both co-located ($F = -0.014$, $p < 0.05$) and MR ($F = -0.016$, $p < 0.05$).

We analyzed the proportions of "this" and "that" across scenarios, suggesting "this" indicates closer proximity and "that" suggests greater distance. For local workers, the average proportions of "that/this" were 0.469, 0.227, and 0.141 (SD = 0.089, 0.429, 0.195) for co-located, MR, and video conferencing, respectively. The Friedman test showed no significant difference across scenarios.

For remote helpers, the average proportions were 0.866, 0.439, and 1.735 (SD = 1.134, 0.332, 1.486), with a significant difference across scenarios ($F = 9.082$, $p < 0.05$). Video conferencing had a significantly higher proportion compared to co-located ($I-J = 0.869$, $p < 0.05$) and MR ($I-J = 1.296$, $p < 0.05$).

In summary, co-located and MR scenarios were more similar in pronoun usage, though local workers used fewer pronouns in MR, indicating greater reliance on these expressions in co-located settings.

Task completion time was the primary performance measure, recorded when participants correctly assembled the blocks. The average times (in seconds) for co-located, MR, and video conferencing were 189.38, 293.44, and 383.33, with standard deviations of 25.296, 40.997, and 60.272, respectively. The Friedman test showed a significant difference in task completion times across scenarios ($F = 4.631$, $p < 0.05$). As shown in Figure 4 VI.2, there was no significant difference between co-located and MR ($I-J = 104.060$, $p > 0.05$), but video conferencing time was significantly higher than co-located ($I-J = 193.949$, $p < 0.05$).

5.2 Qualitative Data Analysis Result

After the experiment, we conducted interviews with remote helpers and local workers to gather their subjective perceptions of presence, actions, speaking, and performance, along with the reasons behind these perceptions. Participants were numbered 1 to 14, with remote helpers labeled as "R" and local workers as "L."

5.2.1 Sense of Co-presence

Perceptions of the Work Environment by Remote Helpers. Twelve of fourteen remote helpers found co-located and MR scenarios more realistic than video conferencing. Six preferred co-

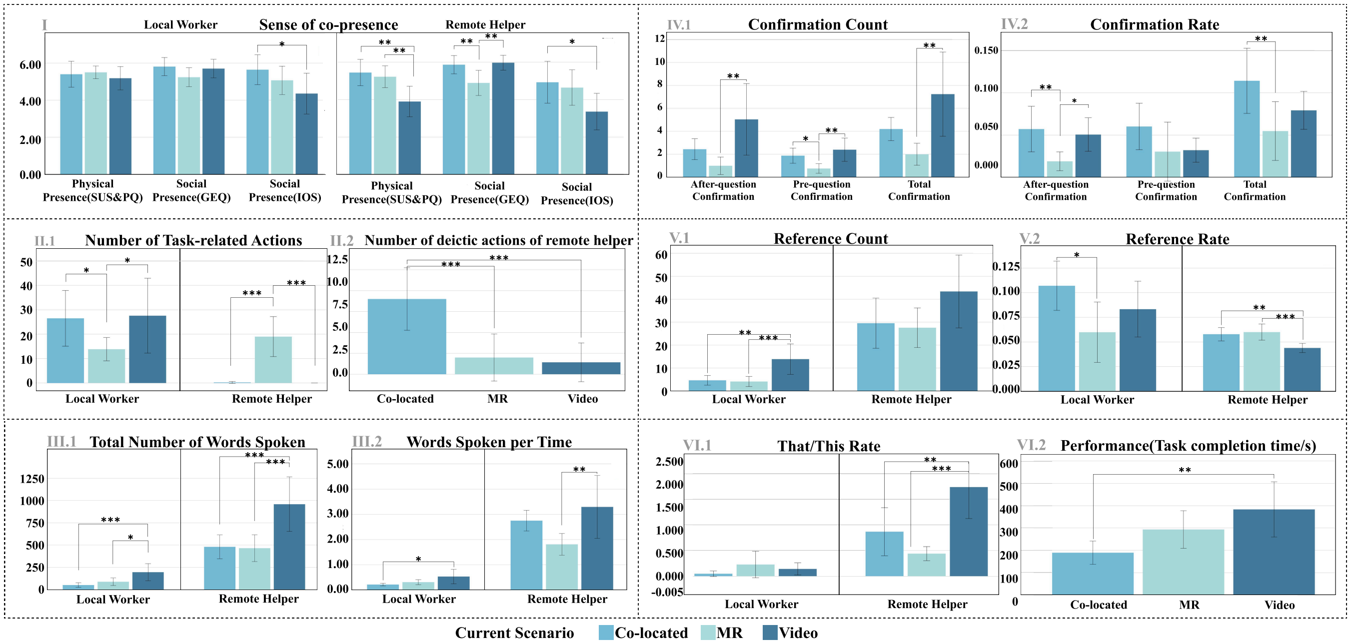

Figure 4: Quantitative Data Analysis: I. - Sense of presence of local worker and remote helper, II.1 - Number of task-related actions, II.2 - Number of deictic actions, III.1 - Total number of words spoken, III.2 - Words spoken per time, IV.1 - Confirmation methods, IV.2 - Confirmation rate, V.1 - Demonstrative pronouns count, V.2 - Demonstrative pronouns rate, VI.1 - "That"/"This" rate, VI.2 - Time taken to complete task (error bars denote standard deviations, * indicates significant differences)

located, three favored MR, and three saw no major difference. Co-located and MR provided a better understanding of the environment, while video conferencing was limited by a restricted field of view (R12: *"My field of view is limited... I can only describe blocks and let them find them."*). MR compensated for the lack of physical presence, but some (R13) felt it emphasized their absence. Five helpers likened MR to VR, noting its sparse point clouds and lack of realism (R13: *"I can walk through the point clouds, which doesn't feel real."*) and missing tactile feedback (R9).

Perceptions of the work environment by local workers. Nine out of fourteen local workers found the co-located and MR scenarios more authentic than video conferencing. Seven preferred co-located, one favored MR, and one saw no significant difference. Four workers perceived no major difference across the three scenarios. Some negative factors in MR included the unrealistic, cartoonish avatars from Meta Horizon Worlds and surreal phenomena like interactions passing through avatars or overlapping blocks, creating a sense of unreality. Workers also noticed transmission delays, with remote helpers' voices arriving faster than images, causing a sense of displacement.

Psychological perception of working together with the other party by remote helpers.

Twelve of fourteen remote helpers felt more collaborative in both co-located and MR scenarios. Six preferred co-located, two favored MR, and four saw no difference. In both scenarios, helpers could see the other party's situation in real time and influence tasks directly through gestures or virtual block demonstrations. In MR, when helpers manipulated virtual blocks while workers assembled physical ones, it felt collaborative: *"I tell him how to assemble it, and he does it with me... I feel like we are working together." (R2)*

Psychological perception of working together with the other party by local workers. Twelve of fourteen local workers felt a stronger sense of collaboration in both co-located and MR scenarios. Four preferred co-located, three favored MR, and five saw no difference. Workers noted that real-time observation of remote helpers provided more intuitive guidance in both scenarios. In MR, they felt more like collaborators rather than being instructed, with the helper's avatar bringing them psychologically closer. Unlike video conferencing, MR allowed helpers to convey information through virtual blocks, enhancing collaboration: *"In MR, I feel more like I am working together because I can see where they place the blocks." (L11)*

Perception of remote helpers and local workers collaborating at the operational level. In the MR scenario, both remote helpers and local workers felt communication relied more on virtual blocks and imagery, with less dependence on each other's actions. *"In MR, I feel like I don't need much cooperation. I just assemble the blocks at my own pace." (R4)* MR conveyed information with a longer lifespan, while co-located and video scenarios required more immediate feedback. *"In MR, he progresses faster, but that's okay. I can quickly know what to do based on his progress, and we complete our parts." (L13)*

5.2.2 Action

Action of remote helpers. Twelve of fourteen remote helpers reported a higher frequency of actions in MR compared to co-located and video conferencing, with ten noting more actions in co-located than in video conferencing. In MR, helpers move and assemble blocks, while in co-located, they guide with gestures. Video conferencing involved fewer actions due to difficulty in conveying information.

R9 mentioned using deictic gestures subconsciously in all scenarios, with the highest frequency in MR and the lowest in video conferencing: *"In MR, I assemble blocks, and she follows. In co-located, I use more gestures, like rotating blocks."* R12 noted that in MR, he demonstrates assembly directly rather than using gestures. Some helpers (R2) preferred gestures in co-located scenarios to simplify verbal descriptions.

Action of Local Workers. Ten of fourteen local workers believed their action frequency was lower in the MR scenario than in co-located and video conferencing. In MR, they mimic the remote helper's block arrangement, reducing trial-and-error attempts, while the other two scenarios required more adjustments. Video conferencing's reliance on verbal descriptions made visualizing outcomes harder, leading to more frequent adjustments. Although co-located actions were slightly higher than in MR, L6 noted smoother actions due to better focus: *"In co-located, I can focus more on completing the task without first observing the remote helper's arrangement." (L6)*

5.2.3 Speaking

Speaking of Remote Helpers. Eleven of fourteen remote helpers felt that language use in the MR scenario was lower than in co-located and video conferencing, with video conferencing requiring the most descriptions. In MR, helpers convey assembly visually by arranging virtual blocks, while video conferencing relies on verbal descriptions: *"It's difficult to describe placement in video conferencing." (R8)* R13 noted that increased language use in video conferencing could enhance precision: *"We established a coordinate system for more precise instructions."* Language use also depends on how each party prefers to describe tasks: *"I express things in 2D and 3D, but he prefers the x, y, z system." (R11)*

Speaking of Local Workers. Eight of fourteen local workers felt that the quantity of language descriptions in the MR scenario was lower than in co-located and video conferencing. In MR, less confirmation is needed as they can see how the remote helper assembles the blocks: *"In MR, I can see how she assembles, so I don't need to confirm as much." (L6)* In contrast, video conferencing requires repeated confirmations. Feedback speed also influences language use; L9 noted that quicker feedback from the remote helper in co-located scenarios reduces the need for questions.

5.2.4 Performance

Performance of Remote Helper. Thirteen of fourteen remote helpers found completing tasks in MR and co-located scenarios easier than in video conferencing. Seven preferred MR, three favored co-located, and three saw both as similar. The key advantage of MR and co-located is the ability to observe the local worker and provide intuitive instructions. However, the collaboration methods differ: *"In MR, I can demonstrate with virtual blocks, while in co-located, the local worker sees my gestures." (R12)*

Performance of Local Worker. Thirteen of fourteen local workers found tasks easier to complete in MR and co-located scenarios than in video conferencing. Seven preferred MR, four found co-located easier, and two saw both as similar. The advantage of co-located is faster feedback: *"In co-located, he can directly point out where I should assemble and quickly correct mistakes." (L3)* MR offers more intuitive information: *"MR is more intuitive, with both verbal descriptions and 3D structures, which I find simpler." (L8)*

5.3 Correlation Analysis between Co-Presence and Behavioral Observation

We conducted a Pearson correlation analysis on the behaviors of remote helpers and local workers and their perceived co-presence, which includes physical presence, psychological social presence, and operational social presence.

5.3.1 Remote helper

The perception of **physical presence** was significantly positively correlated with the demonstrative pronouns rate ($r = 0.456$, $p < 0.05$), suggesting that better physical presence enhances the remote helper's awareness of the environment, increasing pronoun usage.

The perception of **psychological social presence** was negatively correlated with the after-question confirmation rate ($r = -0.406$, $p < 0.05$) and positively correlated with the rate of demonstrative pronouns ($r = 0.383$, $p < 0.05$). This indicates that higher psychological social presence may lead to more pre-question confirmations, while lower presence relies on after-question confirmations.

The perception of **operational social presence** was negatively correlated with task-related actions ($r = -0.462$, $p < 0.05$) and the rate of demonstrative pronouns ($r = -0.410$, $p < 0.05$), but positively correlated with the rate of "that"/"this" ($r = 0.501$, $p < 0.01$). In video conferencing, where operational social presence is highest, remote helpers depend on local workers' feedback, resulting in fewer task-related actions and reduced pronoun usage, with increased use of "that" reflecting greater perceived distance.

5.3.2 Local worker

The perception of **physical presence** was significantly negatively correlated with task-related actions ($r = -0.594$, $p < 0.01$), likely because better awareness of the work environment helps participants select tools and complete tasks more efficiently, reducing unnecessary actions and errors.

The perception of **operational social presence** was significantly negatively correlated with performance ($r = -0.447$, $p < 0.05$), suggesting that higher operational social presence may increase reliance on feedback from the remote helper, thus reducing task efficiency.

6 DISCUSSION

6.1 Relationship Between Remote Helper and Local Worker

In section 5, we observed differences in the psychological and operational aspects of social presence between local workers and remote helpers. Based on the IOS (Inclusion of Other in Self) scale, both groups experienced similar levels of psychological social presence in MR and co-located scenarios, surpassing those in video conferencing. This suggests that MR technology effectively simulates co-presence, aligning with Schroeder's findings on virtual environments emulating face-to-face interactions [58]. Higher psychological social presence in MR may result from real-time visibility of actions and task similarity, as noted by a participant: *"In MR or co-located, I can see her placing the blocks... it feels like we are cooperating in the same room."* (R3). This echoes Lee's research on digital avatars enhancing engagement through visual presence [25].

However, the GEQ (Game Experience Questionnaire) indicated lower operational social presence in MR compared to co-located and video conferencing scenarios. This may be due to the local worker's reduced dependence on the remote helper in MR, where observing block assembly aids understanding. As L13 mentioned, the remote helper's pace can exceed theirs without impacting comprehension.

Participants often perceived MR as more collaborative than guided. The digital MR environment allowed the remote helper to demonstrate tasks, giving local workers a sense of parity in their roles. This supports Whiteside's findings that MR fosters active co-participation in training contexts [77].

6.2 Impact of Environment on Collaborators

Seeing each other's actions is essential for collaboration between remote helpers and local workers, fostering closer connections and enabling rapid feedback. In MR, the three-dimensional, real-time interaction enhances the sense of connection with the workspace and collaborators. L3's observation that seeing each other's images in MR and co-located scenarios promotes closeness supports findings that spatial cues enhance the feeling of shared space [56].

In contrast, video conferencing lacks these spatial cues, complicating communication. R8's experience aligns with Olson's findings that video conferencing limits environmental feedback, leading to communication challenges [49]. Frequent requests to adjust camera angles underscore the importance of perspective, as spatial disconnection can hinder task performance [47].

The use of demonstrative pronouns like "this" and "that" reflects perceived proximity, with a higher frequency of "this" in MR suggesting closeness, supporting studies on how environmental context influences language use in collaborative settings [18].

Lastly, the illusion of presence in MR can lead to errors, such as participants mistaking virtual objects for real ones, echoing Slater's research on how virtual environments can create realistic yet confusing overlays [61]. Similarly, the study by Andrew Irlitti et al. reports that volumetric objects can induce a false sense of control [29].

6.3 Conveying Information through Behavior and Language

In the experiment, we observed two primary modes of information transmission between remote helpers and local workers: behavior and language. Language is convenient for feedback but can lead to misunderstandings. As R6 noted, *"It's hard to describe how the blocks should be transformed... purely verbal communication is very inefficient."* Behavior is more intuitive but depends on copresence and visibility; in co-located scenarios, helpers use gestures more frequently (R2), while video conferencing often relies on language due to limited visibility of actions.

In MR, communication leaned towards non-verbal, action-based interactions like manipulating virtual objects, reducing the need for verbal instructions. This approach proved more efficient for tasks involving physical manipulation, aligning with studies highlighting the importance of non-verbal communication in VR [76]. MR facilitated intuitive task completion by aligning actions with task requirements.

There is a substitutive relationship between actions and language: *"In MR, I talk less and let him follow my actions."* (R6). Referential gestures simplify language complexity: *"In co-located, saying 'this one' or pointing works well, and I can omit a lot of words."* (R3). Conversely, when action-based communication is restricted, verbal descriptions increase.

Referential gestures can simplify language complexity: *"In a co-located scenario, saying 'this one' or 'that one' or just pointing works well, and I can omit a lot of words"* (R3). Deictic terms like "this" and "that" connect actions with verbal descriptions.

6.4 Generalizability Across MR Systems

In this study, we utilized a specific MR system comprising a prescanned environment, real-time point cloud interaction, and the implementation of virtual replicas. Nevertheless, the findings of our experiment demonstrate a degree of generalizability to other MR systems. Notably, virtual replicas constitute a widely adopted feature in remote MR collaboration systems, as evidenced by prior research [9, 27, 51, 71]. These replicas effectively simulate physical environments, thereby enabling remote users to perceive and interact with local workspaces more comprehensively.

Beyond virtual replicas, existing literature [6] indicates that integrating features such as gaze tracking, facial expressions, and virtual avatars [31] into remote collaboration environments can enhance conversational efficiency, modify the use of deictic expressions, and improve co-presence. For example, participants without the pointer used descriptors like color, shape, or size, whereas with the pointer, they simply said "this one" while pointing.

Furthermore, alternative MR modalities, such as 360-degree video and 3D scene reconstruction, exhibit functionalities comparable to those employed in our study's MR system. For instance, these modalities allow users to visualize the local worker's environment and actions, facilitating more efficient communication and fostering a heightened sense of co-presence [70]. These observations underscore the broader applicability of the phenomena identified in this study, suggesting that the insights derived from our MR system can be extended to a range of other MR technologies.

6.5 Limitation and Future Work

MR collaboration platforms provide a more co-located-like experience than video conferencing but face several challenges. First, realism is lacking; participants (R10, R12, R13) reported unclear point clouds and missing objects due to limited room capture via Azure Kinect cameras. Second, basic interactions like grasping lack tactile feedback, affecting accuracy: *"Virtual blocks lack tactility, making control difficult"* (R9). Delays in visual transmission were also noted: *"It feels like slow motion in MR"* (L10). Moreover, the limitations in rendering quality, such as the discrepancy between device resolution (e.g., Oculus Quest Pro) and real-world precision, which diminishes realism, have been consistently highlighted in prior research [30].

Our study focuses on the behavioral differences introduced by MR systems in remote collaboration compared to co-located and video-based interactions, while also examining their relationship with the sense of co-presence. However, our research has certain limitations: the experiment was conducted in a controlled laboratory setting, and the use of a block-building task as a general activity may not fully capture the natural complexity of real-world remote collaboration scenarios. Data analysis in this study primarily relied on manual coding of recorded linguistic and gestural data, which could benefit from more precise measurement methods in future research, potentially incorporating physiological measurement devices. Moreover, this experiment exclusively explored one-onone collaboration processes; future investigations could examine whether similar outcomes emerge in one-to-many or many-to-many collaborative contexts.

7 Conclusion

In conclusion, MR platforms enhance co-presence, offering a more intuitive and effective collaborative environment than traditional video conferencing. By fostering natural interactions with virtual objects and gestures, MR reduces reliance on verbal communication, improving both collaboration efficiency and task performance.

Our findings highlight the advantages of MR in mimicking colocated settings, suggesting its transformative potential for remote collaboration across diverse fields.

8 Acknowledgment

This work is supported by the National Key R&D Program of China under Grant No.2024YFB2808803.

References

[1] D. Anton, G. Kurillo, and R. Bajcsy. User experience and interaction performance in 2d/3d telecollaboration. *Future Generation Computer Systems*, 82:77–88, 2018. 2

[2] A. Aron, E. N. Aron, and D. Smollan. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, 63(4):596, 1992. 4

[3] J. Auda, L. Busse, K. Pfeuffer, U. Gruenefeld, R. Rivu, F. Alt, and S. Schneegass. I'm in control! transferring object ownership between remote users with haptic props in virtual reality. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, pp. 1–10, 2021. 2

[4] J. N. Bailenson, N. Yee, J. Blascovich, A. C. Beall, N. Lundblad, and M. Jin. The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The journal of the learning sciences*, 17(1):102–141, 2008. 2

[5] A. Bayro, Y. Ghasemi, and H. Jeong. Subjective and objective analyses of collaboration and co-presence in a virtual reality remote environment. In *2022 ieee conference on virtual reality and 3d user interfaces abstracts and workshops (vrw)*, pp. 485–487. IEEE, 2022. 4

[6] M. Billinghurst, K. Gupta, M. Katsutoshi, Y. Lee, G. Lee, K. Kunze, and M. Sugimoto. Is it in your eyes? explorations in using gaze cues for remote collaboration. *Collaboration Meets Interactive Spaces*, pp. 177–199, 2016. 3, 9

[7] W. Büschel, P. Reipschläger, R. Langner, and R. Dachselt. Investigating the use of spatial interaction for 3d data visualization on mobile devices. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, pp. 62–71, 2017. 1

[8] J. W. Chastine. *On Inter-Referential Awareness In Collaborative Augmented Reality*. Ph.d. dissertation, Georgia State University, 2007. 1

[9] Y. Cheng, Y. Yan, X. Yi, Y. Shi, and D. Lindlbauer. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 282–297, 2021. 9

[10] H. Clark. Grounding in communication. *Perspectives on socially shared cognition/American Psychological Association*, 1991. 3

[11] H. H. Clark and C. E. Marshall. Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, and I. A. Sag, eds., *Elements of Discourse Understanding*, pp. 10–63. Cambridge University Press, Cambridge, 1981. 3

[12] G. De Leo, L. A. Diggs, E. Radici, and T. W. Mastaglio. Measuring sense of presence and user characteristics to predict effective training in an online simulated virtual environment. *Simulation in Healthcare*, 9(1):1–6, 2014. 2

[13] R. Druta, C. Druta, P. Negirla, and I. Silea. A review on methods and systems for remote collaboration. *Applied Sciences*, 11(21):10035, 2021. 1

[14] T. Duval, T. T. H. Nguyen, C. Fleury, A. Chauffaut, G. Dumont, and V. Gouranton. Improving awareness for 3d virtual collaboration by embedding the features of users' physical environments and by augmenting interaction tools with cognitive feedback cues. *Journal on Multimodal User Interfaces*, 8:187–197, 2014. 2

[15] C. G. Fidalgo, Y. Yan, H. Cho, M. Sousa, D. Lindlbauer, and J. Jorge. A survey on remote assistance and training in mixed reality environments. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2291–2303, 2023. 2

[16] J. Freeman, S. E. Avons, R. Meddis, D. E. Pearson, and W. IJsselsteijn. Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators & Virtual Environments*, 9(2):149–164, 2000. 3

[17] J. Freeman, S. E. Avons, D. E. Pearson, and W. A. IJsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence*, 8(1):1–13, 1999. 3

[18] D. Gergle, R. E. Kraut, and S. R. Fussell. Using visual information for grounding and awareness in collaborative tasks. *Human–Computer Interaction*, 28(1):1–39, 2013. 9

[19] E. Goffman. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. Free Press, 1963. 1

[20] E. Goffman. *Behavior in public places*. Simon and Schuster, 2008. 2

[21] J. E. S. Grønbæk, K. Pfeuffer, E. Velloso, M. Astrup, M. I. S. Pedersen, M. Kjær, G. Leiva, and H. Gellersen. Partially blended realities: Aligning dissimilar spaces for distributed mixed reality meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023. 2

[22] J. E. S. Grønbæk, J. Sánchez Esquivel, G. Leiva, E. Velloso, H. Gellersen, and K. Pfeuffer. Blended whiteboard: Physicality and reconfigurability in remote mixed reality collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2024. 2

[23] K. Gupta, G. A. Lee, and M. Billinghurst. 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):2413–2422, 2016. 2, 4

[24] D. Hein, C. Mai, and H. Hußmann. The usage of presence measurements in research: a review. In *Proceedings of the International Society for Presence Research Annual Conference (Presence)*, pp. 21–22. The International Society for Presence Research Prague, 2018. 2, 3

[25] M. Holzwarth, C. Janiszewski, and M. M. Neumann. The influence of avatars on online consumer shopping behavior. *Journal of marketing*, 70(4):19–36, 2006. 8

[26] X. Huang and R. Xiao. Surfshare: Lightweight spatially consistent physical surface and virtual replica sharing with head-mounted mixed-reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(4):1–24, 2024. 2

[27] X. Huang, M. Yin, Z. Xia, and R. Xiao. Virtualnexus: Enhancing 360-degree video ar/vr collaboration with environment cutouts and virtual replicas. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–12, 2024. 9

[28] W. A. IJsselsteijn, Y. A. de Kort, and K. Poels. The game experience questionnaire. *Technische Universiteit Eindhoven*, 46(1), 2013. 4

[29] A. Irlitti, M. Latifoglu, T. Hoang, B. V. Syiem, and F. Vetere. Volumetric hybrid workspaces: Interactions with objects in remote and co-located telepresence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2024. 9

[30] A. Irlitti, M. Latifoglu, Q. Zhou, M. N. Reinoso, T. Hoang, E. Velloso, and F. Vetere. Volumetric mixed reality telepresence for real-time cross modality collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023. 9

[31] Y. Jiang, Z. Li, M. He, D. Lindlbauer, and Y. Yan. Handavatar: Embodying non-humanoid virtual avatars through hands. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2023. 9

[32] J. G. Johnson, D. Gasques, T. Sharkey, E. Schmitz, and N. Weibel. Do you really need to know where "that" is? enhancing support for referencing in collaborative mixed reality environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021. 3

[33] D. Kahneman and J. Riis. Living, and thinking about it: Two perspectives on life. *The science of well-being*, 1:285–304, 2005. 3

[34] S. Kim, M. Billinghurst, and K. Kim. Multimodal interfaces and communication cues for remote collaboration. *Journal of Multimodal User Interfaces*, 14(4):313–319, 2020. 2

[35] S. Kim, M. Billinghurst, and G. Lee. The effect of collaboration styles and view independence on video-mediated remote collaboration. *Computer Supported Cooperative Work (CSCW)*, 27:569–607, 2018. 3

[36] S. Kim, G. Lee, N. Sakata, and M. Billinghurst. 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)*, pp. 83–92. IEEE, 2014. 3

[37] S. Kimmel, F. Jung, A. Matviienko, W. Heuten, and S. Boll. Let's face it: Influence of facial expressions on social presence in collaborative virtual reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2023. 2

[38] K. Kiyokawa, M. Billinghurst, S. E. Hayes, A. Gupta, Y. Sannohe, and H. Kato. Communication behaviors of co-located users in collaborative ar interfaces. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 139–148. IEEE, 2002. 1

[39] G. Kurillo and R. Bajcsy. 3d teleimmersion for collaboration and interaction of geographically distributed users. *Virtual Reality*, 17:29–43, 2013. 1

[40] 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)*, pp. 9–12. IEEE, 2016. 2, 3

[41] A. F. Monk and C. Gale. A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes*, 33(3):257–278, 2002. 1

[42] J. Müller, R. Rädle, and H. Reiterer. 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*, pp. 6481–6486, 2017. 2, 3

[43] C. Neustaedter, G. Venolia, J. Procyk, and D. Hawkins. To beam or not to beam: A study of remote telepresence attendance at an academic conference. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*, pp. 418–431, 2016.

2
[44] S. Nichols, C. Haldane, and J. R. Wilson. Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies*, 52(3):471–491, 2000. 3
[45] K. Nowak. Defining and differentiating copresence, social presence and presence as transportation. In *presence 2001 conference, Philadelphia, PA*, vol. 2, pp. 686–710. Citeseer, 2001. 2
[46] 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*, pp. 405–415, 2015. 2
[47] T. Ogi and H. Sakon. Distance learning in tele-immersion environment. In *21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07)*, vol. 2, pp. 947–952. IEEE, 2007. 9
[48] M. Ohmi. Sensation of self-motion induced by real-world stimuli. In *Selection and Integration of Visual Information: Proceedings of the International Workshop on Advances in Research on Visual Cognition*, pp. 175–181, 1998. 3
[49] G. M. Olson and J. S. Olson. Distance matters. *Human–computer interaction*, 15(2-3):139–178, 2000. 2, 8
[50] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson. 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*, pp. 1716–1725, 2016. 2, 3
[51] 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, 3, 9
[52] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018. 2
[53] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. 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*, pp. 1–17, 2019. 2
[54] J. D. Prothero, D. E. Parker, T. Furness, and M. Wells. Towards a robust, quantitative measure for presence. In *Proceedings of the conference on Experimental Analysis and Measurement of Situation Awareness*, pp. 359–366, 1995. 2
[55] J. Rasch, V. D. Rusakov, M. Schmitz, and F. Müller. Going, going, gone: Exploring intention communication for multi-user locomotion in virtual reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2023. 3
[56] H. Regenbrecht, T. Lum, P. Kohler, C. Ott, M. Wagner, W. Wilke, and E. Mueller. Using augmented virtuality for remote collaboration. *Presence*, 13(3):338–354, 2004. 8
[57] R. Schroeder. Being there together and the future of connected presence. *Presence*, 15(4):438–454, 2006. 2
[58] R. Schroeder. *Being There Together: Social interaction in shared virtual environments*. Oxford University Press, 2010. 8
[59] N. Schwarz. Self-reports: How the questions shape the answers. *American psychologist*, 54(2):93, 1999. 3
[60] C. P. Scott and J. L. Wildman. Culture, communication, and conflict: A review of the global virtual team literature. *Leading global teams: Translating multidisciplinary science to practice*, pp. 13–32, 2015. 1
[61] M. Slater et al. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: teleoperators and virtual environments*, 8(5):560–565, 1999. 9
[62] M. Slater and A. Steed. A virtual presence counter. *Presence*, 9(5):413–434, 2000. 3
[63] M. Slater, A. Steed, J. McCarthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human factors*, 40(3):469–477, 1998. 4
[64] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018. 3
[65] V. Souza, A. Maciel, L. Nedel, and R. Kopper. Measuring presence in virtual environments: A survey. *ACM Computing Surveys (CSUR)*, 54(8):1–37, 2021. 3
[66] M. Speicher, J. Cao, A. Yu, H. Zhang, and M. Nebeling. 360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, 2(EICS):1–20, 2018. 2
[67] J. Steuer, F. Biocca, M. R. Levy, et al. Defining virtual reality: Dimensions determining telepresence. *Communication in the age of virtual reality*, 33:37–39, 1995. 2
[68] P. M. Strojny, N. Dużmańska-Misiarczyk, N. Lipp, and A. Strojny. Moderators of social facilitation effect in virtual reality: Co-presence and realism of virtual agents. *Frontiers in psychology*, 11:503209, 2020. 1
[69] M. Tait and M. Billinghurst. The effect of view independence in a collaborative ar system. *Computer Supported Cooperative Work (CSCW)*, 24:563–589, 2015. 3
[70] T. Teo, M. Norman, G. A. Lee, M. Billinghurst, and M. Adcock. Exploring interaction techniques for 360 panoramas inside a 3d reconstructed scene for mixed reality remote collaboration. *Journal on Multimodal User Interfaces*, 14:373–385, 2020. 2, 9
[71] 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*, pp. 161–174, 2019. 2, 9
[72] H. Tian, G. A. Lee, H. Bai, and M. Billinghurst. Using virtual replicas to improve mixed reality remote collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2785–2795, 2023. 2
[73] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence*, 9(5):497–503, 2000. 4
[74] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, and H. Ji. Ar/mr remote collaboration on physical tasks: a review. *Robotics and Computer-Integrated Manufacturing*, 72:102071, 2021. 2
[75] X. Wang, P. E. Love, M. J. Kim, and W. Wang. Mutual awareness in collaborative design: An augmented reality integrated telepresence system. *Computers in industry*, 65(2):314–324, 2014. 1
[76] X. Wei, X. Jin, and M. Fan. Communication in immersive social virtual reality: A systematic review of 10 years' studies. In *Proceedings of the Tenth International Symposium of Chinese CHI*, pp. 27–37, 2022. 9
[77] A. L. Whiteside, A. G. Dikkers, and K. Swan. *Social presence in online learning: Multiple perspectives on practice and research*. Taylor & Francis, 2023. 8
[78] J. R. Williamson, J. O'Hagan, J. A. Guerra-Gomez, J. H. Williamson, P. Cesar, and D. A. Shamma. Digital proxemics: Designing social and collaborative interaction in virtual environments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2022. 3
[79] B. G. Witmer, C. J. Jerome, and M. J. Singer. The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 14(3):298–312, 2005. 4
[80] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998. 4
[81] Z. C. Zacharia. Comparing and combining real and virtual experimentation: an effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning*, 23(2):120–132, 2007. 1