



Title	HybridSphere: Enhancing Hybrid Meetings with Avatar-Based VR Environments
Author(s)	Momota, Koji; Shirai, Shizuka; Kobayashi, Masato et al.
Citation	IEEE Transactions on Visualization and Computer Graphics. 2026
Version Type	AM
URL	https://hdl.handle.net/11094/104622
rights	© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

HybridSphere: Enhancing Hybrid Meetings with Avatar-Based VR Environments

Koji Momota¹, Shizuka Shirai^{1†}, Masato Kobayashi¹, Naoya Chiba¹, Photchara Ratsamee², Kiyoshi Kiyokawa³, Yuki Uranishi¹,

¹*The University of Osaka, Japan*

²*Kansai University*

³*Nara Institute of Science and Technology*

[†]shizuka.shirai.cmc@osaka-u.ac.jp

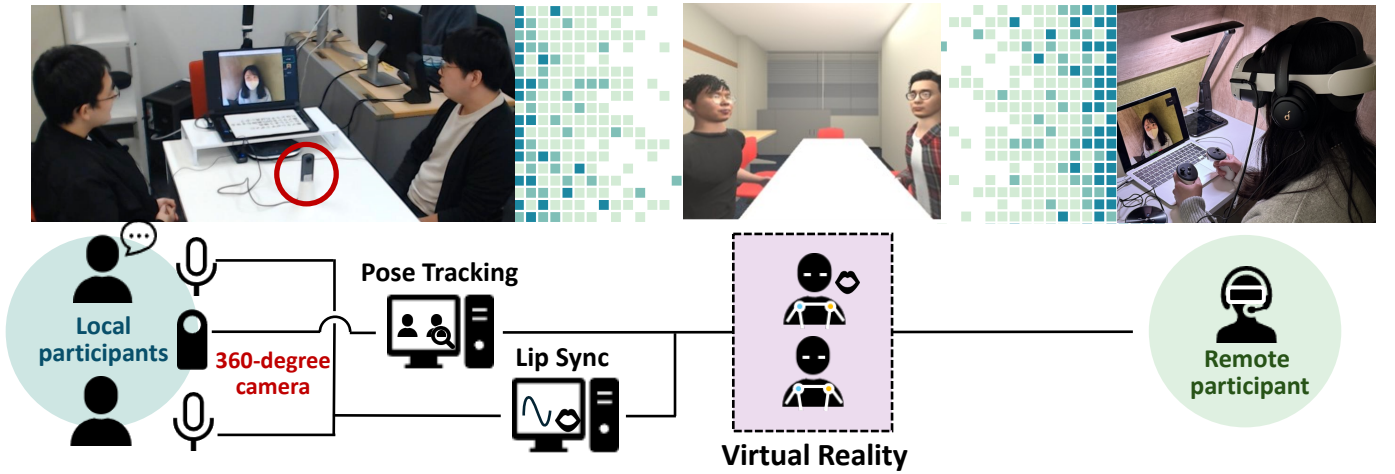


Fig. 1: HybridSphere is a hybrid meeting system that allows remote participants to wear head-mounted displays and experience the meeting environment immersively through avatars representing local participants in a virtual space. Thus, the proposed method addresses the issue that hybrid meetings often result in unequal spatial and nonverbal information distribution.

Abstract—With recent advances in information and communication technologies, Hybrid meetings, where local attendees are physically present and remote participants join virtually, have become increasingly common. However, remote participants often experience reduced contextual awareness and a sense of isolation. To address these issues, we propose HybridSphere, a hybrid meeting system that reconstructs a shared virtual reality environment for remote participants. The system employs a 360-degree camera and pose estimation to generate real-time avatar representations of local attendees to allow remote users with head-mounted displays to experience the meeting as if all participants are in the same virtual space. We conducted a user study comparing HybridSphere to a baseline condition in which remote participants viewed an unmodified 360-degree video. Although the avatars were rated lower in perceived trustworthiness and likability due to limited visual fidelity, participants appreciated the seated-6DoF function. These results suggest that immersive and spatially flexible VR representations can enhance remote engagement in hybrid meetings.

Index Terms—Hybrid Meeting, Virtual Reality, Avatar, 6-DoF viewpoint, 360-degree camera

1. INTRODUCTION

Owing to advancements in digital technologies, remote work and online meetings have become increasingly prevalent. In particular, hybrid meetings, in which local participants convene in person while remote participants join from different locations, have become increasingly common. While such meetings enable flexible participation, they also tend to foster more natural and intimate communication among co-located participants, often leaving remote participants feeling isolated and disengaged [15].

One promising approach to mitigating this disparity is the use of avatars within virtual spaces. With the growing popularity of the metaverse, various efforts have emerged to connect real and virtual environments through avatar-mediated interactions [6], [44]. For instance, Meta’s Horizon Workrooms [2] enables all participants to wear head-mounted displays (HMDs) and convene in fully virtual meeting spaces. However, in hybrid meetings where local participants are already co-located, requiring them to wear HMDs may not be practical and could diminish the convenience of face-to-face interaction. Therefore, there is a growing need for solutions

that allow only remote participants to immerse themselves in a virtual environment without compromising the experience of those attending in person. Among existing approaches for supporting hybrid meetings, Tang et al. [52] proposed a spatial layout that aligns local and remote participants side by side to promote visual equity. However, this method has limitations in conveying body language and spatial information. Also, Kachach et al. [34] proposed The Owl, which presents 360-degree video to remote participants via HMDs, providing a sense of immersion through real-world visuals. Yet, since this method only allows three degrees of freedom (3DoF), users can rotate their view but cannot move forward, backward, or sideways, thus limiting their ability to explore the surrounding space.

To address these challenges of compromising the experience of online participants of hybrid meetings, we proposed “HybridSphere”, a hybrid meeting system that combines both in-person participation physically and remote participation remotely via a VR platform. The proposed system enhances remote participants’ experience by transforming local participants into avatars in a shared virtual space, allowing remote participants wearing HMDs to experience the meeting as if all participants were co-located. Specifically, we capture real-time movements of local participants using 360-degree video combined with pose estimation techniques and reflect these movements onto avatars in a virtual environment. Furthermore, the proposed system provides a seat position changing function that supports limited six-degrees-of-freedom (6DoF) movement, in which users can move their viewpoint among six predefined positions. By doing so, this function allows users to change their viewpoint in a controlled manner and better understand the spatial layout of the meeting environment.

We examine the following three hypotheses regarding the effects of two key factors, the advantages unique to virtual spaces and the fidelity of information reproduction through avatarization, on participants’ experience through the experiment:

- H1: Real-time avatarization enables remote participants to perceive the social presence and characteristics of local participants at a quality level comparable to that achieved using live-action video, thereby facilitating interaction.
- H2: The addition of greater freedom in remote participants’ viewpoints improves their understanding of the spatial relationships within the meeting environment and increases their perceived satisfaction when participating in collaborative group activities, compared to meetings conducted using live-action video.
- H3: The proposed method increases remote participants’ engagement in meetings and leads to a more positive overall meeting experience compared to traditional live-action video-based systems.

In summary, the contributions of this study are as follows:

- Avatar representation was found to be less effective than live-action video in conveying subtle facial expressions and gestures, which may negatively impact remote par-

ticipants’ perceptions of likability and trustworthiness toward local participants. Factors such as the uncanny valley effect and the lack of gaze cues were identified as potential barriers to shared presence and impression formation.

- While changing seat position function with a limited 6DoF feature was positively received in the tasks with activity, a sense of novelty and enjoyment for some users. However, it did not consistently enhance participation or task outcomes. No significant differences were observed in speech time or task accuracy between the two methods, suggesting that the benefits of spatial navigation depend heavily on the nature of the meeting and the user’s familiarity with VR controls.

2. RELATED WORK

In hybrid meetings, spatial context and nonverbal cues tend to be concentrated on the local side because local participants share the same physical space, making it difficult for remote participants to follow the meeting flow and interpersonal interactions. This section summarizes why such asymmetry arises, how prior work has attempted to mitigate it, and how our study is positioned.

2.1. Hybrid meetings and Situational Awareness

Fully remote online meetings can place all participants in similar conditions, but they inherently lack the physical co-presence and informal interactions found in face-to-face settings. Against this backdrop, hybrid meetings, in which co-located local participants and remotely connected participants coexist, have attracted growing research interest [50], [49]. However, hybrid meetings also introduce unique issues. Bos et al. [15] highlight “ingroup cohesion,” where bonding among local participants can inadvertently marginalize remote participants. These issues relate to long-standing theories of awareness and shared spaces in the fields of Computer-Supported Cooperative Work (CSCW) and telepresence.

Remote collaboration depends not only on spoken content but also on situational understanding, such as who is doing what, where, and when. Dourish et al. [22] discuss how actions and their outcomes that naturally appear in a shared workspace can support awareness and reduce the need for explicit reporting. Gutwin et al. [27] systematize workspace awareness and organize the elements required for distributed groupware. Greenberg et al. [26] show that fisheye-view-based presentation can support awareness by providing access to both local detail and global context. Endsley et al. [24] define situational awareness as perception, comprehension, and projection, emphasizing its role in decision-making in dynamic situations. Buxton et al. [18], [17] conceptualize telepresence as the integration of person space and task space, and argue that spatial design can affect the feeling of being together.

In addition, gestures can contribute to establishing reference (e.g., pointing) and coordination, and they can serve as important cues supporting shared situational understanding [25], [36]. Fussell et al. [25] further argue that gestures are not necessarily supported by simply transmitting hand video; instead,

systems should provide visual representations that preserve the functional meaning of gestures. In 3D environments, how much of the body to present is itself a design variable. Benda et al. [12] suggest that full-body representation is not always required and that auxiliary representations (e.g., pointers) can help maintain pointing interpretation when needed. Accordingly, hybrid-meeting support must address not only spatial presentation but also the design of nonverbal cues that enable referencing and joint attention.

These perspectives suggest that asymmetries in access to spatial context and nonverbal cues between local and remote participants are a fundamental challenge in hybrid meetings, particularly for remote participants.

2.2. Recovering Spatial and Non-verbal Cues

Building on these theoretical foundations, numerous studies have explored technical approaches to recovering spatial and nonverbal cues in mediated communication.

For example, Morikawa et al. [42] propose HyperMirror to emphasize sharing a communication space; Nguyen et al. [43] propose MultiView to recover spatial cues such as gaze and gestures in group video conferencing; Pejisa et al. [45] present Room2Room, bringing a remote collaborator into the local physical space; and Tanner et al. [53] explore side-by-side telepresence to support reference and mutual understanding. More recently, Tang et al. [52] propose aligning local and remote video feeds side by side to promote visual parity, but this arrangement does not effectively convey body language or spatial relationships.

With a fixed-view webcam, the overall spatial context is easily cropped out, making remote participants' exploration and confirmation dependent on camera placement. To address this limitation, research on increasing viewpoint freedom has advanced. For example, Tang et al. [51] show that 360° video chat can provide remote participants with view independence, enabling collaborative actions such as exploration, confirmation, and guidance. Young et al. [56] also report that increased viewpoint freedom contributes to spatial presence. Kachach et al. [34] develop The Owl, which delivers 360-degree video to remote participants via an HMD; however, viewpoint control is limited to 3DoF. By contrast, Uson et al. [54] propose Free Viewpoint Video (FVV), reconstructing volumetric representations from multiple depth cameras to enable arbitrary viewpoints, but the required camera count and precise calibration impose substantial deployment burdens.

However, even with free viewing, weak referential cues and joint attention can force orientation alignment and referencing to rely heavily on verbal negotiation [51]. To compensate for this, Jo et al. [33] propose Chili, integrating viewpoint control with stable on-video drawing. Lee et al. [38] propose SharedSphere, overlaying MR-based nonverbal cues onto live 360° video to support remote collaboration. Thus, while a 360° presentation offers the advantage of view independence, additional design is often required to support referencing and joint attention.

Overall, while many hybrid-meeting support methods have been proposed, a persistent trade-off remains between deployment burden and the degree to which spatial and nonverbal

cues can be recovered, particularly in real-world hybrid settings.

2.3. Challenges of VR Meetings

Triggered by the global outbreak of COVID-19, the use of video conferencing has rapidly expanded [16], offering benefits such as reduced travel cost and time [19], [20] and lower environmental impact [5]. However, prior research suggests that face-to-face communication remains the primary mode of human interaction [14], and that the human brain is fundamentally optimized for synchronous interaction in shared physical spaces [37].

As a result, prolonged or intensive use of video conferencing can impose cognitive and psychological burdens, often referred to as "Zoom fatigue," including stress, anxiety, and physical discomfort [7], [48]. Bailenson et al. [11] attribute these effects to characteristics such as sustained close-up gaze, enlarged on-screen faces, and persistent self-view, while limitations in perceiving full-body movements and gestures further hinder emotional understanding [40].

One promising approach to addressing these limitations is the use of VR-based meetings. Communication within virtual environments, often described as the "metaverse," has gained increasing attention, enabling interaction through avatars in shared virtual spaces [6], [44]. Commercial platforms such as VRChat [4], Rec Room [3], and Anyland [1] demonstrate the potential of immersive environments for social interaction. Empirical studies report that VR-based collaboration can improve productivity and creative performance compared to conventional video conferencing [10], [39].

However, fully virtual meeting systems that require all participants to wear HMDs, such as Meta Horizon Workrooms [2], introduce substantial physical and cognitive load, high implementation costs, and the need for user acclimatization [47], [32], [28]. In hybrid meetings in particular, requiring local participants to wear HMDs is often impractical, highlighting the need for hybrid meeting systems that can leverage the strengths of VR for remote participants while minimizing deployment burden and disruption for local participants.

3. HYBRIDSPHERE

As described in the section on related work, Hybrid meetings, in which local participants gather in person while remote participants join via online platforms, often result in an unequal distribution of spatial and nonverbal information. This unequal distribution can make remote participants feel disconnected or excluded from the meeting experience. To address this gap, we propose a new meeting method "HybridSphere."

3.1. Design Principles

The system was designed with the following three design principles to enable high immersion exclusively for remote participants without imposing significant burdens on local participants.

DPI: Local participants should not be required to wear any special equipment.

DP2: A photorealistic avatar design approach is used to represent in-person participants.

DP3: Remote participants are allowed to freely navigate the meeting space with 6DoF.

Regarding the DP1, the proposed system utilizes a relatively lightweight setup, a single 360-degree camera combined with pose estimation, to generate avatars of local participants. In previous research, few studies have focused specifically on enabling high immersion exclusively for remote participants without imposing significant burdens on local participants. Approaches using live-action video are typically limited to 3DoF, while advanced methods such as FVV require extensive camera setups. Moreover, metaverse-based tools often assume that all participants wear HMDs, making them less feasible for hybrid settings. Thus, the proposed system was designed with the principle that local participants should not be required to wear any special equipment, thereby preserving the natural flow of face-to-face interaction.

As for the DP2, the proposed system allows remote participants to feel as if they are joining VR meetings that all participants joined as avatars through real-time avatarization. Also, avatars' visual design and realism play a crucial role in the quality of remote communication. For instance, Han et al. [29] conducted an eight-week study involving 81 participants using a variety of avatars and found that avatars resembling the users' appearance were more effective in fostering synchrony and self-presence compared to standardized avatars. In professional settings, avatars with a more photorealistic appearance are often perceived as more "trustworthy" and "appropriate for business" compared to cartoon-like representations [21], [46]. Furthermore, Phadnis et al. [46] suggest that in contexts where interpersonal relationships have already been established, even non-photorealistic avatars may be more readily accepted. Therefore, our study used an avatar based on facial photographs or appearance data provided by each participant.

Concerning DP3, remote participants are allowed to freely navigate the meeting space with 6DoF, enabling them to better understand spatial relationships and participant behaviors from multiple perspectives. However, if users are unfamiliar with the 6DoF function, there can be a negative effect [9]. Thus, we implement limited 6DoF features, except for head movement. This approach aims to provide a more immersive and shared spatial experience than conventional video conferencing.

3.2. System Architecture

A key feature of the system is that local participants can continue their face-to-face discussions without requiring additional devices, while only remote participants are immersed in the avatar-based virtual space. HybridSphere consists of two primary components: a local-side sensing and capture module and a remote-side display and interaction module. As illustrated in Fig. 3, remote participants enter a virtual meeting space called HybridSphere, where they can engage in discussions with avatarized representations of local participants. A monocular 360-degree camera captures all local participants simultaneously, enabling posture estimation and lip-syncing to represent them as avatars. Regarding the local-side sensing module, a 360-degree camera (RICOH THETA

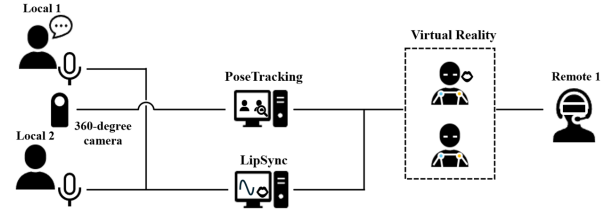


Fig. 2: In this virtual space named “HybridSphere”, remote participants can engage in discussions with the avatarized local participants. A monocular 360-degree camera captures all local participants simultaneously, enabling posture estimation and lip-syncing to accurately represent them as avatars.

V) is placed at the center of the meeting room to capture all local participants simultaneously. The camera provides a live video stream at a resolution of 1920×960 and 29.97 fps. This omnidirectional video feed is ingested into a PC running a Unity-based application (Unity 2022.3.25f1) and processed using a pose estimation library (MediaPipe) to extract joint positions of each participant’s head and upper body. These joint data are then used within Unity to update the avatars in real time.

With regard to the remote-side display and interaction module, remote participants wear an HMD (Meta Quest 3) and enter the Unity-based virtual meeting space. The system supports full 6DoF movement, allowing users to rotate their heads and move freely within the virtual environment. This enables them to virtually change seats or move around the room to improve situational awareness. Voice communication is conducted via a conventional online conferencing platform (Zoom) and synchronized with the avatar system, with speech visually represented through lip-syncing. To minimize the impact of network latency on the visual processing pipeline in our study, both the 360-degree camera and the HMD were directly connected to the same PC used for processing and rendering, while the audio channel was handled separately through Zoom.

3.3. Implementation

To achieve the DP1, a critical challenge in the system implementation was achieving both accurate pose estimation and real-time performance using a single 360-degree camera to capture multiple individuals. Occlusions (e.g., the situation that the lower body is hidden by tables or participants overlapping in the camera view) can reduce estimation accuracy. To address this, we applied coordinate filtering and error correction techniques.

Regarding the DP2, the avatars are pre-generated using an avatar creation tool (Avaturn¹), based on facial photographs or appearance data provided by each participant. Additionally, lip-syncing is applied according to audio input to simulate the local participants’ speech behavior, allowing for a partial but realistic representation of verbal interaction. The choice of avatar style was also carefully considered in light of the

¹<https://avaturn.me>

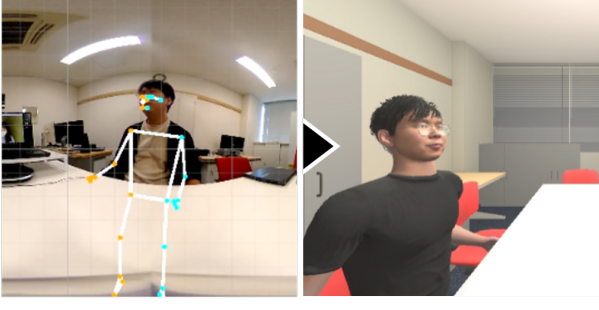


Fig. 3: The left image displays the original spherical image captured by a 360-degree camera, which has been unfolded onto a flat plane. Two-dimensional coordinates, obtained and normalized using MediaPipe, are superimposed on this image. The right image shows the avatar created using the proposed method.



Fig. 4: The remote participant can change their viewpoint using the seat teleportation function. They can move to the desired seat with pointing and selecting the seat via the HMD controller.

uncanny valley phenomenon [41]. While we generated high-fidelity avatars based on participants’ facial photographs, we simplified facial expressions and focused primarily on lip-syncing for verbal cues.

Concerning the DP3, to improve usability for remote participants, we implemented not only free viewpoint movement but also a simplified “seat teleportation” function via HMD controllers, allowing even users unfamiliar with VR to easily navigate the space (Fig. 4). These design considerations aim to minimize cognitive and physical burdens for both local and remote participants while reducing information asymmetry.

4. METHODOLOGY

We conducted a user study to investigate how the presence of remote participants in a virtual space where local participants are represented as avatars affects the overall meeting experience and task performance. This research received ethical approval from the Institutional Review Board (IRB) of the first author’s university (approval code: 2024-13). In this section, we describe the methodology for the user experiment.

4.1. Hypothesis

The user study aims to examine the three main hypotheses described in Section 1.

First, the impact of avatars (H1) focuses on the effects of real-time avatarization using a 360-degree camera on mutual understanding and impression formation between remote and local participants. Specifically, we examine whether the avatars

can reproduce the sense of shared presence with local participants at a level comparable to live-action video and whether the impressions of local participants formed through avatars are maintained to the same degree as those formed through live video.

Second, the impact of 3D spatial navigation (H2) is evaluated to determine whether the function of remote participants to change the seat position with the limited 6DoF features enhances their spatial awareness, their task performance, and their satisfaction. This hypothesis aims to assess not only the effect of avatars but also the extent to which 6DoF viewpoint control contributes to understanding meeting content and facilitating collaboration.

Finally, overall engagement and preference (H3) consider not only task-related outcomes but also broader aspects, such as participants’ willingness to engage in the meeting and the overall usability of the system. This hypothesis includes an investigation of whether the proposed system increases engagement among remote participants and is perceived as a more favorable approach compared to live-action video.

4.2. Participants

To evaluate the remote participant experience in hybrid meetings, we conducted a within-subjects designed experiment involving mock meetings with three-person groups. In each group, two participants attended locally, while one remote participant experienced two meeting formats: the proposed method and a baseline method described in the next subsection. A total of 51 participants were recruited. However, data from one group (three participants) were excluded due to a system error; therefore, data from 48 participants (16 groups of three) were included in the analysis. The participants were undergraduate and graduate students aged between 18 and 26 years (mean age = 21.6, $SD = 2.1$), with an equal gender distribution (24 males and 24 females). All participants were unfamiliar with each other prior to the experiment and were randomly assigned to groups to eliminate the influence of pre-existing relationships on the experimental outcomes.

In each group, one of the three members participated remotely using a head-mounted display (HMD) and experienced both the proposed and baseline methods. The remote participant was randomly selected in front of the other group members to avoid bias and ensure that local participants did not form preconceived impressions about the remote member. Before the experiment, participants were asked to complete a questionnaire regarding their experience with virtual reality (VR). The results showed that 81.3% of participants had “never used VR before” or “a few times a year.” These results indicate that most participants were relatively unfamiliar with VR technologies, suggesting that their experience with the proposed method was largely unconditioned by prior exposure.

4.3. Experimental Setup

This experiment was designed as a within-subjects study in which each remote participant experienced both the proposed method and the baseline method (live-action video with 3DoF). In the baseline method, the remote participant wore

TABLE I: Correspondence between hypotheses and evaluation indicators

Hypotheses	Indicators	Use	
		Task 1	Task 2
H1 Shared Perception via Avatars Impression Formation via Avatars Task Performance	NMM-SPI* ¹ [13]	✓	✓
	Perceived partner characteristics* ¹ [31]	✓	✓
	Speech ratio for Task 1	✓	✓
	Number of ideas for Task 1	✓	
H2 Spatial Awareness Task Performance	MEC-SPQ* ² [55]		✓
	Accuracy rate for Task 2		✓
	Self-evaluation for satisfaction with the outcome of Task 2 [8]		✓
H3 Meeting Engagement Method Preference	Remote participants' speech ratio	✓	✓
	Self-evaluation for satisfaction with group work	✓	✓
	Comparison between methods* ³	✓	✓

*¹The description "other members" in the original text was replaced with "local participants" for remote participants and "remote members" for local participants in the responses.

*²Local participants were asked to answer how well they thought remote participants were able to understand the space.

*³Since local participants were not informed of the methods used by remote participants, they were asked to rank the two meetings held for each task.

an HMD and viewed the unprocessed 360-degree camera footage as a spherical image. This setup allowed for head rotation along 3DoF (pitch, yaw, and roll) but did not permit translational movement such as moving forward, backward, or laterally. Since the local participants were rendered in the live-action video, their facial expressions and body gestures could be faithfully perceived; however, the limited viewpoint mobility made it difficult to explore the space dynamically.

In contrast, the proposed method rendered the local participants as avatars within a virtual environment, enabling the remote participant to navigate the scene with full 6DoF, including both rotational and translational movement. This allowed the user to freely shift viewpoints and observe the meeting space from multiple perspectives. While this approach provided greater flexibility in spatial observation, it also introduced challenges, such as the limited fidelity of facial expression reproduction compared to live-action video. Note that the baseline method resembles The Owl, a system developed by Kachach et al. [34]. However, unlike The Owl, which also includes functionality for visualizing remote participants as avatars on the local side, this feature was intentionally omitted in the present study's baseline method to ensure a clear comparison of the effects of the proposed system.

To reduce bias, local participants were not made aware of which method was being used by the remote participant. This design ensured that remote participants' experiences with navigation and visual representation could be assessed independently while local participants engaged in natural, unmodified face-to-face communication.

4.4. Task

Each group performed two types of tasks twice using different methods. Task 1 was a brainstorming task. Task 2 was a collaborative spatial reasoning quiz, intended to measure spatial understanding and task performance. Each remote participant performed both tasks using both methods, in a counterbalanced order. To prevent learning effects, we prepared two sets of similar but distinct problems for each task. To ensure experimental fairness, we employed a Latin square

design to counterbalance the presentation order of methods and task variants across groups.

Task 1 was designed to examine the effects of avatarization in relation to H1. Accordingly, Task 1 employed a discussion-based activity in which nonverbal cues, such as facial expressions and reactions, play a crucial role. To isolate and evaluate the effects of avatarization, the limited 6DoF seat-changing feature was not utilized in Task 1. In Task 1, participants engaged in a brainstorming session and were asked to reach a conclusion within a time limit. Drawing on the framework proposed by Ide et al. [30], the task topics were designed to be accessible to all participants without requiring prior knowledge. Specifically, the following topics were used: (A) If your toothbrush wears out quickly, how can you make it last 50% longer?, (B) If your room feels cramped, how can you make it feel 50% more spacious? Participants freely proposed ideas and ultimately selected and consolidated a group's consensus.

Task 2 was a collaborative spatial reasoning quiz designed primarily to assess H2. The task required participants to identify a three-dimensional (3D) shape that could be constructed from a given net diagram (Fig. 5). It was presented in a multiple-choice format, and participants worked collaboratively to select the correct answer within a time limit. To evaluate whether providing greater freedom in remote participants' viewpoints improves their understanding of spatial relationships within the meeting environment and their perceived satisfaction, remote participants were allowed to change their seat position within the VR space using the seat-changing function to observe the 3D shape from multiple angles. In contrast, to standardize experimental conditions, local participants were instructed to remain seated. In addition, all participants were instructed not to interact with the 3D objects.

4.5. Measures

Table I shows the correspondence between the hypotheses described in Section 1 and the evaluation metrics used in this experiment. We combined subjective and objective evaluations to verify the proposed method's effects on remote participants'

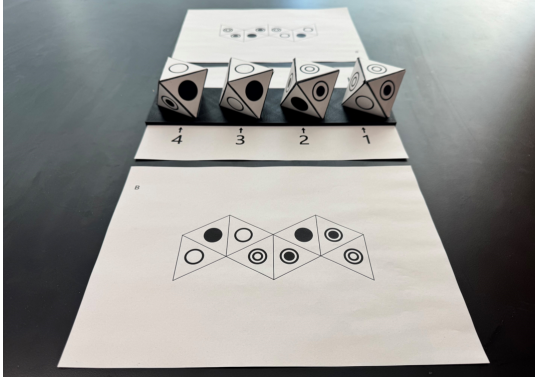


Fig. 5: Task 2 is a quiz question to identify which of the four three-dimensional shapes lined up in the center could be constructed by assembling the net diagram. Participants had to cooperate to share information and discuss the answer.

meeting experience and task performance from multiple perspectives.

Regarding the H1, we used the Networked Minds Social Presence Inventory (NMM-SPI) [13] to measure the spatial and social connections between avatarized local and remote participants. Specifically, 28 items across four categories, Co-presence, Perceived attentional engagement, Perceived emotional contagion, and Perceived comprehension, are presented to remote and local participants, with responses collected on a 7-point Likert scale. We divided the questions into odd numbers (self) and even numbers (others) for each subcategory, calculated the average scores, and used them for analysis. In addition, for questions containing negative expressions, we reversed the values. In addition, we used the questionnaire about perceived partner characteristics [31] to examine impression formation via avatars. This questionnaire consists of nine items across three categories: Likeability, Trustworthiness, and Intelligence. Note that in both questionnaires, for remote participants, “other members” is answered as “local participants,” and for two local participants, “other members” is replaced with “remote members” in their answers. Additionally, to assess task performance, we used the ratio of speech by remote participants and the number of ideas generated by each group in Task 1. For the number of ideas, we counted those expressed during the divergence phase (the first 3 minutes) of Task 1.

Regarding H2, we evaluated spatial awareness and task performance in Task 2. To examine spatial awareness, we used the MEC-Spatial Presence Questionnaire (MEC-SPQ) [55], which employs a 5-point Likert scale. Participants’ understanding of spatial information was assessed using eight items from the Spatial Situation Model category. For local participants, we used the same questionnaire items, but modified the subject from “I” to “the remote participant,” asking local participants to evaluate whether they felt the remote participant was able to perceive the task space.

Additionally, to assess task performance, we evaluated the correctness of the answers that three participants ultimately selected. The accuracy rate was calculated based on whether the answers selected within the time limit of Task 2 were

consistent with the actual correct answers. In addition, Task 2 is a question-and-answer format, so the questionnaire asked about satisfaction with outcomes [8] only when Task 2 was performed. The questionnaire consists of five items about participants’ satisfaction with the answers obtained by the group, their commitment, and their confidence in the validity of the answers, with responses collected on a 7-point scale.

Regarding the H3, to evaluate the engagement in the meeting, we employed two measures: a subjective assessment through a group satisfaction questionnaire and an objective measure based on the speaking ratio of the remote participant. The group satisfaction questionnaire was assessed in terms of enjoyment and engagement in the context of group tasks using a 7-point Likert scale. The questionnaire was administered after the completion of both Task 1 and Task 2. The speaking ratio was calculated by dividing the amount of speech produced by the remote participant by the total amount of speech from all three participants, including the local members. This metric served as an objective indicator of the remote participant’s level of engagement in the meeting.

In addition, for the remote participants, to assess simulator sickness associated with HMD usage, we measured the Simulator Sickness Questionnaire (SSQ) [35], in which participants rate symptoms such as headache, nausea, and visual fatigue on a 4-point scale after each session of each task. Because if avatar-based interaction and 6DoF navigation exacerbate motion sickness, it could present a significant challenge for real-world deployment.

4.6. Analysis

For subjective evaluations excluding method preference, the Wilcoxon signed-rank test, a non-parametric test for paired samples, was conducted. A binomial test was used to examine method preference. For objective evaluations, including speech ratio for Tasks 1 and 2 and the number of ideas for Task 1, the Shapiro–Wilk test confirmed normality; therefore, paired *t*-tests were conducted. For the accuracy rate in Task 2, a McNemar test was conducted. IBM SPSS Statistics (Version 26) was used for all analyses. *P*-values less than 0.05 were considered statistically significant for all tests.

4.7. Procedure

The experimental procedure followed a standardized sequence. Participants first signed consent forms and had photographs taken for avatar creation. They then engaged in face-to-face icebreaker activities to facilitate smooth communication among participants who had not previously met. After the icebreaker, all participants drew lots openly to randomly select three remote participants. This selection procedure was adopted to prevent local participants from forming biased impressions of the remote participants prior to the experimental tasks. Subsequently, a preliminary questionnaire on VR experience was administered only to the remote participants. Task 1 was then explained to all participants. After the explanation, remote participants moved to a separate room and received an explanation of the proposed method or comparison method. After a brief practice session, the remote participants

performed Task 1 according to the experimental conditions. After completing the task, all participants filled out a questionnaire about their meeting experience and the task content. Then, the method was switched, and Task 1 was performed again. After the second Task 1, including a questionnaire, a questionnaire was conducted to evaluate which method the remote participants preferred.

After Task 1, all participants were given an explanation of Task 2 in person, and remote participants again received an explanation of the method and practiced in a separate room. Subsequently, Task 2 was performed using both the proposed method and the comparison method, and participants were asked to fill out a questionnaire after each task. Finally, a method comparison questionnaire was conducted for Task 2, and remote participants were asked to evaluate which method they preferred. The total duration of the experiment, including preparation and completion, was approximately 90 minutes. All sessions were recorded (audio and video) for objective analysis.

5. RESULTS

5.1. Task 1

Table II presents the descriptive statistics for each subjective evaluation in Task 1. Regarding the results of the Wilcoxon signed-rank tests for remote participants, the baseline method yielded significantly higher scores than the proposed method for co-presence (self: $Z = -3.01$, $p = 0.003$, $r = -0.75$), perceived attentional engagement (self: $Z = -2.84$, $p = 0.005$, $r = -0.71$), and perceived emotional contagion on the NMM-SPI (self: $Z = -2.13$, $p = 0.034$, $r = -0.53$; other: $Z = -2.13$, $p = 0.033$, $r = -0.53$). Significant differences were also observed in perceived partner characteristics, including likeability ($Z = 2.96$, $p = 0.003$, $r = 0.74$), trustworthiness ($Z = 2.40$, $p = 0.016$, $r = 0.60$), and intelligence ($Z = 2.25$, $p = 0.024$, $r = 0.56$), as well as in satisfaction with group work in terms of enjoyment ($Z = 2.35$, $p = 0.019$, $r = 0.59$). No significant differences were found for engagement ($Z = 0.80$, $p = 0.426$, $r = 0.20$). In contrast, responses from local participants showed

no significant differences between the two conditions except for the intelligence in perceived partner characteristics ($Z = -2.39$, $p = 0.017$, $r = -0.60$). Median total SSQ scores, administered only to remote participants using HMDs, also showed no significant differences between conditions ($Z = -0.58$, $p = 0.57$, $r = -0.14$).

Regarding objective evaluations, the mean speech ratio for remote participants was 0.43 ($SD = 0.17$) for the baseline method and 0.39 ($SD = 0.15$) for the proposed method. Due to a system failure, speech recordings from one group of remote participants were not available in Task 1; therefore, the speech ratio analysis was conducted using data from 15 groups only. A paired t-test revealed no significant difference between the two methods ($t = 1.94$, $p = 0.07$, $r = 0.46$). Similarly, the mean number of ideas was 9.81 ($SD = 2.96$) for the baseline method and 9.69 ($SD = 3.01$) for the proposed method, with no significant difference observed ($t = 0.13$, $p = 0.896$, $r = 0.03$).

5.2. Task 2

Table III presents the descriptive statistics for each subjective evaluation in Task 2. Regarding the results of the Wilcoxon signed-rank tests for remote participants, the baseline method yielded significantly higher scores than the proposed method for co-presence (self: $Z = -2.35$, $p = 0.019$, $r = -0.59$), perceived attentional engagement (self: $Z = -2.49$, $p = 0.013$, $r = -0.62$), and perceived emotional contagion on the NMM-SPI (self: $Z = -2.20$, $p = 0.028$, $r = -0.55$; other: $Z = -2.45$, $p = 0.014$, $r = -0.61$). Significant differences were also observed in perceived partner characteristics in terms of trustworthiness ($Z = -2.57$, $p = 0.010$, $r = -0.64$). For local participants, the baseline method resulted in significantly higher scores for perceived comprehension (other) than the proposed method ($Z = -2.33$, $p = 0.020$, $r = -0.58$). For the Spatial Situation Model category of the MEC-SPQ, no significant differences were found between the baseline and proposed methods for either remote or local participants. To examine the influence of VR experience, remote participants

TABLE II: Results of Subjective Evaluation of Task 1.

Indicators		Remote participants (N=16)								Local participants (N=32)							
		Baseline			Proposed			Z	r	Baseline			Proposed			Z	r
		Q1	Q2	Q3	Q1	Q2	Q3			Q1	Q2	Q3	Q1	Q2	Q3		
NMM-SPI																	
Co-Presence	self	5.06	5.38	5.94	3.31	4.50	5.94	-3.01**	-0.75	3.81	4.13	5.00	4.00	4.50	5.19	-1.33	-0.33
	other	3.13	4.13	4.75	2.56	3.75	4.44	-1.43	-0.36	3.50	4.00	4.69	3.50	4.00	4.75	-0.63	-0.16
Perceived Attentional Engagement	self	5.33	5.67	5.92	4.33	4.67	5.67	-2.84**	-0.71	4.42	5.50	5.92	4.67	5.33	6.00	-0.01	0.00
	other	4.67	5.17	5.67	4.08	4.67	5.67	-1.27	-0.32	4.75	5.50	5.92	4.67	5.33	6.00	-1.43	-0.36
Perceived Emotional Contagion	self	3.75	4.00	5.00	2.56	3.50	4.63	-2.13*	-0.53	4.00	4.50	4.94	3.75	4.38	5.00	-0.24	-0.06
	other	3.75	4.00	5.00	2.56	3.50	4.63	-2.13*	-0.53	4.00	4.50	5.00	3.75	4.38	4.94	-1.20	-0.30
Perceived Comprehension	self	5.17	6.00	6.00	4.42	6.00	6.00	-1.34	-0.33	5.33	5.83	6.00	5.33	5.67	6.25	-0.66	-0.17
	other	5.33	6.00	6.00	4.75	5.83	6.00	-1.33	-0.33	5.08	6.00	6.00	5.08	6.00	6.00	-0.32	-0.08
Perceived partner characteristics																	
Likeability		5.56	6.00	6.88	3.75	5.13	6.00	-2.96**	-0.74	5.00	6.00	7.00	5.13	6.00	7.00	-0.24	-0.06
		5.00	5.75	6.19	4.38	5.25	6.00	-2.40*	-0.60	5.25	6.00	7.00	5.25	6.00	7.00	-1.06	-0.26
Trustworthiness		5.25	6.50	7.00	4.25	5.50	7.00	-2.25*	-0.56	6.00	6.00	7.00	5.00	6.00	7.00	-2.39*	-0.60
Satisfaction with group work																	
Enjoyment		5.00	6.00	6.00	3.25	5.00	6.00	-2.35*	-0.59	5.25	6.00	6.00	5.00	6.00	6.00	-1.51	-0.38
		5.00	5.50	6.00	4.00	5.50	6.00	-0.80	-0.20	5.00	6.00	6.75	5.00	6.00	6.00	-1.33	-0.33
SSQ total score		59.84	65.45	70.13	59.84	63.58	74.80	-0.58	-0.14	—	—	—	—	—	—	—	—

*, $p < 0.05$, **, $p < 0.01$

TABLE III: Results of Subjective Evaluation of Task 2.

Indicators		Remote participants (N=16)								Local participants (N=32)							
		Baseline			Proposed			Z	r	Baseline			Proposed			Z	r
		Q1	Q2	Q3	Q1	Q2	Q3			Q1	Q2	Q3	Q1	Q2	Q3		
NMM-SPI																	
Co-Presence	self	4.13	4.75	5.50	2.81	3.88	5.25	-2.35*	-0.59	3.00	3.88	4.50	3.50	4.00	4.94	-1.18	-0.29
	other	3.56	4.00	4.88	2.56	3.25	4.44	-1.91	-0.48	3.50	3.75	4.25	3.50	3.75	4.44	-0.41	-0.10
Perceived Attentional Engagement	self	4.42	5.00	5.58	3.42	4.50	5.00	-2.49*	-0.62	3.50	4.67	6.00	3.75	5.00	5.92	-0.21	-0.05
	other	4.33	4.83	5.58	3.75	4.33	5.25	-1.58	-0.39	4.42	4.67	6.00	4.33	5.17	5.92	-0.51	-0.13
Perceived Emotional Contagion	self	4.00	4.13	4.69	3.25	3.75	4.38	-2.20*	-0.55	4.00	4.50	5.00	4.00	4.50	5.00	-0.95	-0.24
	other	4.00	4.00	4.94	3.06	3.88	4.25	-2.45*	-0.61	3.81	4.38	5.00	4.00	4.50	5.00	-0.20	-0.05
Perceived Comprehension	self	3.25	5.00	5.83	2.67	3.17	5.25	-1.89	-0.47	3.67	4.67	5.33	3.00	4.33	5.00	-1.91	-0.48
	other	3.00	4.50	5.58	2.67	3.33	4.92	-1.42	-0.35	3.67	4.83	6.00	3.33	4.17	5.00	-2.33*	-0.58
Perceived partner characteristics																	
Likeability		5.31	6.00	6.44	4.75	5.38	6.00	-1.86	-0.46	5.25	6.13	7.00	5.13	6.13	7.00	-0.08	-0.02
Trustworthiness		6.00	6.00	6.44	4.63	5.75	6.19	-2.57*	-0.64	5.06	6.38	7.00	5.25	6.38	7.00	-0.35	-0.09
Intelligence		6.00	6.50	7.00	5.25	6.00	7.00	-1.73	-0.43	5.00	6.00	7.00	5.00	6.00	7.00	-0.29	-0.07
Satisfaction with group work																	
Enjoyment		3.50	5.00	6.00	3.00	5.00	6.00	-0.58	-0.15	5.00	6.00	6.00	5.00	6.00	6.00	-0.68	-0.17
Engagement		3.00	5.00	6.00	3.00	4.50	5.75	-0.84	-0.21	5.00	6.00	6.75	5.00	6.00	6.00	-0.40	-0.10
SSQ total score		59.84	63.58	70.13	59.84	63.58	70.13	-0.74	-0.19	—	—	—	—	—	—	—	—
MEC-SPQ		2.28	2.56	3.88	2.44	3.25	3.59	-0.88	-0.22	2.25	3.06	3.59	2.66	3.19	3.50	-0.45	-0.11
Satisfaction with outcome		3.80	4.00	5.20	3.25	3.80	4.40	-1.34	-0.33	3.45	4.40	5.50	4.00	4.90	5.55	-0.20	-0.05

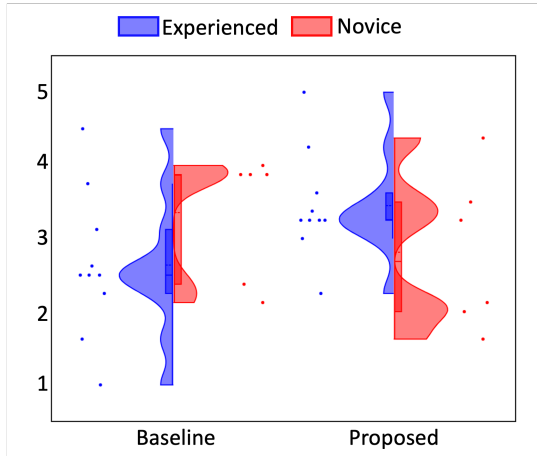
*: $p < 0.05$ 

Fig. 6: The violin plot shows the results of the MEC-Spatial Presence Questionnaire (MEC-SPQ) for each method with a separate VR experience. The scores of VR-experienced users using the proposed method are higher than those of the others.

TABLE IV: Cross-tabulation table for the correct number of quizzes.

		Proposed		Total
		Correct	Incorrect	
Baseline	Correct	4	4	8
	Incorrect	3	5	8
Total		7	9	16

were divided into two groups based on self-reported VR usage experience (Experienced: $N = 9$, Novice: $N = 7$), and the scores were visualized in Fig. 6. Among participants with VR experience, scores for the proposed method tended to be higher. Median total SSQ scores showed no significant differences between conditions ($Z = -0.74$, $p = 0.459$, $r =$

-0.19).

Regarding objective evaluations, the mean speech ratio for remote participants was 0.38 ($SD = 0.14$) and for the baseline method, and 0.36 ($SD = 0.19$) for the proposed method. A paired t-test revealed no significant difference between the two methods ($t = 0.83$, $p = 0.42$, $r = 0.22$). For quiz accuracy, a McNemar test was conducted using the contingency table shown in Table IV, which indicated no significant difference in the distribution of correct and incorrect responses between the two methods.

5.3. Preferences of the method

After each task, remote participants completed a comparative questionnaire asking which method they found more enjoyable, easier to use in future meetings (Loyalty), and more conducive to participation (Efficiency). As shown in Table V, a significantly higher number of participants preferred the baseline method in all three categories in Task 1 (Enjoyable: $p = 0.021$; Loyalty: $p = 0.001$; Efficiency: $p = 0.021$). However, in Task 2, there was a significant increase in preference for the proposed method in terms of enjoyment ($p = 0.021$). Many participants described the ability to freely explore the virtual space as “novel” and “fun.”

Responses from local participants showed no significant differences in either task, indicating that they were largely unaware of which method the remote participant was using.

6. DISCUSSION

Based on the experimental results, this section discusses verifying the three hypotheses formulated in this study.

6.1. H1: Effects of Avatar Representation

To evaluate whether avatar-based representation can provide a shared sense of presence and favorable impressions comparable to live-action video, this study investigated whether avatars

TABLE V: Number of responses on preferences of method (or session) for each task.

	Indicators	Remote participant (N=16)			Local participants (N=32)		
		Baseline	Proposed	<i>p</i>	Baseline	Proposed	<i>p</i>
Task 1	Enjoyable	13	3	0.021*	18	14	0.596
	Loyalty	15	1	0.001**	20	12	0.216
	Efficiency	13	3	0.021*	19	13	0.377
Task 2	Enjoyable	3	13	0.021*	18	14	0.596
	Loyalty	4	12	0.077	16	16	1.000
	Efficiency	10	6	0.454	18	14	0.596

*: $p < 0.05$, **: $p < 0.01$

can facilitate a similar level of shared perception with local participants (shared perception via avatars), and whether the impressions of local participants perceived through avatars are as favorable as those perceived through live video (impression formation via avatars).

Regarding the shared perception via avatars, the results of the NMM-SPI indicated that the baseline method using live-action video outperformed the avatar-based proposed method on several metrics, such as co-presence, emotional contagion, and attentional engagement in both tasks. These results suggest that the avatar-based virtual space may not have conveyed the same strong sense of “being together” as the live video condition. One primary reason for this outcome is the technical limitations of avatar representation. While the system in this study combined 360-degree video with pose estimation to reproduce upper-body movements in real time, it was difficult to capture subtle gestures and rapid facial expressions with high fidelity. In some cases, minor movements were filtered out during processing. Particularly in brainstorming discussions where participants rely heavily on facial expressions and reactions to gauge each other’s intentions, the lack of nonverbal feedback through avatars likely made it harder for remote participants to interpret the social dynamics of the meeting. Another possible factor behind the lower co-presence scores is the uncanny valley phenomenon. When avatars closely resemble humans but fail to match live-action realism in behavior or expression, even minor discrepancies can lead to discomfort. If the appearance of an avatar is nearly, but not perfectly, realistic, users may experience a loss of perceived togetherness rather than a sense of immersion.

With regards to impression formation via avatars, the baseline method yielded higher scores on all categories (likeability, trustworthiness, and intelligence) of partner perception traits for Task 1, and trustworthiness for Task 2. These results indicate that avatarization reduced local participants’ likability, trust, and perceptions of intelligence. In particular, even with photorealistic appearances, the limited range of reproducible movements and expressions may have triggered the uncanny valley [41], potentially increasing discomfort.

Previous studies have shown that in knowledge-based tasks such as meetings, a more realistic appearance is preferred over an abstract one [21], [46]. However, the uncanny valley phenomenon [41], where a realistic appearance is perceived as unsettling, becomes more pronounced, potentially causing discomfort for remote participants. Addressing these challenges will require further advancements in avatar technology. While

the current system implemented a combination of voice-based lip-syncing and pose estimation, future work must explore more naturalistic movement generation and design approaches to mitigate the uncanny valley. Alternatively, one potential direction is the use of intentionally abstract avatars that avoid hyperrealism, which may help reduce user discomfort and improve social acceptance.

6.2. H2: Effects of 3D Spatial Navigation

Next, we examined the extent to which the seat position control using a limited 6DoF feature provided by the proposed method contributes to remote participants’ spatial understanding and task performance. Regarding the spatial awareness, the results of the MEC-Spatial Presence Questionnaire (MEC-SPQ) showed no significant overall differences between the baseline and the proposed method. However, among participants with prior VR experience, scores were slightly higher when using the proposed method. These findings suggest that users who are already comfortable with VR navigation (e.g., familiar with movement controls) are more likely to effectively utilize 6DoF capabilities to enhance spatial understanding. Conversely, these results indicate that the proposed method requires a certain level of familiarity with VR environment.

With respect to task performance, no significant improvements were observed under the proposed method. While some remote participants reported that spatial navigation was helpful for specific quiz problems, local participants often led the discussions and collectively supplemented the required information. As a result, the ability of a single remote participant to change viewpoints did not have a decisive impact on overall group outcomes. These findings further suggest that the effectiveness of viewpoint movement depends strongly on the nature and difficulty of the task. In particular, the spatial quiz used in this study may not have sufficiently emphasized spatial exploration, embodiment, or nonverbal cues, factors that are central to the potential advantages of 6DoF navigation. Future work should therefore consider task designs that more directly assess these aspects in order to more thoroughly evaluate H2.

6.3. H3: Engagement and Method Preference

Finally, we assessed whether the proposed method increases remote participants’ willingness to engage in meetings and whether they perceive it as a more favorable interaction modality.

With regards to the meeting engagement, the results of group satisfaction and the speech ratio, the proposed method

did not significantly enhance engagement. In Task 1, the enjoyment scores for the proposed method were lower than baseline, likely due to limited facial expression reproduction and the uncanny valley effect, making it difficult for remote participants to gauge the atmosphere. In contrast, Task 2 showed no such difference, suggesting that the novelty of 6DoF navigation may have offset these issues to some extent. However, given that speech ratios remained stable across methods, it can be inferred that the proposed method did not substantially reduce engagement. The proposed method may be considered at least equivalent to the baseline in terms of maintaining engagement. Regarding the preference of the method, post-task comparison surveys revealed that, for Task 1 (discussion), most remote participants preferred the baseline method, citing reasons such as ease of understanding facial expressions and gestures. For Task 2, however, more participants favored the avatar-based proposed method, appreciating the ability to freely navigate the space. These results indicate that user preferences are task-dependent.

Accordingly, while live-action video was preferred for tasks emphasizing interpersonal cues, the avatar-based approach was more favorably received in tasks benefiting from spatial exploration. This suggests that future systems should be designed to match interaction modalities with task types. Additionally, features that compensate for avatars' current limitations, such as expression recognition or symbolic emotional feedback, may further enhance their effectiveness.

6.4. Limitations and Future Work

The findings of this study reveal that numerous factors influence user experience in hybrid meetings, including the fidelity of avatar reproduction, the uncanny valley phenomenon, the operational burden of spatial navigation, task characteristics, and users' VR experience. Three issues, in particular, should be addressed in future work: (1) how naturally avatar behavior and facial expressions can be reproduced, (2) which types of tasks benefit most from 6DoF navigation, and (3) how to design onboarding experiences for VR novices. Regarding avatar design, various factors such as application context, ethnicity, and geographic background may influence users' preferences and experiences. Not all application domains require photorealistic avatars. In educational or entertainment contexts, stylized or abstract avatars may enhance creativity and engagement. In contrast, business or formal meetings often demand higher fidelity in facial representation, making the realism of avatars a critical factor. In addition, this study was conducted with participants recruited in the authors' country and within a relatively narrow age range, which may limit ethnic and geographic diversity. These demographic constraints should be considered when interpreting the results. Future work should therefore evaluate the applicability of the findings to participants with different cultural, ethnic, and geographic backgrounds and across different age groups, as well as examine long-term usage in diverse contexts.

Moreover, this study primarily focused on investigating the impact of the proposed method on remote participants, which required placing local participants in standard online

meeting settings. Consequently, the proposed method does not address how local participants perceive the presence of remote participants beyond vocal communication, such as through facial expressions, gestures, or reactions. In addition, where and how a remote participant is placed in the local physical space can directly shape the experience. For instance, Drago et al. [23] reported that seated remote avatars are preferred in social AR/AR teleconferencing, affecting perceived meeting appropriateness, social presence, and etiquette. This aligns with telepresence perspectives emphasizing that experience can be shaped by embedding meaning into spatial structure, distance, and arrangement [18], [17]. Accordingly, examining how awareness of remote participants' presence and nonverbal behavior can be supported for local participants remains an important direction for future work.

While avatar-based virtual spaces still present various challenges, they offer promising opportunities to expand the possibilities of hybrid meetings. When implemented with appropriate technical refinements and task-aligned system design, they have the potential to deliver benefits that conventional video conferencing cannot.

7. CONCLUSION

In this study, we proposed a novel approach to enhancing the experience of remote participants in hybrid meetings by constructing a virtual space in which local participants, captured via a 360-degree camera, are represented as avatars, and remote participants can navigate the environment with the seat position control using a limited 6DoF feature. We conducted the user study to clarify the advantages and limitations of integrating avatars and free-viewpoint interaction into hybrid meeting contexts. Future work should improve avatar expressiveness, optimize task design to enhance the remote participant experience, and develop support mechanisms that enable local participants to better perceive remote participants' presence and nonverbal behavior.

ACKNOWLEDGMENTS

The author acknowledges the use of ChatGPT, developed by OpenAI, to assist with proofreading the manuscript in English. This work was supported by JSPS KAKENHI Grant Number JP23K25716.

REFERENCES

- [1] Anyland. <https://anyland.com/>. (Accessed on 01/22/2025).
- [2] "Horizon Workrooms Virtual Office and Meetings — Meta For Work". <https://forwork.meta.com/jp/horizon-workrooms/>. (Accessed on 01/12/2025).
- [3] Rec Room. <https://recroom.com/>. (Accessed on 01/22/2025).
- [4] VRChat. <https://hello.vrchat.com/>. (Accessed on 01/22/2025).
- [5] A. Aguilera. Business travel and mobile workers. *Transportation Research Part A: Policy and Practice*, 42(8):1109–1116, 2008.
- [6] A. M. Al-Ghaili, H. Kasim, N. M. Al-Hada, Z. B. Hassan, M. Othman, J. H. Tharik, R. M. Kasmani, and I. Shaye. A review of metaverse's definitions, architecture, applications, challenges, issues, solutions, and future trends. *Ieee Access*, 10:125835–125866, 2022.
- [7] K. Anderson and J. C. Looi. Chronic zoom syndrome: emergence of an insidious and debilitating mental health disorder during covid-19. *Australasian Psychiatry*, 28(6):669–669, 2020.

- [8] C. S. Ang, A. Bobrowicz, P. Siriaraaya, J. Trickey, and K. Winspear. Effects of gesture-based avatar-mediated communication on brainstorming and negotiation tasks among younger users. *Computers in Human behavior*, 29(3):1204–1211, 2013.
- [9] P. Atsikpasi and E. Fokides. A scoping review of the educational uses of 6dof hmds. *Virtual reality*, 26(1):205–222, 2022.
- [10] L. Aufegger and N. Elliott-Deflo. Virtual reality and productivity in knowledge workers. *Frontiers in Virtual Reality*, 3:890700, 2022.
- [11] J. N. Bailenson. Nonverbal overload: A theoretical argument for the causes of zoom fatigue. 2021.
- [12] B. Benda and E. D. Ragan. The effects of virtual avatar visibility on pointing interpretation by observers in 3d environments. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 50–59. IEEE, 2021.
- [13] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence*, Philadelphia, PA, pp. 1–9, 2001.
- [14] N. T. Boaz and A. Almquist. *Biological Anthropology*. Harcourt Brace College Publishers, 1996.
- [15] N. Bos, N. S. Shami, J. S. Olson, A. Cheshin, and N. Nan. In-group/out-group effects in distributed teams: an experimental simulation. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 429–436, 2004.
- [16] E. Brynjolfsson. Covid-19 and remote work: An early look at us data, 2020.
- [17] B. Buxton. Mediaspace–meaningspace–meetingspace. In *Media space 20+ years of mediated life*, pp. 217–231. Springer, 2009.
- [18] W. Buxton. Telepresence: Integrating shared task and person spaces. In *Proceedings of graphics interface*, vol. 92, pp. 123–129. Canadian Information Processing Society Toronto, Canada, 1992.
- [19] J. M. Denstadli. Impacts of videoconferencing on business travel: the norwegian experience. *Journal of Air Transport Management*, 10(6):371–376, 2004.
- [20] J. M. Denstadli, T. E. Julsrud, and R. J. Hjorthol. Videoconferencing as a mode of communication: A comparative study of the use of videoconferencing and face-to-face meetings. *Journal of Business and Technical Communication*, 26(1):65–91, 2012.
- [21] G. C. Dobre, M. Wilczkowiak, M. Gillies, X. Pan, and S. Rintel. Nice is different than good: Longitudinal communicative effects of realistic and cartoon avatars in real mixed reality work meetings. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–7, 2022.
- [22] P. Dourish and V. Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pp. 107–114, 1992.
- [23] D. Drago, J. O’Hagan, F. Kawsar, and S. A. Brewster. Exploring avatar seat allocation strategies in social ar. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2025.
- [24] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. In *Situational awareness*, pp. 9–42. Routledge, 2017.
- [25] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19(3):273–309, 2004.
- [26] S. Greenberg, C. Gutwin, and A. Cockburn. Awareness through fisheye views in relaxed-wysiwiw groupware. In *Graphics interface*, vol. 96, pp. 28–38, 1996.
- [27] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3):411–446, 2002.
- [28] K. S. Hale and K. M. Stanney. *Handbook of virtual environments: Design, implementation, and applications*. CRC Press, 2014.
- [29] E. Han, M. R. Miller, C. DeVeaux, H. Jun, K. L. Nowak, J. T. Hancock, N. Ram, and J. N. Bailenson. People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. *Journal of Computer-Mediated Communication*, 28(2):zmac031, 2023.
- [30] M. Ide, S. Oshima, S. Mori, M. Yoshimi, J. Ichino, and S. Tano. Effects of avatar’s symbolic gesture in virtual reality brainstorming. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction*, OzCHI ’20, 8 pages, p. 170–177. Association for Computing Machinery, New York, NY, USA, 2021. <https://doi.org/10.1145/3441000.3441081> doi: 10.1145/3441000.3441081
- [31] C. Jensen, S. D. Farnham, S. M. Drucker, and P. Kollock. The effect of communication modality on cooperation in online environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 470–477, 2000.
- [32] J. Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.
- [33] H. Jo and S. Hwang. Chili: viewpoint control and on-video drawing for mobile video calls. In *CHI’13 Extended Abstracts on Human Factors in Computing Systems*, pp. 1425–1430, 2013.
- [34] R. Kachach, S. Morcuende, D. Gonzalez-Morin, P. Perez-Garcia, E. Gonzalez-Sosa, F. Pereira, and A. Villegas. The owl: Immersive telepresence communication for hybrid conferences. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 451–452. IEEE, 2021.
- [35] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
- [36] D. Kirk, A. Crabtree, and T. Rodden. Ways of the hands. In *ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work, 18–22 September 2005, Paris, France*, pp. 1–21. Springer, 2005.
- [37] N. Kock. Media richness or media naturalness? the evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. *IEEE transactions on professional communication*, 48(2):117–130, 2005.
- [38] G. A. Lee, T. Teo, S. Kim, and M. Billinghamurst. A user study on mr remote collaboration using live 360 video. In *2018 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 153–164. IEEE, 2018.
- [39] J. Li, Y. Kong, T. Rögglä, F. De Simone, S. Ananthanarayan, H. De Ridder, A. El Ali, and P. Cesar. Measuring and understanding photo sharing experiences in social virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2019.
- [40] H. K. Meeren, C. C. Van Heijnsbergen, and B. De Gelder. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, 102(45):16518–16523, 2005.
- [41] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2):98–100, 2012.
- [42] O. Morikawa and T. Maesako. Hypermirror: toward pleasant-to-use video mediated communication system. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pp. 149–158, 1998.
- [43] D. Nguyen and J. Canny. Multiview: spatially faithful group video conferencing. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 799–808, 2005.
- [44] K. L. Nowak and J. Fox. Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, 6:30–53, 2018.
- [45] T. Pejisa, 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.
- [46] V. Phadnis, K. Moore, and M. Gonzalez-Franco. The work avatar face-off: Knowledge worker preferences for realism in meetings. In *2023 IEEE international symposium on mixed and augmented reality (ISMAR)*, pp. 960–969. IEEE, 2023.
- [47] 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.
- [48] R. Riedl. On the stress potential of videoconferencing: definition and root causes of zoom fatigue. *Electronic Markets*, 32(1):153–177, 2022.
- [49] B. Saatçi, K. Akyüz, S. Rintel, and C. N. Klokmoose. (re) configuring hybrid meetings: Moving from user-centered design to meeting-centered design. *Computer Supported Cooperative Work (CSCW)*, 29(6):769–794, 2020.
- [50] B. Saatçi, R. Rädle, S. Rintel, K. O’Hara, and C. Nylandsted Klokmoose. Hybrid meetings in the modern workplace: stories of success and failure. In *Collaboration Technologies and Social Computing: 25th International Conference, CRIWG+ CollabTech 2019, Kyoto, Japan, September 4–6, 2019, Proceedings 25*, pp. 45–61. Springer, 2019.
- [51] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration in 360 videochat: Challenges and opportunities. 2017.

- [52] J. C. Tang, K. Inkpen, S. Junuzovic, K. Mallari, A. D. Wilson, S. Rintel, S. Cupala, T. Carbary, A. Sellen, and W. A. Buxton. Perspectives: creating inclusive and equitable hybrid meeting experiences. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2):1–25, 2023.
- [53] P. Tanner and V. Shah. Improving remote collaboration through side-by-side telepresence. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pp. 3493–3498. 2010.
- [54] J. Usón, C. Cortés, V. Muñoz, T. Hernando, D. Berjón, F. Morán, J. Cabrera, and N. García. Untethered real-time immersive free viewpoint video. In *Proceedings of the 16th International Workshop on Immersive Mixed and Virtual Environment Systems*, pp. 45–49, 2024.
- [55] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, L. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, et al. Mec spatial presence questionnaire. *Retrieved Sept*, 18(2004):2015, 2004.
- [56] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE transactions on visualization and computer graphics*, 25(5):1908–1918, 2019.