Hybrid Robot Integrating Physical and Virtual Body Parts: Effects of Head and Arm Modalities on Social Presence

Ikkaku Kawaguchi
University of Tsukuba
Tsukuba, Ibaraki, Japan
kawaguchi@cs.tsukuba.ac.jp

Keiichi Ihara
University of Tsukuba
Tsukuba, Ibaraki, Japan
University of Colorado Boulder
Boulder, Colorado, USA
kihara@iplab.cs.tsukuba.ac.jp

Ayumi Ichikawa
University of Tsukuba
Tsukuba, Ibaraki, Japan
aichikawa@iplab.cs.tsukuba.ac.jp

Aoi Sakata
University of Tsukuba
Tsukuba, Ibaraki, Japan
asakata@iplab.cs.tsukuba.ac.jp

Yusuke Ashizawa
University of Tsukuba
Tsukuba, Ibaraki, Japan
ashizawa@iplab.cs.tsukuba.ac.jp

Shintaro Mori
University of Tsukuba
Tsukuba, Ibaraki, Japan
smori@iplab.cs.tsukuba.ac.jp

Miki Hasegawa
University of Tsukuba
Tsukuba, Ibaraki, Japan
hasegawa@iplab.cs.tsukuba.ac.jp

Kosuke Fujikawa
University of Tsukuba
Tsukuba, Ibaraki, Japan
fujikawa@iplab.cs.tsukuba.ac.jp

Abstract

This paper investigates a hybrid robot that integrates physical and AR body parts, focusing on how the presentation modalities of the head and arms impact human perception of social presence. We implemented a hybrid robot system that allows switching the presentation modalities of the head and arms between physical and AR, and compared four conditions: both physical, physical head and AR arms, AR head and physical arms, and both AR. Two experiments were conducted with different tasks: material explanation and multi-party discussion. These experiments demonstrate that the hybrid robot with a physical head and AR arms effectively overcomes AR device limitations, achieving a social presence comparable to or greater than that of a fully physical robot. The findings also show that the specific AR-presented body parts significantly influence evaluations, underscoring the importance of careful design in hybrid robots.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**.

Keywords

Augmented Reality, Embodiment, Social Presence

ACM Reference Format:
Ikkaku Kawaguchi, Keiichi Ihara, Ayumi Ichikawa, Aoi Sakata, Yusuke Ashizawa, Shintaro Mori, Miki Hasegawa, and Kosuke Fujikawa. 2025. Hybrid Robot Integrating Physical and Virtual Body Parts: Effects of Head and Arm Modalities on Social Presence. In *13th International Conference on Human-Agent Interaction (HAI '25), November 10–13, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3765766.3765783

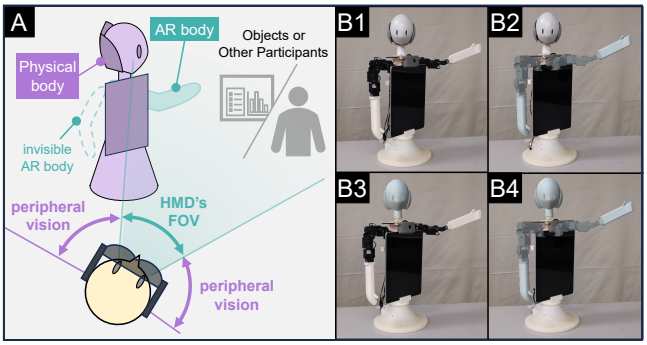

Figure 1: (A)Concept of hybrid robots. (B)Four conditions we compared: (B1) Both physical, (B2) physical Head and AR Arms, (B3) AR Head and Physical Arms, and (B4) Both AR.

1 Introduction

In interactions with others in the physical world, conveying social presence [8, 18] is essential for achieving smooth and meaningful communication. Social presence refers to the sense of "being together with another" [8], and a key foundation of it is the use of nonverbal cues such as gaze, posture, and gestures. These bodily expressions help social interaction—for instance, regulating conversational flow, indicating attention, and supporting shared understanding [2, 15, 23, 28]. The ability to engage in such bodily-based communication is referred to as embodiment [11].

To reproduce this social capability, the concept of robotic embodiment has been explored [10]. For example, telepresence robots like kubi [41] or Double [39] convey remote person's gaze and proxemics through display rotation and its mobility [7, 9, 14, 31, 34, 36–38, 45]. Furthermore, robots with more human-like physicality, such as head or arms, can convey detailed bodily states through

HAI '25, Yokohama, Japan

ACM ISBN 979-8-4007-2178-6/25/11
https://doi.org/10.1145/3765766.3765783

the robot's physical body parts [1, 3, 6, 13, 24, 25, 29, 35]. However, using physical robots involves significant costs in system design, implementation, and maintenance.

An alternative approach is virtual embodiment, using CG avatars presented through AR devices [19–21, 27, 30, 32, 33]. CG avatars are not constrained by physical limitations and offer advanced embodiment at a relatively lower cost. However, the currently available AR headsets have a limited field of view(FOV) [26], which affects peripheral perception and causes a decrease in social presence [21].

We focus on a hybrid embodiment that integrates an optical see-through AR device with a physical robot to address the trade-offs between robotic embodiment and virtual embodiment. In this paper, we define a hybrid embodiment as one that integrates different modalities for each body part, such as a physical head and AR arms, and call the robot with hybrid embodiment as hybrid robot. By using an optical see-through AR device that does not block peripheral vision, physical body parts can be continuously visible even outside of the AR device's FOV, while AR offers advanced embodiment within the FOV(Fig.1(A)). This configuration is expected to compensate for the limitations of the AR's FOV and improve social presence compared to using AR alone. Additionally, replacing some physical components with AR can reduce the costs associated with the design, implementation, and maintenance of physical robots. It should be noted that there are already studies evaluating robots that match our definition of hybrid robots [16, 17]. However, previous studies only evaluated the hybrid robot positioned directly in front of participants and entirely within the FOV of the AR device. Thus, the effects of hybrid embodiment have not been investigated in situations where the targets of interactions (other participants or objects) are located away from the robot, causing the robot to fall outside the AR device's FOV. In addition, the body parts adopted for AR presentation were limited to the arms. In a hybrid configuration, other body parts, such as the head, could also be adapted for AR presentation, and effectiveness may vary depending on the chosen body part. Based on these limitations, this study set the following research questions(RQs).

RQ1 How does the hybrid configuration affect social presence when AR parts occasionally fall outside the FOV?

RQ2 Do evaluations of the hybrid robot differ based on which body part is adopted for AR presentation?

To address the RQ1 and RQ2, we adopt the following approaches: First, we evaluate the hybrid robots in two experimental settings where participants need to interact with elements located away from the robot(e.g. material explanation task, multi-party discussion task). Second, we focus on the arms and the head as the target of AR presentation in hybrid robots, and compare four conditions combining physical and AR presentation for each body part (Fig.1(B1-B4)).

In this paper, we describe the hybrid robot system we implemented for our investigation, which is capable of switching between physical and AR modalities for the head and arms. We then describe the two experiments and their results, and report the findings of this research.

2 Related work

2.1 Robotic Embodiment

A method using robots has been proposed to convey body status and enhance social presence. Telepresence robots, such as kubi [41]or Double [39], allow remote participants to look around the local site by rotating a display with a camera and conveying their references to the local participants or objects. The rotation of the display can be combined with video images and has been adopted by many systems [7, 9, 14, 31, 34, 36–38, 45]. While display rotation conveys the remote participant's spatial reference and improves conversational flow [7, 38], rotating a display showing a face image leads to gaze misperception [4, 22]. As an alternative approach, systems using more human-like physical bodies, such as heads or arms, have proposed [1, 3, 6, 24, 25, 29, 35]. However, using physical robots involves significant costs in system design, implementation, and maintenance. In this study, we explore the potential of hybrid embodiment, which may help reduce such costs.

2.2 Virtual Embodiment

There are researches that present a virtual body through AR in remote communication. For example, a method using simple body parts (head, hand) [5], whole body CG avatar [21, 27], and point cloud of the actual user [19, 20, 30] has been proposed. The current head-mounted displays (HMDs) have a limited FOV [26], therefore proposals to address the FOV issue, such as a method to make the avatars smaller to fit them within the FOV [32, 33]. In addition, compared to objects in real space, luminance and resolution are also constrained in optical see-through AR devices. This study mitigates the effects of these AR limitations by using a robot with a physical body combined with AR. In addition, some systems combine physical robots with AR [19, 21, 27]. However, in these systems, the embodiment is entirely presented through AR, so we distinguish them from hybrid embodiment, where both a physical body and AR are used together for embodiment.

2.3 Hybrid Embodiment

As an example of hybrid embodiment, a method to improve expressiveness by giving arms to a social robot without arms has been proposed [16]. However, this study compares the robot with and without AR arms, but does not assess the differences between physical and AR modalities for the same body part. In contrast, Han et al. conducted a study comparing the effects of deictic gestures performed with physical arms versus AR arms [17]. However, the target of the AR presentation was limited to the arms. Furthermore, these studies only evaluated the hybrid robot positioned directly in front of participants and entirely within the FOV of the AR device.

Based on these limitations, this study aims to further evaluate the effectiveness of hybrid robots. First, we assess the effectiveness of the hybrid robot in situations where participants need to interact with elements other than the robot. Such situations are precisely where robotic embodiment is needed, but the limited FOV becomes a critical issue, so evaluating the hybrid robot in these situations is valuable for understanding its effectiveness. In this study, we select two scenarios: one in which a robot explains materials in the real world, and another in which a robot participates in a multi-party discussion. We conduct experiments in each of these situations(Section

4, 5). Second, we investigate whether the hybrid robot's evaluations vary depending on the body part presented in AR. We adopt not only arms but also the head for AR presentation, as the head plays a crucial role in social interaction by conveying gaze. We compare four conditions by altering the presentation modality of the head and arms.

3 System

We implemented a hybrid robot system capable of switching between physical and AR modalities for the head and arms. The system includes the physical robot, Microsoft HoloLens 2, the AR control PC, and the remote control PC(Fig. 2 (A)).

3.1 Physical Robot Implementation

The physical robot used in this study is a telepresence robot equipped with a humanoid head, developed in our previous research. The robot's appearance is shown in Fig. 2 (B). The physical robot has a head with 2 degrees of freedom (DoF) and arms with 4DoF(3 DoF in the shoulder and 1 DoF in the elbow). Both the head and arms are independently detachable. Joint movements are operated by servo motors, each controlled by serial communication from the Surface Go 3embedded in the robot's torso. The angles of each joint are controlled by data received via UDP communication from the remote control PC. A Python program running on the Surface Go 3 manages communication with the remote control PC and the servo motors' serial control. For remote communication, a video call is connected between the Surface Go 3 and the remote control PC. A 120-degree wide-angle web camerawas used to capture the local environment. In this study, we focused only on the robot's body movements; hence, the display was deactivated, and the participant's face was not shown.

3.2 AR Implementation

HoloLens 2 is utilized to overlay AR-rendered head and arms onto the physical robot. HoloLens 2's resolution is 1440 x 936 per eye (47 px/deg), and the field of view is approximately 52 degrees diagonally (28.5 degrees vertically, 43 degrees horizontally). The frame rate was about 60 fps during the system operation. The AR presentation program is implemented by Unity and played on the AR control PC, and shows AR body parts on HoloLens 2 through Microsoft's

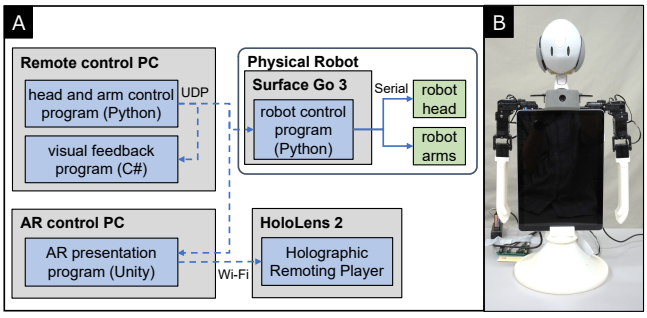

Figure 2: System configuration, (A)block diagram showing system components, (B)appearance of the physical robot.

Holographic Remoting Player [1]. The 3D data for the AR head and arms is exported from the 3D CAD data used for the robot's physical implementation. The angles of each joint are controlled by data received via UDP communication from the remote control PC. In addition, a function to record the user's gaze during the task in the experiment was implemented using Microsoft's Mixed Reality Toolkit (MRTK) [2]. On the AR control PC,the AR presentation program is executed that receives data from the remote control PC via UDP and controls the movements of the body parts presented in AR. In Experiment 1, instead of using the data from the remote control PC, we controlled the robot using the Python program that plays pre-recorded explanations.

3.3 Remote Control

On the remote control PC, the head and arms control program is executed. The direction of the robot's head and arms are controlled by the remote operator's body motion. The motions of head and arms of the remote operator are estimated from the video obtained from the web camera, using the OpenCV [3] and dlib [4] for head direction, and the MediaPipe [5] for arm position. The information of the head and arms is sent to the AR control PC and Surface Go 3 on the robot via UDP communication. The same information is also sent to the visual feedback program which is also run in the remote control PC. Visual feedback is usedto check the current direction of the robot's head and the arms, and is displayed on the video of the local environment. The feedback of the head direction is expressed by a red square frame that moves corresponding to the head of the physical robot on the local site, and the arm direction is expressed by changing the background color of the square on the screen placed at the position corresponding to the current arm direction.

4 Experiment 1: Material Explanation Task

In this study, we evaluate the proposed system in two different experimental settings (e.g., a material explanation task and a multi-party discussion task). This paper first describes the content and results of each experiment, and then discusses the characteristics of the hybrid robot based on the results of both experiments.

In this section, we describe the experiment to evaluate the effects of the presentation modalities of head and arms in material explanation tasks. The experiment focused on explaining two materials placed in physical space. This experiment was conducted with the approval of the ethics review committee of the authors' affiliated organization.

4.1 Experiment Design

We set a task in which a remote operator(experimenter) explained two materials placed in the physical space to one participant. These materials were photographs of two local souvenir items from a specific prefecture in Japan. The explanations covered the names, features, and prices of each item. In the task, to control speech and motion in each trial, prerecorded voice and predefined motion

[1] https://apps.microsoft.com/detail/9nblggh4sv40
[2] https://github.com/microsoft/MixedRealityToolkit-Unity
[3] https://github.com/opencv/opencv
[4] https://github.com/davisking/dlib
[5] https://github.com/google/mediapipe

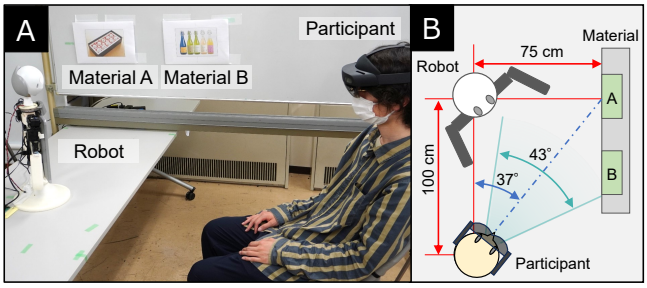

Figure 3: Experimental setup. (A)actual environment, (B)overview of the layout

were used to explain the two materials. Motion during the task included pointing with the arm for materials and directing the head towards the materials and the participant. Since this experiment was designed as a within-participant design with four conditions, we prepared scripts for the four prefectures. After the explanation, the remote operator asked the participant which souvenirs they preferred, and the participant orally provided their choice and the reasons. The average time required for a task(explanations and responses) was 80.61 seconds (SD = 8.72), with variations in time due to differences in the participants' responses. The experimental setup is shown in Fig.3. The angle between the material and the robot was approximately 37 degrees at minimum. The FOV of HoloLens 2 is 43 degrees horizontally(21.5 degrees for each side), which causes the robot to be outside the AR presentation range when participants are directly facing the materials. To mitigate the influence of motor noise, the explanation was presented through a noise-canceling headset.We recruited 16 voluntary participants(1 female, 15 males; mean age = 22.4, SD = 1.26) from undergraduate and graduate students at our local university. All participants were affiliated with an undergraduate or graduate department in computer science. Regarding familiarity with AR/VR, participants responded using a 7-point Likert scale (1: beginner, 7: expert), with an average score of 3.25 (SD = 1.69). The experiment consisted of pre-experiment instruction, a practice task, a main task and questionnaires in each condition, and an interview after all conditions were completed. The overall duration of the experiment was approximately 1 hour, and participants received compensation following the guidelines of the authors' affiliated organization.

4.2 Conditions

In this study, we established four conditions that combine physical and AR presentations for both the head and arms: **C1: Both Physical, C2: Physical Head and AR Arms, C3: AR Head and Physical Arms, C4: Both AR** (Fig.1(B1)-(B4)). Although the AR body is not presented in C1, participants wore HoloLens 2 to standardize the effect of wearing HoloLens 2 (visibility in the experimental environment, physical burden, etc.). We adopted a within-participant design. Considering the order effects of each condition and the scripts, we counterbalanced the order for the conditions and the scripts used in each trial by Latin squares.

4.3 Measures

We set the following measures to evaluate the effects of each body part's presentation modality on social presence and interaction.

Social Presence. We employed the Networked Mind Social Presence questionnaire [18] to assess social presence in our study. From the original questionnaire, We selected three subscales most relevant to our task: Co-presence (CoP), Attentional Allocation (AA), and Perceived Message Understanding (PMU). Each subscale consisted of 6 items, and their average values were used as scores for the respective categories.

Clarity of Body Status. To investigate how well users could recognize the state of each body part, we set a questionnaire to evaluate the clarity of body status. The questionnaire includes questions about awareness and direction clarity of gaze and pointing (Q1-Q5), as well as questions about the sense of eye contact (Q7) and looking at the same object (Q7). Participants answered each question on a 7-point Likert scale.

Gaze Direction. To evaluate the references to the physical space from the system, we analyzed the participants' gaze direction during the experiment. We utilized HoloLens 2's gaze tracking feature to measure the time participants spent looking at the robot, Material A, Material B, and other areas during the task. To determine whether participants looked at each object, we defined detection regions for the robot, Material A, and Material B. We considered the user's gaze ray (invisible to the user) from HoloLens 2's gaze recognition feature hitting the detection region as an indication of looking at the respective object. The temporal resolution for gaze direction acquisition was approximately 30 Hz.

Preference. After completing all conditions, a semi-structured interview was conducted. In the interview, participants ranked the four conditions in terms of preference based on the viewpoint of receiving explanations about materials placed in the physical space from a remote participant. Participants were also asked to explain the reasons for their ranking.

Quality of AR Presentation. Additionally, as a post-survey, we conducted a questionnaire on the quality of AR presentations. Participants rated resolution, brightness, and smoothness of motion on a 7-point Likert scale (1: AR presentation felt lower quality than physical body, 4: No difference, 7: AR presentation felt higher quality than physical body).

4.4 Results

For the analysis, a two-way repeated-measures ANOVA with Aligned Rank Transform (ART) [44] was conducted, considering the factors of head and arm presentation (physical vs. AR). We set the significance level at 0.05 for all analyses.

Social Presence. The results of each subscale are shown in Fig. 4(A). Statistical analysis revealed significant main effects of head presentation on Co-presence (CoP) and Perceived Message Understanding (PMU) ($F(1, 15) = 14.0$, $p = 0.002$, $\eta^2_G = 0.16$; $F(1, 15) = 10.8$, $p = 0.005$, $\eta^2_G = 0.077$). A significant main effect of arm presentation on Attentional Allocation (AA) was also found($F(1, 15) = 5.71$, $p = 0.031$, $\eta^2_G = 0.017$).

Clarity of Body Status. The results of each question are shown in Fig.4(B). Statistical analysis revealed significant main effects of head presentation on Q3, Q5, Q6, and Q7 ($F(1, 15) = 6.51$, $p = 0.022$,

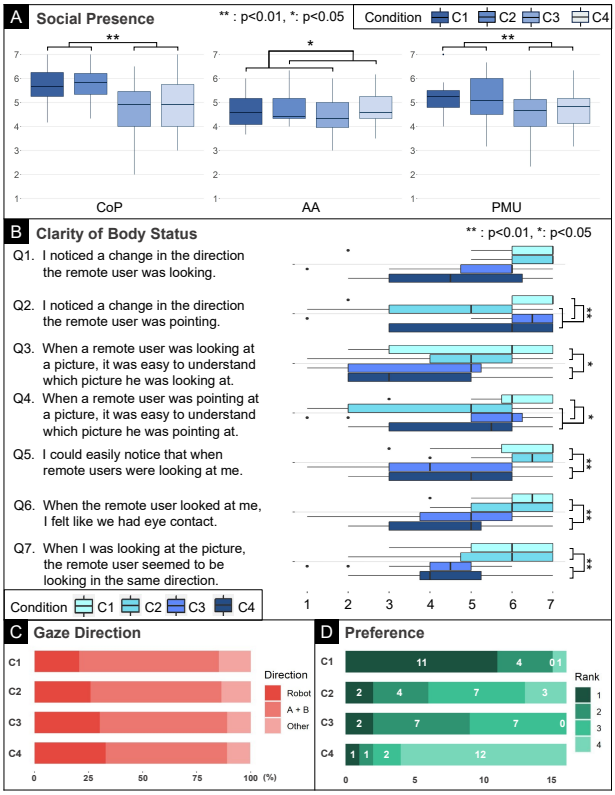

Figure 4: Result of experiment 1. (A)social presence, (B)clarity of body status, (C)gaze direction, (D)preference.

$\eta_G^2 = 0.068$; $F(1,15) = 17.6$, $p < 0.001$, $\eta_G^2 = 0.25$; $F(1,15) = 17.2$, $p < 0.001$, $\eta_G^2 = 0.26$; $F(1,15) = 47.4$, $p < 0.001$, $\eta_G^2 = 0.17$). Significant main effects of arm presentation on Q2 and Q4 were also found($F(1,15) = 17.5$, $p < 0.001$, $\eta_G^2 = 0.14$; $F(1,15) = 8.07$, $p = 0.012$, $\eta_G^2 = 0.092$). Furthermore, there was a significant interaction for Q1 between head and arm presentations ($F(1,15) = 5.82$, $p = 0.029$, $\eta_G^2 = 0.013$). Post hoc multiple comparisons were conducted for Q1, but no significant differences were observed between conditions.

Gaze Direction. The proportions of time during each trial spent with a gaze directed toward the robot, the materials (total for Materials A and B), and other directions are shown in Fig. 4(C). Errors in HoloLens 2 gaze tracking occurred in 4 trials. As we adopt a within-participant setting, the data for 4 participants with errors were entirely excluded, including trials without errors, resulting in the analysis being conducted on data from the remaining 12 participants. However, for evaluation items other than gaze direction analysis, the data from all 16 participants were included since stimulus presentation did not change based on the failure of gaze direction recognition.

In the gaze direction analysis, we use the percentage of time spent with the gaze directed toward the robot. The mean percentage for each condition were as follows: C1: 20.7% ($SD = 7.28$), C2: 25.9% ($SD = 9.09$), C3: 30.1% ($SD = 8.10$), and C4: 32.9% ($SD = 9.12$). Statistical analysis revealed a significant main effect of head presentation ($F(1,10) = 17.2$, $p = 0.002$, $\eta_G^2 = 0.095$).

Preference. The ranking of preference for the four conditions obtained from interviews is shown in Fig. 4(D). Wilcoxon signed-rank tests with Bonferroni correction revealed significant differences between conditions for both C1 - C4 ($p = 0.026$) and C3 - C4 ($p = 0.027$).

Quality of AR Presentation. The result of the questionnaire about AR presentation quality is as follows: Resolution was rated at an average of 2.81 ($SD = 0.81$), brightness at an average of 3.38 ($SD = 1.49$), and smoothness of movement at an average of 4.31 ($SD = 1.26$).

5 Experiment 2:Multi-party Discussion Task

In this section, we describe the experiment to evaluate the effects of the presentation modalities of head and arms in multi-party discussion tasks in which interaction with other participants is important. As with Experiment 1, this experiment was conducted with the approval of the ethics review committee of the authors' affiliated organization.

5.1 Experiment Design

In the experiment, we use the system described in Section 3 to set up a task where a remote participant (experimenter) and two local participants have a discussion. The discussion topic was consensus games, in which participants discussed the priority of three items under a given scenario. For each trial, participants first read a paper describing the scenario and three candidate items and decided the personal priority of items. After that, participants initiated the discussion. Participants first explained the items they had chosen and the reasons for their selection, then had a discussion to decide the priority of the three items as a group consensus within five minutes. Since this experiment was designed as a within-participant design with four conditions, we prepared four scenarios. During the task, the remote participant (experimenter) was instructed to follow predefined guidelines for verbal and nonverbal information presentation. The verbal guidelines were to select an item that was not chosen by the other two participants at the beginning and agree with their opinions at the end. The nonverbal guidelines were to gaze at the speaker when someone else was speaking, point at the paper when explaining the selected item, and look at both the paper and the participants approximately equally. The head and arm tracking functions of the system were utilized for controlling the robot. To enhance the stability of movement and accuracy in the presentation direction, thresholds were established around each target. When the head or arms were oriented within these thresholds, the robot's head and arms were adjusted to align with the corresponding target.

The experimental setup is shown in Fig.5(A, B). The three candidate items were placed in the center of the participants as cards. From the local participant, the angle between the other local participant and the robot was approximately 64 degrees, causing the robot to be outside the AR presentation range when viewing another participant. We recruited 16 voluntary participants(9 females, 7

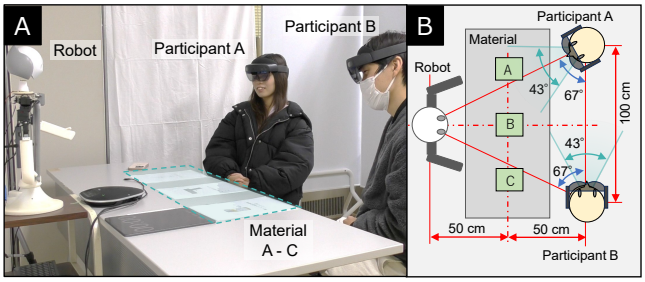

Figure 5: Experimental setup. (A)actual environment, (B)overview of the layout

males; mean age = 21.1, $SD = 1.45$) from undergraduate and graduate students at the local university. All participants were affiliated with an undergraduate or graduate department in the computer science field. Regarding familiarity with AR/VR, participants responded using a 7-point Likert scale (1: beginner, 7: expert), with an average score of 3.43 ($SD = 1.55$). The experiment consisted of pre-experiment instruction, a practice task, a main task and questionnaires in each condition, and an interview after all conditions were completed. The overall duration of the experiment was approximately 90 minutes, and participants received compensation following the guidelines of the authors' affiliated organization.

5.2 Conditions

In this experiment, we set four conditions same as in Experiment 1. The appearance of each condition is also the same (shown in Fig.1(B1)-(B4)). We adopted a within-participant design, and considering the order effects of each condition and the scenario, we counterbalanced the order for each condition and the scenario used in each condition by Latin squares.

5.3 Measures

The measures are also the same as in Experiment 1: social presence, clarity of body status, gaze direction, semi-structured interview about preference, and the quality of AR Presentation. For the questionnaire on the clarity of body status, we added two additional items related to other local participants. In this experiment, the gaze detection and recording features of HoloLens 2 were used to capture each user's gaze direction. Specifically, the video of what the user was looking at was recorded with gaze points detected by the device. Three annotators then annotated these videos, and the time each user spent looking at each target (the robot, cards, another participant, or others) was calculated.

5.4 Results

For the analysis, a two-way repeated-measures ANOVA with Aligned Rank Transform (ART) [44] was conducted, considering the factors of head and arm presentation (physical vs. AR). We set the significance level at 0.05 for all analyses.

Social Presence. The results of each subscale are shown in Fig.6(A). Statistical analysis revealed significant main effects of head presentation on Co-presence (CoP), Perceived Message Understanding (PMU)($F(1,15) = 12.81$, $p = 0.003$, $\eta_p^2 = 0.46$; $F(1,15) = 7.38$, $p = 0.016$, $\eta_p^2 = 0.33$,). A significant main effect of arm presentation was also observed in Attentional Allocation (AA) ($F(1,15) = 9.04$, $p = 0.009$, $\eta_p^2 = 0.38$).

Clarity of Body Status. Errors related to the experimental setting(card position error) occurred in two groups, so we excluded the data of these groups, and analysis was conducted on data from the remaining 12 participants. The results of each question are shown in Fig.6(B). Statistical analysis revealed significant main effects for head presentation on Q1, Q3, Q5, Q6, Q7, and Q8 ($F(1,11) = 13.84$, $p = 0.0034$, $\eta_p^2 = 0.56$; $F(1,11) = 4.98$, $p = 0.047$, $\eta_p^2 = 0.31$; $F(1,11) = 8.17$, $p = 0.016$, $\eta_p^2 = 0.43$; $F(1,11) = 6.75$, $p = 0.025$, $\eta_p^2 = 0.38$; $F(1,11) = 9.08$, $p = 0.012$, $\eta_p^2 = 0.45$; $F(1,11) = 7.94$, $p = 0.017$, $\eta_p^2 = 0.42$). The main effects of arm presentation were not found.

Gaze Direction. The proportions of time during each trial spent with a gaze directed toward the robot, the materials (total for Materials A, B, and C), another local participant, and other directions are shown in Fig.6(C). Errors related to the recording function of HoloLens 2 occurred in two groups. As we adopted a within-participant experimental design, the data for the two groups(4 participants) were entirely excluded in the analysis of gaze direction, including trials without errors, resulting in the analysis being conducted on data from the remaining 12 participants.

In the gaze direction analysis, we use the percentage of time spent with the gaze directed toward the robot, the same as in Experiment 1. The mean percentage for each condition were as follows: C1: 19.1 % ($SD = 10.3$), C2: 23.3 % ($SD = 15.8$), C3: 41.0 % ($SD = 24.8$), C4: 24.1 % ($SD = 14.7$). Statistical analysis revealed significant main effects of head presentation and interaction between head and arm presentations($F(1,11) = 15.1$, $p = 0.0025$, $\eta_p^2 = 0.58$; $F(1,11) = 6.83$, $p = 0.024$, $\eta_p^2 = 0.38$). Post hoc multiple comparisons revealed a significant difference between the C1 and C3 conditions ($p = 0.024$).

Preference. The ranking of preference for the four conditions obtained from interviews is shown in Fig.6(D). Wilcoxon signed-rank tests with Bonferroni correction revealed significant differences between conditions for C1-C4 ($p = 0.0066$).

Quality of AR Presentation. The result of the questionnaire about AR presentation quality is as follows: Resolution was rated at an average of 2.19 ($SD = 0.81$), brightness at an average of 4.0 ($SD = 1.54$), and smoothness of movement at an average of 3.06 ($SD = 1.75$).

6 Discussion

6.1 Summary of Results

Social Presence. The results for social presence were consistent across both Experiment 1 and Experiment 2. In both experiments, physical head led to significantly higher ratings for Co-Presence (CoP) and Perceived Message Understanding (PMU) than AR. On the other hand, in Attentional Allocation (AA), physical arms resulted in significantly lower ratings compared to AR. In line with this result, an interview comment such as, "The arms drew attention to the robot, making it harder to focus on the picture" (Exp.1-P7), was noted.

Clarity of Body Status. For the clarity of body status, items related to gaze showed a significant main effect of head presentation in

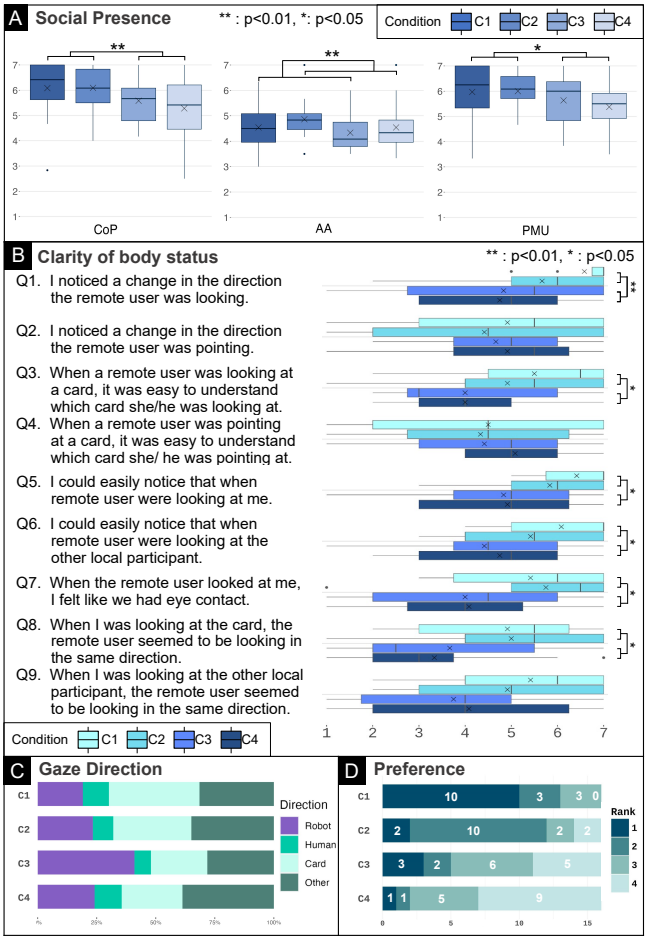

Figure 6: Result of experiment 2. (A)social presence, (B)clarity of body status, (C)gaze direction, (D)preference.

both experiments(Exp.1: Q3, Q5-Q7; Exp.2: Q1, Q3, Q5-Q8), with physical head presentation receiving a higher score than AR. These items are related to mutual gaze and joint attention, which play important roles in human interaction. The results indicate that physically presenting the head improves their transmission. For pointing gesture, there was a main effect of the arms in Experiment 1, where physically presenting the arms improved the awareness and clarity of pointing gestures (Q2, Q4). In contrast, no significant main effect of the arms was found in Experiment 2, likely due to the limited number of pointing(only once per trial) and the reduced significance of pointing, as participants could comprehend the target through the context of the conversation.

Gaze Direction. In both experiments, there was a significant main effect of the head on the proportion of time spent looking at the robot. When the head was physically present, participants spent significantly less time looking at the robot, likely because they could gather information from their peripheral vision. This was further supported by interview comments, such as "The FOV was narrow, and the head was out of view, so I couldn't tell where the robot was looking, making it harder to stay focused on the conversation.(Exp.1-P1)"

Preference. For preferences, Condition C1 (Both Physical) was the most preferred, while Condition C4(Both AR) was the least. There was some variation in ratings between Conditions C2(Physical Head and AR Arms) and C3(AR Head and Physical Arms). Considering the total number of times ranked first or second, C3 ranked higher in Experiment 1 (C3 = 9, C2 = 6), whereas C2 ranked higher in Experiment 2 (C2 = 12, C3 = 5). These differences are likely due to task variations. In Experiment 1, the physical arms were important for explaining materials, as shown by the clarity of body status results. In Experiment 2, where the arms were less critical, the physical head became more important, especially in multi-person interaction where gaze played an important role.

Quality of AR Presentation. In both experiments, the resolution of the AR presentation was rated lower than the physical presentation (Exp.1 = 2.81, Exp.2 = 2.19). For brightness, the AR presentation received similar or slightly lower ratings than the physical presentation (Exp.1 = 3.38, Exp.2 = 4.0). For smoothness, the AR presentation was rated slightly higher than the physical presentation in Experiment 1 but lower in Experiment 2 (Exp.1 = 4.31, Exp.2 = 3.06). In the interview results, 11 comments in Experiment 1 and 7 comments in Experiment 2 mentioned the negative effect of limited FOV. Additionally, 2 comments in Experiment 1 and 6 comments in Experiment 2 referenced the quality of the CG presentation. These results indicate that the challenges of optical see-through HMDs, highlighted in previous studies[26], also occurred in this study's experimental setup.

6.2 Discussion on Research Question

Based on the results from the two experiments, we first discuss RQ1: "How does the hybrid configuration affect social presence when AR parts occasionally fall outside the FOV?" In both experiments, the main effects of the head were significant for the social presence subscales of CoP and PMU, while the main effects of the arms were significant for AA. These findings suggest that the hybrid configuration, where the head is physically presented and the arms are presented via AR (C2), is optimal for conveying social presence. Specifically, the physical presence of the head enhances CoP and PMU, while the AR presentation of the arms reduces AA. These results suggest that, at least in the two experimental environments we set, the hybrid robot successfully compensates for the challenges of AR and achieves a social presence comparable to or exceeding that of a fully physical robot. Crucially, it is necessary to note that the physical arms negatively affected AA potentially because the importance of pointing gestures was low in the tasks used in this study. In situations where the significance of pointing increases—such as continuous and frequent pointing in device operation instructions, which demands users' strict attention to objects—physical arms may not negatively impact AA. Therefore, evaluations in different contexts are needed. As a design implication based on the results obtained in this study, it is desirable to physically present the head when social interaction with people is emphasized. Conversely, in situations where social connection is less critical but guidance of attention via pointing is crucial (such as when explaining complex information with minimal participant interaction), it is desirable to physically present the arms.

Next, we address RQ2: "Do evaluations of the hybrid robot differ based on which body part is adopted for AR presentation?" As previously noted, the physical head significantly enhances social presence, whereas the physical arm does not. As the term "social gaze [12]" indicates, gaze plays an important role in social interaction with others, so the result indicates the importance of the physical head is not surprising. Additionally, the results for clarity of body status indicate that the effects of each body part's presentation modality differ depending on the environment and situation, and user preferences also change accordingly. These findings demonstrate that evaluations of the hybrid robot are influenced by which body part is presented via AR. Therefore, it is crucial to consider which body parts should be presented in AR when designing hybrid robots.

Beyond the aforementioned discussion on RQs, several constraints related to the experimental configuration must be considered. In this study's experimental setup, the torso (including the display and base) of the robot was presented physically in all conditions. Consequently, even Condition C4 (Both AR) is configured as a hybrid presentation combining a physical body and AR overlays, rather than an AR-centric presentation such as VROOM[21]. Therefore, we did not compare our conditions with a condition where the entire body is presented solely via AR, which warrants future investigation. However, based on the current results, it is likely that adopting a fully AR body would lead to a further decrease in social presence compared to C4, potentially demonstrating the effectiveness of the hybrid configuration more prominently.

Furthermore, the requirement for participants to wear an AR device (HoloLens 2) even in C1 (Both Physical) deviates from the ordinary usage scenario of physical robots. This configuration was necessary to maintain experimental control across all conditions in this study. However, to anticipate real-world usage, a comparison with a condition where participants interact with the fully physical robot without wearing an AR device is needed.

6.3 Limitation

A notable limitation of this study is that the results obtained are closely tied to the quality of the AR presentation. Specifically, the research can be regarded as a case study using an optical see-through AR device with relatively low quality—not only in terms of its limited FOV but also its resolution. The use of higher-quality AR-HMDs, such as the JVC HMD-VS1W [40] with a 120-degree horizontal FOV, may reduce the disadvantages of AR and potentially yield different results. However, commercially available AR devices like the Xreal Air 2 [42], which prioritize portability, lightweight design, and casual use, hold a significant market share. We expect an increase in devices like the Xreal Air 2 Ultra [43], which balances casual usability with spatial recognition capabilities. Therefore, this study's results can be considered valuable insights into devices aimed at casual use rather than high-end.

The experimental setting is also a limitation of this study. In this study, evaluations were mainly based on questionnaires, and the impact of the system on participants' behavior and conversation in actual conversational situations was not evaluated. This was due to controlling factors other than condition as much as possible to investigate the effects of arm and head modality in detail. Such an experimental design has been employed in a related study [38], and we believe it was valid for our purpose. However, it is also important to evaluate the impact of the system on the participants' behavior, participation, and conversation content, so we need to conduct evaluations in more realistic settings. Additionally, since the age, gender, and affiliation of the participants in the experiments were limited, it will be necessary to conduct evaluations with diverse participants in the future.

6.4 Futurework

In this study, we limited the information presented in AR to 3D data with exactly the same form as that of a physical robot. However, AR presentation is not limited by physical constraints, so more human-like expressions, such as nuanced facial expressions or hand gestures, or superhuman expressions, such as extending arms, can be added by AR. Furthermore, it is also possible to place AR objects, such as text information, images, video, etc., around the robot. In our future work, we will examine how the robot's physical body should be combined with such advanced AR expressions.

7 Conclusion

In this paper, we investigate a hybrid robot that integrates physical and AR body parts, focusing on how the presentation modalities of the head and arms impact human perception of social presence. We implemented a hybrid robot system that can switch between physical and AR modalities for the head and arms, and conducted two experiments with different tasks: material explanation and multi-party discussion. The results of experiments indicate that, in the two experimental environments examined, the hybrid robot effectively addresses the challenges of AR device and achieves a social presence comparable to or exceeding that of a fully physical robot. Additionally, the findings reveal that evaluations of the hybrid robot are influenced by the specific body parts presented via AR, highlighting the importance of careful consideration regarding which body parts to include in AR during the design of hybrid robots. As a limitation, findings from these experiments were mainly based on subjective evaluations and did not assess the impact on participants' behavior or the content of the conversation. Furthermore, participants were limited in age, gender, and affiliation. Therefore, targeting a more diverse range of participants and conducting evaluations in a more realistic setting will be necessary. Furthermore, We only used AR data with the exact same shape as a physical robot, but since AR enables more nuanced expressions that are closer to humans or superhuman expressions, so we will work on such applied expressions in Future work.

Acknowledgments

The authors acknowledge the use of generative AI in preparing this manuscript and developing the system application. Specifically, ChatGPT and Claude were utilized to assist in the writing process and certain aspects of application development. These AI tools contributed to refining the manuscript's language and aided in generating code snippets for the system application. However, all core research ideas, experimental design, data analysis, and critical interpretations were conducted by the human authors. This work was supported by JSPS KAKENHI Grant Number 24K20808.

References

[1] Sigurdur Orn Adalgeirsson and Cynthia Breazeal. 2010. MeBot: A robotic platform for socially embodied telepresence. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 15–22.

[2] Henny Admoni and Brian Scassellati. 2017. Social eye gaze in human-robot interaction: a review. *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63.

[3] Aldebaran. 2020. *Telepresence by Pepper.* Retrieved September 12, 2024 from https://www.aldebaran.com/en/blog/videos/telepresence-pepper

[4] Stuart M Anstis, John W Mayhew, and Tania Morley. 1969. The perception of where a face or television'portrait'is looking. *The American journal of psychology* 82, 4 (1969), 474–489.

[5] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. 2020. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[6] Giulia Barbareschi, Midori Kawaguchi, Hiroaki Kato, Masato Nagahiro, Kazuaki Takeuchi, Yoshifumi Shiiba, Shunichi Kasahara, Kai Kunze, and Kouta Minamizawa. 2023. "I am both here and there" Parallel Control of Multiple Robotic Avatars by Disabled Workers in a Café. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[7] Jacob T Biehl, Daniel Avrahami, and Anthony Dunnigan. 2015. Not really there: Understanding embodied communication affordances in team perception and participation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1567–1575.

[8] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments* 12, 5 (2003), 456–480.

[9] Andriana Boudouraki, Joel E Fischer, Stuart Reeves, and Sean Rintel. 2021. " I can't get round" Recruiting Assistance in Mobile Robotic Telepresence. *Proceedings of the ACM on Human-computer Interaction* 4, CSCW3 (2021), 1–21.

[10] Eric Deng, Bilge Mutlu, Maja J Mataric, et al. 2019. Embodiment in socially interactive robots. *Foundations and Trends® in Robotics* 7, 4 (2019), 251–356.

[11] Paul Dourish. 2001. *Where the action is: the foundations of embodied interaction*. MIT press.

[12] Nathan J Emery. 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews* 24, 6 (2000), 581–604.

[13] Charith Lasantha Fernando, Masahiro Furukawa, Tadatoshi Kurogi, Sho Kamuro, Kouta Minamizawa, Susumu Tachi, et al. 2012. Design of TELESAR V for transferring bodily consciousness in telexistence. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 5112–5118.

[14] Naomi T Fitter, Luke Rush, Elizabeth Cha, Thomas Groechel, Maja J Matarić, and Leila Takayama. 2020. Closeness is key over long distances: Effects of interpersonal closeness on telepresence experience. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*. 499–507.

[15] Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction* 19, 3 (2004), 273–309.

[16] Thomas Groechel, Zhonghao Shi, Roxanna Pakkar, and Maja J Matarić. 2019. Using socially expressive mixed reality arms for enhancing low-expressivity robots. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–8.

[17] Zhao Han, Yifei Zhu, Albert Phan, Fernando Sandoval Garza, Amia Castro, and Tom Williams. 2023. Crossing Reality: Comparing Physical and Virtual Robot Deixis. In *2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[18] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence. In *Seventh annual international workshop: Presence*, Vol. 2004. Universidad Politecnica de Valencia Valencia, Spain.

[19] Keiichi Ihara, Mehrad Faridan, Ayumi Ichikawa, Ikkaku Kawaguchi, and Ryo Suzuki. 2023. HoloBots: Augmenting Holographic Telepresence with Mobile Robots for Tangible Remote Collaboration in Mixed Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–12.

[20] Andrew Irlitti, Mesut Latifoglu, Thuong Hoang, Brandon Victor Syiem, and Frank Vetere. 2024. Volumetric Hybrid Workspaces: Interactions with Objects in Remote and Co-located Telepresence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.

[21] Brennan Jones, Yaying Zhang, Priscilla NY Wong, and Sean Rintel. 2021. Belonging there: VROOM-ing into the uncanny valley of XR telepresence. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.

[22] Ikkaku Kawaguchi, Hideaki Kuzuoka, and Yusuke Suzuki. 2015. Study on gaze direction perception of face image displayed on rotatable flat display. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1729–1737.

[23] Adam Kendon. 1990. *Conducting interaction: Patterns of behavior in focused encounters*. Vol. 7. CUP Archive.

[24] Hideaki Kuzuoka, Jun'ichi Kosaka, Keiichi Yamazaki, Yasuko Suga, Akiko Yamazaki, Paul Luff, and Christian Heath. 2004. Mediating dual ecologies. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 477–486.

[25] Hideaki Kuzuoka, Yuya Suzuki, Jun Yamashita, and Keiichi Yamazaki. 2010. Reconfiguring spatial formation arrangement by robot body orientation. In *2010 5th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 285–292.

[26] Lik-Hang Lee, Tristan Braud, Simo Hosio, and Pan Hui. 2021. Towards augmented reality driven human-city interaction: Current research on mobile headsets and future challenges. *ACM Computing Surveys (CSUR)* 54, 8 (2021), 1–38.

[27] Le Luo, Dongdong Weng, Jie Hao, Ziqi Tu, and Haiyan Jiang. 2023. Controllable Telepresence: A Robotic-Arm-Based Mixed-Reality Telecollaboration System. *Sensors* 23, 8 (2023), 4113.

[28] Gerhard Sigurd Nielsen. 1964. *Studies in Self Confrontation: Viewing a Sound Motion Picture of Self and Another Person in a Stressful Dyadic Interaction.* Munksgaard.

[29] Yuya Onishi, Kazuaki Tanaka, and Hideyuki Nakanishi. 2014. PopArm: A robot arm for embodying video-mediated pointing behaviors. In *2014 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 137–141.

[30] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*. 741–754.

[31] Eric Paulos and John Canny. 1998. PRoP: Personal roving presence. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 296–303.

[32] Thammathip Piumsomboon, Gun A Lee, Jonathon D Hart, Barrett Ens, Robert W Lindeman, Bruce H Thomas, and Mark Billinghurst. 2018. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[33] Thammathip Piumsomboon, Gun A Lee, Andrew Irlitti, Barrett Ens, Bruce H Thomas, and Mark Billinghurst. 2019. 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*. 1–17.

[34] Mose Sakashita, Hyunju Kim, Brandon Woodard, Ruidong Zhang, and François Guimbretière. 2023. VRoxy: Wide-Area Collaboration From an Office Using a VR-Driven Robotic Proxy. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[35] Mose Sakashita, Tatsuya Minagawa, Amy Koike, Ippei Suzuki, Keisuke Kawahara, and Yoichi Ochiai. 2017. You as a puppet: evaluation of telepresence user interface for puppetry. In *Proceedings of the 30th annual ACM symposium on user Interface software and technology*. 217–228.

[36] Mose Sakashita, E Andy Ricci, Jatin Arora, and François Guimbretière. 2022. RemoteCoDe: Robotic Embodiment for Enhancing Peripheral Awareness in Remote Collaboration Tasks. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–22.

[37] Mose Sakashita, Ruidong Zhang, Xiaoyi Li, Hyunju Kim, Michael Russo, Cheng Zhang, Malte F Jung, and François Guimbretière. 2023. ReMotion: Supporting Remote Collaboration in Open Space with Automatic Robotic Embodiment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[38] David Sirkin and Wendy Ju. 2012. Consistency in physical and on-screen action improves perceptions of telepresence robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. 57–64.

[39] Double Robotics, Inc. 2024. *Double Robotics - Telepresence Robot for the Hybrid Office.* Retrieved September 12, 2024 from https://www.doublerobotics.com/

[40] JVCKENWOOD Corporation. 2021. *HMD-VS1DW.* JVCKENWOOD Corporation. Retrieved September 12, 2024 from https://www.jvc.com/usa/pro/projectors/head-mounted-display/hmd-vs1dw/

[41] Xandex, Inc. 2024. *KUBI Telepresence Robot.* Retrieved September 12, 2024 from https://www.kubiconnect.com/

[42] XREAL, Inc. 2023. *XREAL Air 2.* Retrieved September 12, 2024 from https://www.xreal.com/air2

[43] XREAL, Inc. 2024. *XREAL Air 2 Ultra.* Retrieved September 12, 2024 from https://www.xreal.com/air2ultra

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

[45] Nicole Yankelovich, Nigel Simpson, Jonathan Kaplan, and Joe Provino. 2007. Porta-person: Telepresence for the connected conference room. In *CHI'07 extended abstracts on Human factors in computing systems*. 2789–2794.