Evaluating Remote Virtual Hands Models on Social Presence in Hand-based 3D Remote Collaboration

Boram Yoon* KAIST UVR Lab
Hyung-il Kim† KAIST UVR Lab
Seo Young Oh‡ KAIST UVR Lab
Woontack Woo§ KAIST UVR Lab

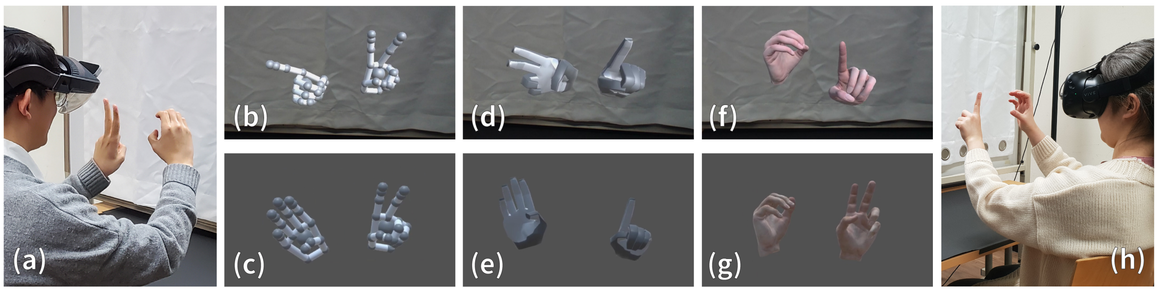

Figure 1: Our user study setup: (b)-(g) The participant's augmented reality (AR) and virtual reality (VR) views through a head-mounted display (HMD), showing different remote virtual hands representations. The hands in (b)-(g) are posing different letters of American Sign Language (ASL) and specific numbers during our user task: (a) Local user under AR conditions, (b) AR-skeleton hands, (c) VR-skeleton hands, (d) AR-low polygon hands, (e) VR-low polygon hands, (f) AR-realistic (female) hands, (g) VR-realistic (male) hands, and (h) remote user under VR conditions.

Abstract

This study investigates the effects of a virtual hand representation on the user experience including social presence during hand-based 3D remote collaboration. Although a remote hand appearance is a critical parts of a hand-based telepresence, it has been rarely studied in comparison to studies on the self-embodiment of virtual hands in a 3D environment. Thus, we conducted a user study comparing the three virtual hands models (Skeleton, Low Polygon and Realistic) while performing a remote collaborative task based on the American Sign Language (ASL) using both Augmented Reality (AR) and Virtual Reality (VR) environments. We found that the realistic type was perceived as the most sense of being together, human-like, and trustable representation. The low polygon model could also convey a clear sign and moderate level of social presence. Although the system was configured asymmetrically in AR and VR, little difference in perception was found except for the participant's mental load and message understanding. We then discuss the results and suggest design implications for future hand-based 3D telepresence systems.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed/augmented reality; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

1 Introduction

This study explores the effects of a virtual-hand representation on user experience including social presence, presence, trust, and preference in a hand-based 3D remote collaboration. Technical advances in head-mounted displays (HMDs) and tracking devices have increased the potential and interest in 3D-based remote collaboration over the last decades. This interest has increased not only in the research field but also in many commercial technology companies, and a variety of systems have been developed, from realistic 3D image reconstructed telepresence to avatar-mediated telepresence.

More avatar-mediated 3D telepresence methods have been introduced commercially in recent years because they require relatively low cost and fewer devices compared to 3D image reconstruction. Furthermore, with the introduction of Facebook Social VR [10], Magic Leap Social [38], and Spatial [59] with Microsoft Hololens 2, social networking services have been extended to the 3D platforms such as Virtual Reality (VR) and Augmented Reality (AR). Mixed Reality (MR) based services also have the potential to become popular in the future due to the interoperability between heterogeneous devices and the environmental setups among participanting users.

To be specific, an asymmetric MR setup can overcome the disadvantages of a specific environmental configuration of both types of symmetric 3D remote collaboration. For instance, it is critical to synchronize real environmental conditions of each collaborator in a symmetric AR setup, and a VR setup only supports limited applications that are unaffected by their use in virtual reality. Therefore, an asymmetrical AR and VR setup can successfully utilize VR for remote sites to reduce environmental limitations and AR for local sites to cover more collaborative tasks in reality.

Unlike traditional 2D-based online social activities, a 3D telepresence system enables more diverse collaborations for users by augmenting their virtual body parts on an HMD, similar to real life. A remote expert can support local users by delivering action guides to perform the proper tasks, as demonstrated by several companies [9, 12]. Moreover, remote education [66] or collaborative games that involve physical activities, which are also challenging tasks in 2D videoconferencing, can also be performed.

Although 3D telepresence has the advantage of capturing and transmitting body movements in 3D data [45], showing an entire avatar requires high computational cost and system overhead. In some scenarios, however, expressing only certain body parts is sufficient to accomplish the desired task. For example, by utilizing only the hands, a wide range of applications can be supported, including remote assistance [2, 26, 51, 64], medical surgery [1], educational

*e-mail: boram.yoon1206@kaist.ac.kr
†e-mail: hyungil@kaist.ac.kr
‡e-mail: seoyoung.oh@kaist.ac.kr
§e-mail: wwoo@kaist.ac.kr

978-1-7281-8508-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ISMAR50242.2020.00080

scenarios, and board games [3,26,62,71]. In such hand-based tasks, the user's attention is mainly focused on the collaborator's hand expressions. Thus, it is essential for a system to express the hand movements properly.

The selection of an appropriate virtual hand model, particularly with a proper understanding of each 3D environment, is an important factor for achieving a satisfying user experience. However, little research has focused on remote hand embodiment on both the AR and VR sides; several previous studies on telepresence explored only the effect of a full or half body avatar [16, 24, 72]. Although there have been many studies related to self-presence, sense of ownership or agency of one's own virtual hands in VR [4, 28, 44, 56, 57], but such studies do not cover the remote partner's virtual hands and have been conducted only in a VR environment. Since 3D telepresence is applied to both environments and can be connected asymmetrically under MR conditions, it is necessary to investigate also in AR.

Thus, we explored the effect of a visual representation of remote hands based on subjective measurements including social presence and task performance. To undertake this research, an MR-based remote collaboration environment was constructed, and collaborative user tasks were conducted with pairs of recruited participants. Three virtual hands models derived from the different degrees of realistic expression were evaluated in both AR and VR environments (in Figure 1(a)-(h)); the experimental conditions were set as a combination of two main factors: Virtual Hands Type and 3D Environment Type. From the user experiment, we found valuable insights for the design and implementation of a future hand-based MR telepresence system. As a result, the novel contributions of this study are as follows:

- A user study (*n*=48) was conducted on the remote hands embodiment of the collaborator, which has been rarely explored in previous studies.
- We compared the overall user experience in both AR and VR environments during synchronous communication, and investigated the possibility of a future implementation for a 3D asymmetric remote collaboration.
- The implications for designing an avatar-mediated 3D collaborative system interface were discussed, particularly for hand-based scenarios.

2 Related Work

This research extends earlier studies on 3D remote collaboration including examples for hand-based scenarios, and the various perceptions of users on both a full body remote avatar and a virtual hands representation.

2.1 3D Remote Collaboration and Hand-based Scenario

As the interest in 3D remote collaboration has increased throughout the academic and commercial fields, telepresence systems have been developed for supporting simple social activities as well as targeting more specific purposes. 3D environmental conditions such as VR, AR, and MR have been utilized to deliver an immersive social experience, in which the users are shown as virtual avatars or through reconstructed images using their full or partial bodies.

Previous research on VR collaborative systems have enabled remote users to perform social activities through shared virtual spaces in avatar-mediated telepresence [11, 13, 46, 58, 60], life-size projections [55], or multi-user collaboration platforms [20, 54, 68]. AR-based telepresence augments remote users in the real-world of the local user by applying 3D video image reconstruction [6,45,47,53], or a virtual avatar on an HMD [38, 40, 59]. In the case of MR-based remote collaboration, recent studies have introduced telepresence systems augmenting a remote VR user into the real space of a local AR user [41,48–52,63,72]. Many companies have also demonstrated 3D remote collaboration systems by leveraging VR [10, 34, 39] or AR platforms [38, 40, 59] to help the users communicate with distant friends and family.

In these 3D remote collaborations, a full-body avatar or 3D volumetric image is usually selected to represent the real user virtually, and thus it involves space−time costs because a significant amount of information needs to be displayed. However, in a more specialized collaboration scenario, remote collaboration can also be implemented by representing only the essential parts of the partner's body, such as the hands, rather than the full body [2,3,18,19,22,27,62,71]. One example of this is a situation in which a remote expert provides instructions, training, or support for a particular task to a local user (or novice). To make these scenarios possible, many previous studies have developed systems displaying a remote user's hands during collaboration on a 2D screen at the local user side [2, 18–20, 62, 71], or showing the virtual hands through a projector [22, 27] or HMD [3, 26, 50, 64, 65, 73].

Moreover, it has been suggested by many researchers using the user's virtual hands as a hand gesture cue by transmitting the image of the hands to the view of the local user [3, 26, 27, 64, 65], and also applied to commercial services such as DAQRI Worksense [9] and Google Glass Enterprise [12]. Remote 3D education for hand gesture-based activities such as playing musical instruments, handicrafts, cooking, and sign language can be effective examples of hand-based collaboration in the future. Game playing such as building blocks, or putting together puzzles can also be developed by augmenting the participated user's hands [3, 26, 62, 71]. However, there have been few studies on evaluating remote virtual hands or their communication efficiency, particularly at the user side, despite numerous future scenarios that can be utilized through hand-based 3D remote collaboration systems.

2.2 Effects of Avatar Appearance in 3D Environment

Because avatar-mediated 3D telepresence delivers the remote embodiment of another party, it is essential to design the virtual avatar to provide a better perception to local users participating in a remote collaboration. Body embodiment research conducted in a 3D environment has been actively investigating both self- and remote embodiment [16, 23, 24, 30, 72, 74]. Self-embodiment measures the feeling of self-body representation, whereas remote embodiment indicates one's feeling of the partner's virtual body, and its effectiveness has been evaluated based on social presence or trust.

Latoschik et al. [30] evaluated the difference of the full-body avatar divided by the avatar realism in immersive VR with social presence and trust and concluded that the avatar with the highest fidelity achieved a strong acceptance by users despite evoking an uncanny feeling. Jo et al. [24] explored whether a cartoon or a realistic avatar causes higher co-presence and trust in real and virtual background environments. They found that co-presence was higher for a cartoon model, although the realistic model delivered a higher level of trust to users.

In addition, Yoon et al. [72] compared social presence of avatar models in an AR-based remote collaboration. Six different conditions divided by the variables of body part visibility and character style were set in their study, and it was found that the whole-body condition increased the user's feeling of social presence. Moreover, Yoon et al. [72] found no difference in social presence between cartoon and realistic avatars, and thus they suggested that such models can be selected based on collaboration context.

There have also been other interesting studies related to the remote avatar models [16,23,74], but most of these studies have investigated based on full body avatars. Thus, studies that focus more on partial bodies, which can convey communication cues, such as the hands, also need to be explored. In addition, a study above described the concern regarding the use of only head and hands type avatar [72], yet they conducted tasks focused on seeing the whole-body avatar. In situations using hand-based communication, different outcomes might likely be derived.

2.3 Virtual Hand Representation and User Perception

There is still a lack of research on how to best demonstrate the remote hand embodiment in a 3D remote collaboration, although many methods have been developed that allow the user's hands to be transmitted to the shared space. In a hand-gesture based scenario, communication between collaborating users can be sufficiently supported by sharing only their virtual hands [3, 19, 20, 22, 26, 27, 62–65, 71]. The participating user can perceive a virtual representation of their own hands and the hands of the collaborator at the same time, and thus, the study should investigate both sides. However, previous studies have mostly focused on a subjective perception when expressing one's actual hands in a first-person view through a virtual hands model in a VR environment [4, 17, 28, 32, 33, 35–37, 44, 56, 57].

Argelaguet et al. [4] evaluated the effectiveness among three hand models with different degrees of realism in VR, and concluded that a less realistic model induces a higher agency of the virtual hand and a realistic model has higher ownership. Knierim et al. [28] conducted a virtual keyboard typing task in VR, and found that the realistic model has the lowest task load, whereas the user presence did not show any differences. By contrast, Grubert et al. [14] investigated the effect of virtual hands in the same typing task, but they concluded that a minimalist model such as a fingertip enhanced the task performance whereas the most realistic model decreased the performance. Ogawa et al. [44] found that a realistic hand had a better perception in estimating virtual object sizes within the VR environment. The user's feeling of presence towards the virtual hands was also investigated by Schwind et al., including abstract and realistic styles, in terms of gender perception [57] or a reduced number of fingers [56].

Other earlier studies also measured hand illusion, ownership, and self-presence based on experimental factors such as a structural distortion of the hands [17], visual transparency [25], and more diverse properties [32, 33, 37]. However, most of these studies were mainly focused on an evaluation of one's own hands and also assumed a VR environment. ShadowHands [71] and ShowMe [3] each suggested using toon-style rendered hands, and a 3D hand mesh mapping or blob styles as a way of representing the hands of a remote partner under a remote collaboration. However, these two studies explored the system configuration itself rather than performing quantitative/qualitative assessments on the user side.

Therefore, to implement a suitable remote hands embodiment for a near future 3D hand-based telepresence, we explored how the user's evaluation differs in terms of social presence, presence, trust, and preference when the collaborator's virtual hands are provided differently. A difference in performance was observed by measuring the mental effort and completion time when evaluating which condition showed the highest communication effectiveness. Although previous hand studies have been mostly conducted using VR, but the application of a telepresence system in AR or MR is also expected in the future. Thus, this study also focuses on an evaluation of both AR and VR for comparing the difference between the user responses.

3 Methodology

The main motivation of this study is to explore how the representation of a collaborator's hands affects the overall user experience. To achieve this, we set the following research questions.

- RQ 1. How is the overall hand-based 3D remote collaboration experience different within various remote hands representations, and what features should be considered in the future?
- RQ 2. Under what condition would users feel the highest social presence and positive feelings when the collaborator's virtual hands are represented with different levels of visual appearance in terms of realism?
- RQ 3. How would a user evaluation regarding social presence differ between AR and VR conditions, and it is possible to connect in an MR configuration asymmetrically?

3.1 Experimental Conditions and Hypotheses

According to previous research, we set two experimental factors: 1) Virtual Hands Type (*hands*) and 2) 3D Environment Type (*environment*). The first factor refers to a virtual hand representation style, and thus we differentiated it based on the degree of visual appearance. Based on the first factor, we derived three experimental conditions: (1) Skeleton (*Skeleton*), (2) Low Polygon (*LowPoly*), and (3) Realistic (*Realistic*) (see in Table 1). In many previous studies, virtual avatar realism has been investigated as a factor of virtual hands representation in both self and remote embodiment studies [4, 16, 24, 28, 30, 44, 56, 57, 72, 74]. Zibrek et al. also mentioned that the realism of an avatar can be a positive choice for its virtual expression [74]. The realism of remote virtual hands has not been evaluated to a large extent in 3D hand-based remote collaboration, especially in the case of remote embodiment, thus, we set as the first factor. Second, we also found the necessity of exploring remote hand embodiment in both AR and VR 3D environments according to earlier studies. We therefore set two more experimental conditions based on the second factor, i.e., *AR* and *VR* environments. Consequently, we derived six experimental conditions from the combination of a 3x2 factorial design. Based on these conditions and earlier studies, the following hypotheses were assumed.

- H1. Virtual hands type (*hands*) will affect the social presence, presence, and trust in hand-based MR remote collaboration.
- H2. A realistic hands type will have the highest social presence, presence and trust.
- H3. A realistic hands type will have the lowest mental effort and the highest task performance.
- H4. A 3D environment (*environment*) will affect the social presence, presence, and trust in hand-based MR remote collaboration.
- H5. The participants in AR will show higher social presence, presence, and trust than the participants in VR.
- H6. The participants in VR will require lower mental effort than the participants in AR.

H1, **H2**, and **H3** are related to the first experimental factor (*hands*), and are established based on previous studies. Jo et al. [24] found that the realistic model had a higher trust and positive user perception, and Yoon et al. [72] also mentioned that social presence would be higher if the model could provide a more detailed communication cue. Latoschik et al. [30] and Knierim et al. [28] both revealed that user acceptance was stronger on the most high-fidelity avatars, and that the lowest task load was observed toward realistic hands. In our hypotheses, the user's preference regarding each hands model is not supposed because it was expected to have less difference due to its ambivalently accepted nature depending on the participant's personal cognition [24, 30, 57].

H4, **H5**, and **H6** are related to our second factor (*environment*). It was previously revealed that a real background would provide a higher co-presence to a local user during a remote collaboration [24]. According to Witmer and Singer's definition [69], presence indicates the subjective experience of being in a one place. AR condition provides the real environmental surroundings, and the participants can work with their real hands; thus, **H5** is assumed. Although we expected that the real environment might show better scores in terms of user perception, the previous studies have argued that the AR condition has a cognitive effect on the user's mental effort due to the complexity of the environment [29, 31]: In AR, the participants should face both virtual and real heterogeneous information. Because both AR and VR conditions require wearing an HMD, except for the load caused by the device, AR therefore may result in a higher overload. **H6** only supposes mental effort because we can collect and compare the task completion time for different hands types owing to the paired communication between the AR and VR conditions.

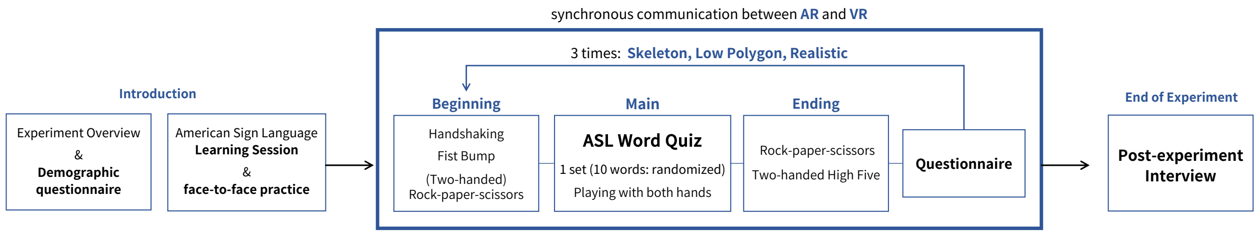

Figure 2: Flow chart of overall study procedure

	Virtual Hands Type			
	(1) Skeleton	(2) Low Polygon	(3) Realistic	
Hands Model			Male	Female

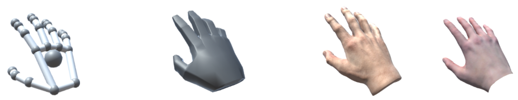

Table 1: Three experimental conditions derived from the first factor, virtual hands type (*hands*).

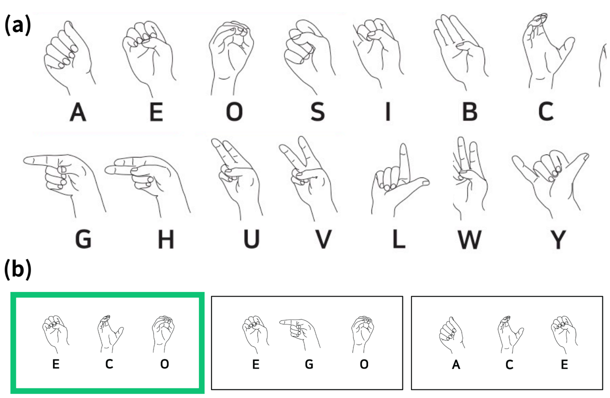

Figure 3: (a) Representation of 14 alphabet letters in American Sign Language (ASL) and (b) example of an ASL word quiz: The answer is shown in the green box and all three words were provided in multiple choice format

3.2 Study Design

The experiment was in a mixed factorial design with a within-subject variable virtual hands type (*hands*) and between-subject variable 3D environment type (*environment*). For *hands*, we had three different levels (skeleton, low polygon, and realistic); the second independent variable *environment* had two levels (AR and VR) (Figure 1). The experiment was conducted with a pair of recruited participants, one at the local AR side and the other at the remote VR side of the implemented MR remote collaboration system. Recruiting participants in pairs is a widely adopted method to observe the user perception on virtual body parts [26, 47, 58, 67]. Since the *environment* was a between-subject factor, each participant was experienced either AR or VR conditions.

As dependent variables, we measured the social presence, presence, trust, preference, and mental effort after each condition had finished. Social presence was the main measurement because it is an important factor indicating how much the mediating system effectively conveys "the feeling of communicating together" to users [7, 21]. In this study, three main social presence questionnaires were utilized based on the Networked Minds Measure of Social Presence [15], Bailenson's Social Presence [5], and Nowak's Social Presence [43]. Harms and Biocca's Social Presence focuses on co-presence, attention, and mutual understanding under the interaction during a collaboration, and Nowak's Social Presence focuses on feelings such as intimacy, or involvement in the interaction with their remote partner, while Bailenson's Social Presence consists of more direct questions related to the perception of images.

In other factors, presence [69], trust [8, 30], and likability [33] were also measured to evaluate how the participants differently perceived the virtual image of their remote partner's hands. For presence, we utilized Witmer and Singer Presence Questionnaire [69], which is a commonly adopted method [28, 56, 57], to evaluate the participant's experience with different virtual models and environmental conditions in terms of naturalness and engagement. The trust and likability measurements adopted for this experiment were derived from previous studies focused on the representation of a virtual avatar [30, 33]. Social presence, trust, likability were rated on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree) and presence was rated on a 7-point semantic differential scale.

In terms of the task performance, a Subjective Mental Effort Questionnaire (SMEQ) [75] was evaluated to observe whether the different hand representations affect the collaboration performance. SMEQ was rated at between 0 to 150 and considered the mental load while performing the task. Furthermore, to gather more objective data, the task completion time was recorded and measured during the tasks as another performance factor. After the participants finished performing all three *hands* conditions, a post-experiment interview was also conducted. During the post-task interview, we asked questions related to social presence, the participant's general feelings, and the overall feedback.

3.3 System Implementation and Setup

We implemented a prototype MR remote collaboration system for our user study. The system was developed using the Unity (ver 2017.4.25f1) game engine. In our study, the AR participants wore a Meta 2 Optical See-Through (OST) HMD (in Figure 1(a)), which has a 60Hz refresh rate, 90-degrees field of view (FoV), and 1280x1440 resolution per eye. The remote VR participants wore an HTC Vive HMD (in Figure 1(h)), which has a 90Hz refresh rate, 110-degrees FoV, and 1080x1200 resolution per eye. A Leap Motion Orion 4.0.0 SDK hand tracker was mounted on both the AR and VR HMDs for the real-time hand tracking of the users. We used two separate PCs connected on a wired LAN. For the AR side, we used a Windows 10 PC with an Intel Core i7-7700K, 32GB of RAM, and an NVIDIA GTX1080Ti. For the VR side, we used a Windows 10 PC with an Intel Core i7-6700K, 16GB of RAM, and an NVIDIA GTX1080. The user's head position and orientation as well as the hand pose data, were synchronized using Unity Networking (UNet).

The tracked local hand pose of each side is encoded and sent the other side, and the received data from the other side are decoded into remote hand pose data. The rigged hand models are controlled with the hand pose data. For the VR side, the hand movements of both the VR and AR users are rendered and animated in VR. For the AR side, only the remote (VR) user's hand movements are augmented because AR users can see their own hands. For realistic hands (Table 1(3)), both female and male models were utilized to reduce the gender impact [57]. The overall color schemes of the low Polygon and skeleton were prepared using an achromatic color as much as possible to avoid any effects caused by a color preference (Table 1(1) and (2)). The color was also selected based on our preliminary study to reduce the color overlapping with the VR background.

For the 3D environmental conditions, a desk and chair were placed in each space, and one side was configured as the AR side and the other was configured as the VR side. We set up the surroundings of the AR participant with minimum components to make them more focused on the task and hands representation as well as avoid potential cognitive load by the complexity of the real background [29, 31]. To control the differences in both conditions, the VR environment was configured to reflect the real environment of the AR. Each space was set up in the same room for seamlessly transmitting their voice, but the spaces were separated by a partition so that the participants could not see each other, which has been a widely adopted implementation in previous remote collaboration studies [26, 42, 72].

3.4 Participants

A total of 48 participants were recruited in pairs through our university's online community recruiting board. There were 22 females and 26 males ranging from 19 to 36 years in age (*M*=24.58, *SD*=3.65). To avoid any biased results derived from the gender difference, we balanced the gender ratio of the recruited pairs as much as possible (10 m-m, 8 f-f, and 6 f-m pairs). Furthermore, we recruited pairs in a close relationship such as friends, family, or colleagues to avoid any unexpected negative feelings that might occur between strangers (7-point self-reported closeness score among pairs, *M*=6.15, *SD*=.94).

The participants were asked their level of familiarity with VR/AR (wearing an HMD), hand-gesture interaction in VR/AR, and 2D/3D remote collaboration systems based on a 7-point rating scale (1 = novice, 7 = expert). The resulting average level of familiarity was 2.85 (*SD* = 1.64) and 2.15 (*SD* = 1.64) for VR/AR and 3D hand-gesture interaction experience, respectively, and 5.33 (*SD* = 1.6) and 1.35 (*SD* = .84) for 2D videoconferencing systems and 3D Social VR/AR systems. The experiment took an hour on average and the participants were compensated with $10; and IRB approval was obtained in advance.

3.5 Tasks and Procedures

To evaluate the difference in the virtual hands representation based on user perception and performance in AR and VR, we conducted a user study using a hand-gesture based collaborative game scenario (see in Figure 2). The main task was a Word Quiz Solving task, which involved an American Sign Language (ASL). It was regarded to be suitable for observing the main purpose of this study because sign language requires sufficient hand movements and various changes in hand motion. In addition, because the hand-based telepresence system is applicable for a future user scenario, such as for use in sign language or remote handicraft education, the user task was also designed to utilize more enriched hand movements based on a simplified version of ASL. Thus, each pair of participants were asked to solve the ASL word quiz sets through a collaboration. In this study, we only adopted 14 (four vowels and ten consonants) among the 26 letters because the other 12 are difficult to detect using a leap motion tracker due to the finger occlusions (Figure 3). Sub-tasks, including a handshake, rock-paper-scissors, and high five, were also added at the beginning and the end of the main task procedure to induce more hand interactions between the two participants.

Task: The ASL Word Quiz Solving task needed two players, one participant giving the quiz, and the other participant guessing the answer. When the team finished solving a one-word question, they switched their roles of quizzer and solver. The quiz consisted of a set of ten different words per condition, so they changed roles five times for each condition (each participant had to guess the answer five times). Three different virtual hands conditions per pair were set in this study, and a total of 30 words (10 words (1 set) x 3 *hands* conditions) were given to each pair of participants. The order of the three quiz sets was randomized based on a Latin Square design. The quiz words were designed to consist of a certain percentage of the alphabet to control between the conditions as much as possible. The study also included adequate ASL learning and a face-to-face practice session (conducted using two sample words) to avoid learning effects from the iterative trials and to fully understand the task and procedure. During the collaborative solving task, one participant was on a local AR side, and the other participant experienced a remote VR side. The role of the AR/VR side collaborator was randomly assigned by the experimenter, but the number of participants per AR/VR condition was balanced (AR = 11 female and 13 male; VR = 11 female and 13 male). The environmental conditions were set differently between participants in the AR and VR, but each participant performed as an equal collaborator at the same level of the role (repeatedly took turns equally between giving the quiz and guessing the answer) because we tried to minimize biased results derived from the repeated measures.

Experimental Procedure: The user study was conducted with a pair of participants, who first filled out a demographic questionnaire asking their age and previous experience with AR/VR, hand-gesture interactions, and a 2D/3D telepresence system. Next, we explained the purpose of the study and the overall experimental task and procedure (see Figure 2). Our main task utilized ASL, so the participants learned and fully practiced it before the experimental trials. The researcher showed an image of 14 sign language letters (Figure 3) to the participants and demonstrated the exact hand pose. After the participants practiced the ASL with their own hands until they became familiar with it, they tried two sample ASL word quizzes while collaborating face-to-face to ensure an adequate understanding of the experimental task. When the face-to-face practicing session ended, we guided each participant based on a random assignment to sit in an appropriate 3D environment (in Figure 1(a) and (h)). Although the two participants sat facing each other to hear the other's voice allowing them to have a conversation together, they were separated by a front partition and could not see each other. The participant sitting at the local AR side wore an AR OST-HMD and viewed the real environment and their own real hands. The VR side participant wore a VR HMD and could only see the virtual scene and virtual hands of the partner and their own hands.

The main task session was composed of three parts: the beginning, ASL quiz, and ending (see in Figure 2). When the main session began, the pair greeted each other with a handshake and fist bump and played the traditional two-handed rock-paper-scissors game to see who would start first. The winner of the beginning session started the ASL quiz by giving a question to the remote partner. When the question word appeared on their HMD, the participant made the ASL hand pose in front of the leap motion attached to the HMD. The remote partner looked at the partner's hand pose and guessed the correct answer. The quiz giver was instructed to use both hands when they gave the quiz; one hand should make an alphabet letter, and the other should show the alphabet letter's ordinal number. After 10 seconds from the time the question appeared on the quizzer's side, three possible answers of multiple-choice answers also appeared on the solver's side to help them get the answer easily (because the task was not focused on memorizing the exact ASL alphabet letters). The participants were also instructed to apply a two-handed high-five motion when they got the correct answer. They acted as the quiz giver and solver five times equally because they switched their role for each word. When they finished answering all ten words under the single virtual hands type conditions, they played a one-handed rock-paper-scissors during the ending session. Furthermore, they gave a two-handed high-five if they tied, and the main task session was ended when they tied three times.

Once they finished the main interactive task, they took off the AR/VR HMD to avoid dizziness or headache and answered the prepared questionnaires. We set three conditions of the virtual hand representation in this study, and thus they repeated the above main ASL session three times. The duration for each task and questionnaire was 10 to 15 minutes per condition (the blue box in Figure 2). The order of the virtual hands conditions was counter-balanced based on a Latin Square method to avoid ordering effects in a within-subject design. The post-experiment interview was conducted when the participants finished the tasks with all three virtual hands conditions. The researcher conducted the interview separately in a different space on each side to prevent the participants from being affected by the responses of their partner.

4 RESULTS

We present the results from the quantitative and qualitative data in Figures 4 to 6. For the quantitative data, we obtained answers regarding social presence, presence, trust, likability, and SMEQ for every condition. To obtain objective data from the quantitative results, the task completion time was measured during the user task, excluding the beginning and ending sessions (Figure 2). We used the same questionnaires for all 48 participants. For all subjective measurements and SMEQ, 24 (pairs) x 3 (hand types) x 2 (environment types) = 144 data points were collected. A total of 24 (pairs) x 3 (hand types) x 10 (quiz trials per condition) = 720 data points for the task completion time were recorded. For both a subjective and objective data analysis, we excluded outliers such as extreme values, contaminated trials, and system errors. The task completion time was collected through 10 experimental trials, but we eliminated the first and second trial data of all pairs reducing any potential learning effect of the repeated trials. The participant's general feedback about the overall virtual hand representation was collected after all experimental tasks had ended, which contributed to the data points of the 48 participants.

4.1 Subjective Measures

We used a two-way repeated measures ANOVA procedure with the Aligned Rank Transform(ART) for non-parametric factorial analysis (α = 0.05), as proposed by Wobbrock et al. [70]. All pairwise comparisons were Bonferroni corrected. The internal consistency among Likert items examined by Composite Reliability based on Confirmatory Factor Analysis(CFA), which accepts the calculated value of more than 0.6 as an acceptable reliability level.

Social Presence and Presence

Harms and Biocca's Social Presence [15] questionnaires included the following three sub-scales: Co-presence(*CP*), Attentional Allocation(*AA*), and Perceived Message Understanding(*PMU*). We also used the aggregated social presence(*HSP*) score which combines all three sub-scales to compare the results. The internal consistency of *HSP* and the three sub-scales all showed good reliability ranging from 0.811 to 0.924.

We found a significant effect of *hands* ($F(2,95) = 3.291$, $p =$ .041), and *environment* ($F(1,95) = 3.996$, $p = .048$) on the aggregated Harms and Biocca's Social Presence(*HSP*), however, the post-hoc analysis with pairwise comparison revealed no significant difference between any pair of *hands*: *Skeleton-LowPoly* ($p = .151$), *LowPoly-Realistic* ($p = 1.0$), *Skeleton-Realistic* ($p = .054$). There was also no effect of *hands*×*environment* interaction ($F(2,95) =$ .474, $p = .624$). We analyzed each sub-scale: *CP* showed a significant effect of *hands* ($F(2,105) = 5.572$, $p = .005$), and the post-hoc pairwise comparisons only indicated a significant difference between *Skeleton-Realistic* ($p = .004$). *PMU* showed a significant effect of *environment* ($F(1,105) = 5.885$, $p = .017$). The other factors, interaction effect, and the rest of *hands* pairs regarding *CP* and *PMU* were not significantly different. Furthermore, *AA* revealed no significant main effects on any factors nor interaction effect.

Bailenson's Social Presence [5] questionnaires consisted of five questions with no sub-scales. In this study, we only asked the three questions among five because we only focused on the representation of the hands; the two excluded items asked for a whole image of the person. Bailenson's Social Presence (*BSP*) was also aggregated from the scores of the three questions. The internal consistency of the *BSP* indicated good reliability among Likert items (0.856). We found a significant main effect of *hands* on *BSP* ($F(2,115) =$ 5.487, $p = .005$). Pairwise comparisons of *hands* revealed significant differences between *Skeleton* and *Realistic* ($p = .004$). The other *hands* pairs showed no significant differences on *BSP* ($p > 0.05$). *Environment* or *hands*×*environment* interaction had no significant effect on *BSP* ($BSP_{env.}$: $F(1,115) = .599$, $p = .441$; $BSP_{hands\times env.}$: $F(2,115) = .600$, $p = .550$).

Among *Nowak and Biocca's Social Presence* [43] questionnaire sub-scales, we only adopted four questions of the perceived other's and self-reported copresence to avoid redundant questions with *HSP* as well as focusing more on intimacy, engagement and relationship during the interaction between the participants. Composite Reliability passed the internal consistency of collected scores of Nowak's Social Presence (*NSP*) (0.777), and the aggregated value of *NSP* was also used to evaluate each factors. A significant main effect on *NSP* was found in *hands* ($F(2,110) = 3.662$, $p = .029$). The post-hoc analysis with pairwise comparison showed significant differences between *Skeleton* and *Realistic* ($p = .035$), while there were no significant differences neither *Skeleton-LowPoly* nor *LowPoly-Realistic* pairs (all $p > .134$). Moreover, we found no significant effect of *environment* ($F(1,110) = 2.074$, $p = .153$) or *hands*×*environment* interaction ($F(2,110) = .031$, $p = .970$).

Presence Questionnaire suggested by Witmer and Singer [69] measured effectiveness and experience of the virtual environment. In this study, we extracted the three questions under naturalness and involvement sub-scales, which were the most related items with our experimental setup. The internal consistency of answers was also in the acceptance level (0.637). The aggregated Presence Questionnaire (*PQ*) showed a significant effect of *hands* ($F(1,110) =$ 4.681, $p = .011$), pairwise comparisons revealed that *Skeleton-Realistic* ($p = .009$) was significantly different. Other pairwise comparisons of *hands* factor on *PQ* were not significantly different (all $p > .151$). The *environment* ($F(1,110) = 2.001$, $p = .160$) and *hands*×*environment* had no significant effect ($F(2,110) = 1.459$, $p = .237$).

Trust, Likability and Mental Effort

The questionnaire of trust consisted of the three items, and scores obtained from the participants were also aggregated to analyze the results. The internal consistency of trust (*TRST*) with Composite Reliability was also in an adequately accepted range (0.848). We found a significant main effect of *hands* on *TRST* score ($F(2,100) =$ 3.864, $p = .024$). Pairwise comparisons of *hands* revealed a significant difference in *Skeleton* and *Realistic* hands pair ($p = .022$). Other pairwise comparisons of *hands* on *TRST* were not significantly different (all $p > .219$). The other factor *environment* ($F(1,100) =$ .006, $p = .938$) and *hands*×*environment* had no significant effects ($F(2,100) = .625$, $p = .537$).

We used a single question (*"I liked the appearance of the remote partner's virtual hands."*) to compare the likability (*LIKE*), and it was found that there was a significant main effects of *hands* ($F(2,115) = 4.863$, $p = .009$). The post-hoc analysis with pairwise comparison showed a significant difference between *Skeleton* and *Realistic* ($p = .008$), and rest of the pairs had no significant differences (all $p > .163$). The other factor as well as interaction showed no significant effects on *LIKE* (*environment*: $F(1,115) = .206$, $p =$ .651; *hands*×*environment*: $F(2,115) = .079$, $p = .924$).

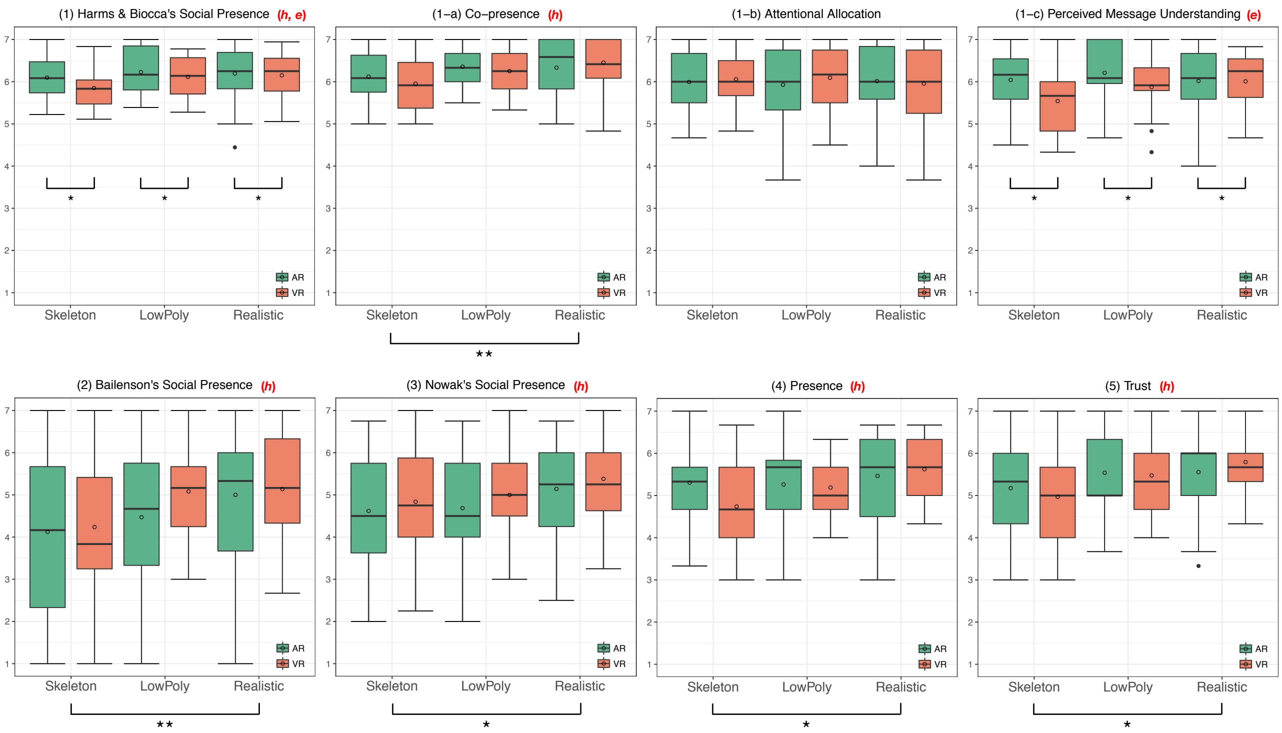

Figure 4: (1)-(5) Results of subjective measures on the user perception (h and e: a significant effect of hands and environment, respectively): (1) Aggregated Harms & Biocca's Social Presence (*HSP*), (1-a) Co-presence (*CP*), (1-b) Attentional Allocation (*AA*), (1-c) Perceived Message Understanding (*PMU*), (2) Bailenson's Social Presence (*BSP*), (3) Nowak's Social Presence (*NSP*), (4) Witmer & Singer Presence (*PQ*), (5) Trust (*TRST*).

Subjective Mental Effort Questionnaires (*SMEQ*) was a single question, asking a mental load while performing the task. A significant main effect on *SMEQ* was found in *environment* ($F(1,90) = 7.407$, $p = .008$), between *AR* and *VR* conditions. There were no significant main effects on *hands* ($F(2,90) = 2.640$, $p = .077$) nor *hands*×*environment* interaction ($F(2,90) = 1.168$, $p = .316$).

4.2 Objective Measure

Since the prototype system was an MR-based telepresence as well as the pair of participants in each AR and VR side performed the main quiz task together, the virtual hands type (*hands*) was only used as a factor for collecting the task completion time. In addition, the purpose of this study was to observe whether there was a significant difference in AR and VR, rather than comparing the exact value scored in each environmental condition. Hence, the completion time recorded under the three virtual hands conditions (skeleton (*Skeleton*), low polygon (*LowPoly*), realistic (*Realistic*)) was analyzed.

Task Completion Time

The ten data points per each condition had collected because the quiz task consisted of a set of ten words, so it was aggregated to analyze statistical differences between hands conditions. However, the eight data points out of ten were used to analyze owing to the training effects. We firstly checked the normality with a Shapiro-Wilk test, and it was found that our data was not following normal distribution (*Skeleton*: $W = .953$, $p < .001$; *Low Polygon*: $W = .920$, $p < .001$; *Realistic*: $W = .966$, $p = .001$). Therefore, we used a Friedman Test for non-parametric analysis ($\alpha = 0.05$) followed by a post-hoc Wilcoxon Signed Rank Test, and found that there was a significant difference in the task completion time depending on which type of virtual hands was used ($\chi^2(2) = 8.455$, $p = .015$). The post-hoc analysis with Bonferroni correction found that the pair between *LowPoly* and *Realistic* showed a significant difference ($p = .016$). The rest *hands* pairs had no significant differences (all $p > .133$).

4.3 Post-experiment Interview

During the post-experiment interview, we asked additional questions utilized in a prior study [72] to each participant. Specifically, there were two questions regarding social presence, two questions on the easiness of understanding the message, and two questions on preference to better understand their subjective feeling (see Figure 6). We categorized their answers according to the three virtual hands model, and these opinions revealed in the same way in both AR and VR environmental conditions. Furthermore, we also summarized other observations and feedback obtained from the participants.

Realistic Hands: As shown in Q1 and Q5, most of the participants selected realistic type in terms of social presence (the feeling of doing something together) and preference (most preferred) (in Figure 6). The participants who gave a positive answer to realistic hands responded with statements such as "like a real human hand," "felt familiar," and "Realism." For instance, *AR_P15 stated "It was similar to a human hand, so I felt like I was actually interacting with my experiment partner,"* and *VR_P22 claimed "It was very familiar human hands, I felt the other person's presence the most,"* were common responses. By contrast, there were some subjects who answered a negative opinion (Q2, Q6): "It looks different from my partner's real hand," or "it sometimes posed or moved unrealistically different from the gesture of the real person's hand made." *VR_P2* mentioned – *"It was quite disgusting when the joints were sometimes bent in a strange way, and I did not like it because it was different from my partner's real hands that I knew."*

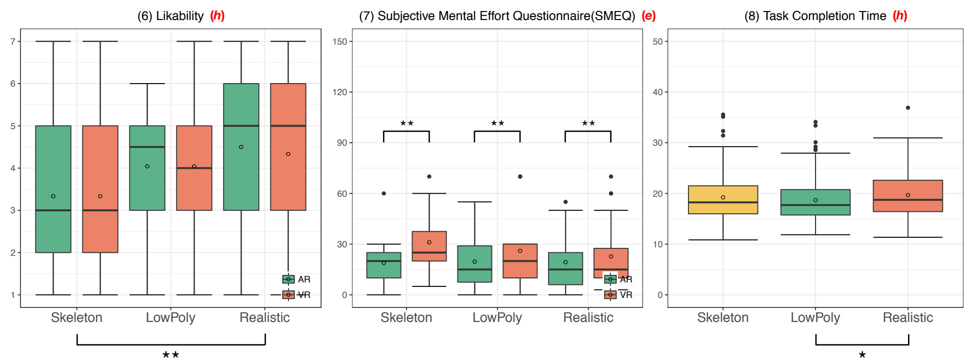

Figure 5: (6)-(8) Results of both subjective and objective measures (h and e: a significant effect of hands and environment, respectively): (6) Likability (*LIKE*), (7) Subjective Mental Effort (*SMEQ*), and (8) Task Completion Time (in seconds)

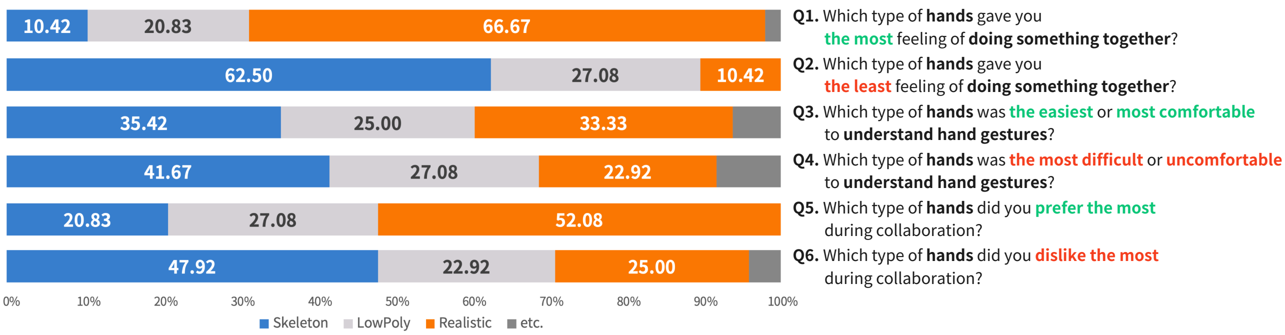

Figure 6: Percentage of answers regarding the three virtual hands models collected from the post-experiment interview questions

Low Polygon Hands: In the case of a low polygon, the positive/negative response rate on each question appeared similarly because the two main reasons were perceived with ambivalence: "Neutral" and "Symbolic." The participants stated that they felt neutral, which meant the model was not too realistic nor unrealistic, and it was comfortable because of a moderate level of realism. However, some participants perceived this neutral feeling as rather vague and robotic: *AR_P18 – "It seemed moderately real; in other words, it also seemed to be moderately fake,"* or *VR_P1 – "There was an indescribable discomfort that was not as concise as skeleton or human-like hands."* In addition, for the meaning of "Symbolic", it was mentioned that the sign made by the hand gesture was clearly transmitted due to the minimal and clean skin of the low polygon (*VR_P21 – "The shape was the cleanest, and it was a solid color, not a skin color, so I could focus more on the shape of the hand itself."*), but each division of the joint was unclear owing to the absence of an expression such as a finger joint (e.g., *VR_P16 – "It was unclear because hand knuckles were the most indistinguishable."*).

Skeleton Hands: It was revealed in Q2 and Q6 (as shown in Figure 6) that most of the participants negatively perceived the skeleton in terms of social presence or preference among the three models. The skeleton shape and unrealistic appearance seemed most affected by the perception because the answers repeatedly mentioned: "not human-like hands," "fear," or "discrepancy." For example, *AR_P21 stated "Even though I noticed the letter the gesture indicated, I did not feel like I was working with a real human because the shape was too separated from the hands of a real person."*. *VR_P6* also said *"It was difficult to catch the movements of the hands, and I was less focused because I felt emotional disparateness."* At the same time, however, some participants responded positively because it was "intuitively recognizable," indicating that the simplified skeletal structure attracted their attention: *VR_P7 – "I felt like the skeleton could effectively deliver key information such as the hand movements and poses with the most minimal element."*

Other Perceived Factors on Virtual Hands Appearance: During the interview, the factors related to the appearance such as texture, volume, scale, and color of the hands were mentioned repeatedly. For instance, the participants specified the feelings of personalization (e.g., skin tone, hair, or wrinkles) that the texture of the realistic hand can express, or the model color of the low polygon or skeleton: *VR_P7 – "The realistic texture such as the hair and protruding bone joints, was unfavorable and rather hindered the hand recognition."* In addition, there were more responses to the volume or hand/finger scale, an opinion that the volume was recognized as being too thick or thin, and comments regarding a three-dimensional effect. For example, *AR_P22 – "The low polygon had the thickest finger, and was therefore the most clearly communicated, and the design was also the most comfortable. However, the realistic type had a thin finger, and it was difficult to tell whether it was a full grip or a slight grip."*

Observation on AR and VR participants: In addition, the observations by the researchers showed that there was a specific behavioral difference created by the fundamental system characteristics between participants in each 3D environment condition when conducting the complete experimental tasks. Since the VR participant's virtual self-hands were shown in front of them, they behaved in a self-correcting way when the hands were not properly displayed due to the leap motion's limited tracking range or a temporal delay. However, there was some difficulty in correcting the wrong expression for the AR participants because they saw their real hands rather than virtual self-hands. Thus, it was frequently observed that the VR participants needed to directly deliver extra feedback to their partner to correct the hand model displayed or had to wait for the proper virtual hand gesture to appear.

5 Discussion

We verify the six hypotheses based on statistical analysis and further discuss on designing a future MR remote collaboration, particularly for a hand-gesture based collaborative scenario.

In the case of **H1**, it was verified that the virtual hands type (*hands*) will affect the user experience during an MR remote collaboration because of significant effects of the *hands* type, as supported by the social presence (*HSP*, *BSP*, and *NSP*), presence (*PQ*), and trust (*TRST*). The second hypothesis (**H2**) that the realistic hands type will have the highest social presence, presence, and trust was partially accepted because the pairwise comparison revealed that the *hands* pair between the realistic and skeleton had a significant difference, and the realistic showed higher *BSP*, *NSP*, *PQ*, and *TRST* than the skeleton. However, although *HSP* revealed a significant effect on *hands*, there were no significant differences found in the post-hoc analysis among the *hands* pairs. In addition, it was not verified that the realistic had a higher value than the low polygon because the *hands* pair (*LowPoly-Realistic*) did not show any significant difference. We inferred these results for the following reasons: First, in the case of *HSP*, it could be that the sub-scale *AA* and *PMU* have an effect on the aggregated value. Although the degree of realism differed under the three conditions, they had common information (e.g., skeletal structure), which is necessary to distinguish gestures; the voice was also transmitted. Second, although the feeling of realism was perceived as the highest in the realistic, a certain number of participants answered positively regarding the low polygon because of its moderate impression, which might diminish the significant difference.

The experimental results rejected our third hypothesis (**H3**). We found no significant effects of *hands* on *SMEQ*, and thus it was not verified that the realistic would induce the least amount of mental effort. Furthermore, **H3** also assumed that the realistic would achieve the highest task performance; however, the low polygon hands scored better than the realistic in terms of the task completion time (Figure 5(8)). Thus, **H3** was not supported. As already mentioned in a previous study [61, 72], we speculated that it could be possible to have no significant difference in mental effort if the virtual hands could fulfill the role of a communication cue (e.g., proper hands structure or fingers). Furthermore, as the participant's feedback indicated, the simplified hand expression of the low polygon may convey the sign more clearly so that the task performance was better with this model. In addition, when the realistic posed an unrealistic shape or when the posed shape was different from what the partner gestured, it could take away the participant's concentration or disturb their ability to catch the message more often. This result conflicts with that by Knierim et al. [28], who argued that the lowest task load occurs when applying a realistic hands. However, Grubert et al. [14] confirmed that a simplistic virtual hands representation can be more advantageous in reducing the time for certain tasks. We assessed that a remote embodiment in comparison with previous studies evaluating a self-embodiment [28] also might arrive at a different result.

We hypothesized that a 3D environment (*environment*) will affect the user experience in a MR remote collaboration (**H4**). From the statistical analysis, we partially verified the **H4** because only *HSP* and its sub-scale *PMU* had significant effects on the *environment*. The results on both *HSP* and *PMU* showed that the three virtual hands under the *AR* conditions were higher than those for the *VR* conditions, and thus, the fifth hypothesis (**H5**) was also partially accepted. Jo et al. [24] revealed that the co-presence was higher in a real background, and trust showed no difference between background types. We found that our results did not align with theirs except for the *TRST* measurement; however, Jo et al. did not compare both sides of the MR system at the same time during a remote communication. In addition, they evaluated the image of the remote expert represented by a full-body avatar between pairs of participants who did not know each other. As a result, we concluded that no differences were shown in the other values of social presence, presence, and trust because the participant's concentration was mostly on the communication itself rather than the effect of the 3D environmental difference when collaborating at the same time. Furthermore, a simple office environment was constructed to minimize potential noise from the messy background in our settings, and the participants would have been able to focus more on the hand movements, tasks, and communication; thus the possible effect caused by the different 3D environment might be reduced.

Finally, **H6** supposed that the participants of VR will show a lower mental effort than the participants of AR. We found a significant effect of the *environment* on mental effort regardless of the virtual hands type (*hands*), but the results showed the opposite, i.e., the mental effort of the VR participant was higher than that of the AR participant. Thus, we could not confirm our sixth hypothesis (**H6**). This validation seemed to be related to our observation that the different amount of load demanded particularly more from the VR participants. Also, even though we expected that the cognitive load of the AR condition would be higher, the sickness of the virtual experience might have had more influence than the overload from the perception switching in the AR due to the simplified experimental setup.

5.1 Implications for Hand-based 3D Remote Collaboration

The experiment found that the virtual hands type and 3D environment of the remote partner had an effect on the subjective factors, as well as the performance during a remote MR collaboration. We also discovered meaningful feedback and qualitative observations gathered through the overall experimental process. From our findings, we summarize the design implications for deciding an adequate hands model and environmental setup to achieve an effective hand-based 3D collaboration.

Remote Virtual Hands Appearance in Hand-based 3D Telepresence: When expressing the partner's remote hands using a virtual avatar, the amount of realism of the model used should be considered. Based on the validation of **H1**, it was found that the virtual hands type affected the user's subjective experience, and the realistic hands type had a particularly higher value than the skeleton based on the following indicators: social presence(*CP of HSP, BSP, and NSP*), presence(*PQ*), trust(*TRST*), and likability(*LIKE*). We also revealed that there were no differences between the *Realistic-Low Polygon* hands in social presence or other subjective indicators in a pairwise comparison, and the task completion time showed a lower value in the low polygon than the realistic.

It was found that the realistic and the low polygon models seemed to be able to convey different perceptions and experiences to users, as mentioned in the qualitative feedback. For example, all the participants, who chose the realistic hands, left common comments such as it achieved a "human-like appearance," which could directly lead to feelings of familiarity, intimacy, and realism regarding the remote virtual hands. By contrast, the majority opinion of the low polygon was "neutral," so there were no particularly skewed values among the dependent factors toward this model. In addition, because the surface of the low polygon was smooth and simple, it was regarded that the message could be delivered more clearly with a more symbolic expression.

Consequently, we expected both the low polygon and the realistic approaches could be utilized for the virtual hands of a remote partner; however, each model would be selected differently depending on the main focus of the collaboration context, i.e., message delivery or resemblance to the human partner, or the system setup. When the main purpose is communication "performance" and message "delivery" rather than the image of the people, a low polygon approach could be selected. In addition, if there is a limited variation in the choice of advanced realistic hands options due to the system configuration, such an approach is also expected to be utilized. In such cases, it may be helpful to adopt a low polygon style for a moderate design, particularly in terms of the image aspects such as the volume, scale, and color. By contrast, if the communication 'partner' is emphasized and the task needs to fulfill the user's perception related to the human resemblance, relationship, and trust between collaborators, a realistic approach would be a suggested option. Moreover, it would be helpful to consider a more personalized model since the importance of consistency with their partner's real hands was repeatedly mentioned.

In addition, the skeleton hands showed the lowest results in social presence and other indicators, and thus we would not suggest its use. We found that doing so could cause some information loss or confusion due to a self-occluded finger expression: it could also induce negative feelings as reported by the participants. As a result, a skeleton approach may be considered when it is not necessary to consider the subjective emotion of the user during collaboration, or if a simple task is conducted by conveying minimal information. However, although it did not show much difference from the other models in terms of message understanding owing to its minimalist design, it would be better to avoid its implementation in MR remote collaboration.

Collaboration between Different 3D Environments: In our study, the difference between AR and VR in overall user evaluation indicators was not statistically significant, excluding only a few measurements. Thus, a successful utilization of an asymmetrically connected MR remote collaboration is expected, especially for hand-based tasks. According to the validation of **H4**, it was interestingly found that the environmental difference affected the communication-related part, i.e., perceived message understanding (*PMU*) and mental effort, of the user's experience, and the VR participants felt more difficulty in this regard than the AR participants.

When interpreting the user responses comprehensively, one possible reason could be that more information was provided to the VR participants than to the AR participants. For instance, the aforementioned self-correction, which was unconsciously conducted by the VR participants, occurred when their hands went out of the tracking area or made an incorrect or unintended expression. On the contrary, the AR participants could not have any information regarding what their hands were currently representing because they saw their real hands, so the task might be easier for them if disregarding a self-correction. Therefore, 1) when correcting the hand gestures by themselves, and 2) when giving feedback to the AR participants to make them fix their posture, additional efforts of the VR participants might be demanded.

In conclusion, we would like to recommend that when configuring an MR-based environmental setup, the limitations and disadvantages discussed earlier derived from the fundamental differences between the AR and VR should be carefully improved to reduce any difficulties and support an effective and seamless interaction.

5.2 Limitations

Although the present study covers interesting points, but there are also a number of limitations that should be considered in a future study. First, we tracked the hand gestures and movements by using a leap motion. The limited tracking area, as well as technical limitations on processing hand occlusions or jitters, may have affected the results, or a feeling of discomfort could have been captured from an unrealistic posture of the realistic hands. Although we excluded outliers as much as possible in our analysis, extra sensors or cameras should be utilized for a more robust hand tracking.

Second, based on information gathered during the interview, other appearance factors of the remote partner's virtual hands might draw more insight. In our study, however, the criterion used for selecting the hands model was more focused on the level of visual appearance. Furthermore, the participants who were on the VR side only experienced realistic type of self-hands. Thus, the personalized factor or an ideal pair of both remote- and self-hands type would also be investigated as an independent factors in further research.

Finally, we only conducted a task involving ASL to focus more on the complex changes in hand pose and motion. However, a further investigation with more varying user tasks reflecting our implications is required, such as collaboration while interacting with a real/virtual object, or using a remote assist scenario. In addition, we focused on the pairs in a close relationship, and different results derived through collaboration between strangers would be beneficial.

6 Conclusion and Future Work

In this study, we investigated the effects of the virtual hands type of a remote collaborator on the overall user experience in both AR and VR remote collaboration conditions. To verify our hypotheses and investigate the research questions, we conducted a user study and analyzed several user measurements including social presence, presence, trust, likability, and performance. The results showed that the realistic and skeleton hands types were significantly different in most factors excluding one social presence factor (*HSP*) and mental effort (*SMEQ*), and no significant differences were found between other pairs with a low polygon. Furthermore, the 3D environment did not show any significant differences in the aforementioned factors except for message understanding under social presence (*PMU of HSP*) and mental effort (*SMEQ*) among the virtual hands types.

We concluded that among the three types of virtual hands, which were distinguished by their realism (different levels of visual appearance), both the realistic and low polygon could be considered for implementing a hands-based 3D remote collaboration. However, if the collaboration needs to provide intimacy, trust, and the feeling of realism to the users, then the realistic type might be more recommended. By contrast, if the system needs a moderate level of social presence and focuses more on clearer communication, then the low polygon should be considered. Moreover, when implementing a hands-based collaboration with an asymmetric MR telepresence system, the effects from the differences between the AR and VR would be diluted in terms of the user's subjective feelings. However, it would be beneficial for a future system if it could effectively support the interactions between users in a heterogeneous environment and reduce the potential difficulties in communication.

In future research, we would like to conduct further user studies with a more improved system, different interactive tasks, and collaborators roles. In a more robust system, different insights from the user-side or other accurate objective data might be discovered, and thus a system enhancement can be achieved with additional implementation. In this study, we set the collaboration type with equal participation of the users and only conducted a single task; however, it will be necessary to explore more various collaborative scenarios such as a trainer and trainee. Furthermore, we would like to investigate other determining factors related to personalization, such as skin tone, age, volume, or size. Finally, we would also like to define the best pair of remote hands between participants; we plan to examine the proper match between virtual remote hands and the user's own hands under the VR condition.

Acknowledgments

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2019-0-01270, WISE AR UI/UX Platform Development for Smartglasses) and Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT (NRF-2017M3C4A7066316).

REFERENCES

[1] I. Abibouraguimane, K. Hagihara, K. Higuchi, Y. Itoh, Y. Sato, T. Hayashida, and M. Sugimoto. Cosummary: adaptive fast-forwarding for surgical videos by detecting collaborative scenes using hand regions and gaze positions. In *IUI*, pp. 580–590, 2019.

[2] L. Alem and J. Li. A study of gestures in a video-mediated collaborative assembly task. *Advances in Human-Computer Interaction*, 2011:1, 2011.

[3] J. Amores, X. Benavides, and P. Maes. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1343–1348. ACM, 2015.

[4] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*, pp. 3–10. IEEE, 2016.

[5] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal Distances in Virtual Environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003. doi: 10.1177/0146167203253270

[6] S. Beck, A. Kunert, A. Kulik, and B. Froehlich. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):616–625, April 2013. doi: 10.1109/TVCG.2013.33

[7] F. Biocca, C. Harms, and J. K. Burgoon. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments*, 12(5):456–480, oct 2003. doi: 10.1162/105474603322761270

[8] K.-T. Chun and J. B. Campbell. Dimensionality of the rotter interpersonal trust scale. *Psychological Reports*, 35(3):1059–1070, 1974.

[9] DAQRI. DAQRI Worksense, Retrieved July 26, 2020 from https://daqri.com/worksense/, 2019.

[10] Facebook. Facebook Horizon, Retrieved July 26, 2020 from https://www.oculus.com/facebook-horizon/, 2020.

[11] R. Fribourg, F. Argelaguet, L. Hoyet, and A. Lécuyer. Studying the Sense of Embodiment in VR Shared Experiences. In *VR 2018 - 25th IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 1–8. IEEE, Reutlingen, Germany, Mar. 2018.

[12] Google. Google Glass Enterprise, Retrieved July 26, 2020 from https://www.google.com/glass/start/, 2019.

[13] S. W. Greenwald, Z. Wang, M. Funk, and P. Maes. Investigating social presence and communication with embodied avatars in room-scale virtual reality. In *International Conference on Immersive Learning*, pp. 75–90. Springer, 2017.

[14] J. Grubert, L. Witzani, E. Ofek, M. Pahud, M. Kranz, and P. O. Kristensson. Effects of hand representations for typing in virtual reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 151–158. IEEE, 2018.

[15] C. Harms and F. Biocca. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annual International Workshop: Presence 2004*, pp. 246–251, 2004.

[16] P. Heidicker, E. Langbehn, and F. Steinicke. Influence of avatar appearance on presence in social vr. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 233–234, March 2017. doi: 10.1109/3DUI.2017.7893357

[17] L. Hoyet, F. Argelaguet, C. Nicole, and A. Lécuyer. "Wow! i have six fingers!": Would you accept structural changes of your hand in vr? *Frontiers in Robotics and AI*, 3:27, 2016.

[18] W. Huang and L. Alem. Supporting hand gestures in mobile remote collaboration: a usability evaluation. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, pp. 211–216. British Computer Society, 2011.

[19] W. Huang, L. Alem, F. Tecchia, and H. B.-L. Duh. Augmented 3d hands: a gesture-based mixed reality system for distributed collaboration. *Journal on Multimodal User Interfaces*, 12(2):77–89, 2018.

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

[21] W. A. IJsselsteijn, H. de Ridder, J. Freeman, and S. E. Avons. Presence: concept, determinants, and measurement. In *Human Vision and Electronic Imaging V*, vol. 3959, pp. 520–530. International Society for Optics and Photonics, 2000.

[22] D. Iwai, R. Matsukage, S. Aoyama, T. Kikukawa, and K. Sato. Geometrically consistent projection-based tabletop sharing for remote collaboration. *IEEE Access*, 6:6293–6302, 2017.

[23] D. Jo, K.-H. Kim, and G. J. Kim. Effects of avatar and background representation forms to co-presence in mixed reality (mr) tele-conference systems. In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, p. 12. ACM, 2016.

[24] D. Jo, K.-h. Kim, and G. J. Kim. Effects of Avatar and Background Types on Users' Co-presence and Trust for Mixed Reality-Based Teleconference Systems. In *Proceedings the 30th Conference on Computer Animation and Social Agents*, pp. 27–36, 2017.

[25] S. Jung, P. J. Wisniewski, and C. E. Hughes. In limbo: The effect of gradual visual transition between real and virtual on virtual body ownership illusion and presence. In *Proceedings of the 25th IEEE Conference on Virtual Reality and 3D User Interfaces, IEEE VR*, 2018.

[26] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghurst. Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 173. ACM, 2019.

[27] D. Kirk and D. Stanton Fraser. Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1191–1200. ACM, 2006.

[28] P. Knierim, V. Schwind, A. M. Feit, F. Nieuwenhuizen, and N. Henze. Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 345. ACM, 2018.

[29] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12. IEEE, 2010.

[30] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, pp. 39:1–39:10. ACM, New York, NY, USA, 2017. doi: 10.1145/3139131.3139156

[31] C. Lee, G. A. Rincon, G. Meyer, T. Höllerer, and D. A. Bowman. The effects of visual realism on search tasks in mixed reality simulation. *IEEE transactions on visualization and computer graphics*, 19(4):547–556, 2013.

[32] L. Lin and S. Jörg. Need a hand?: how appearance affects the virtual hand illusion. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 69–76. ACM, 2016.

[33] L. Lin, A. Normovle, A. Adkins, Y. Sun, A. Robb, Y. Ye, M. Di Luca, and S. Jörg. The effect of hand size and interaction modality on the virtual hand illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 510–518. IEEE, 2019.

[34] Linden Research. Sansar, Retrieved July 26, 2020 from https://www.sansar.com/, 2019.

[35] J.-L. Lugrin, M. Ertl, P. Krop, R. Klüpfel, S. Stierstorfer, B. Weisz, M. Rück, J. Schmitt, N. Schmidt, and M. E. Latoschik. Any "body" there? avatar visibility effects in a virtual reality game. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 17–24. IEEE, 2018.

[36] J.-L. Lugrin, J. Latt, and M. E. Latoschik. Anthropomorphism and illusion of virtual body ownership. In *ICAT-EGVE*, pp. 1–8, 2015.

[37] J.-L. Lugrin, J. Latt, and M. E. Latoschik. Avatar anthropomorphism and illusion of body ownership in vr. In *Virtual Reality (VR), 2015 IEEE*, pp. 229–230. IEEE, 2015.

[38] Magic Leap. Social, Retrieved July 26, 2020 from https://www.magicleap.com/experiences/social/, 2018.

[39] Microsoft. AltspaceVR, Retrieved July 26, 2020 from https://altvr.com/, 2015.

[40] Nasa JPL. OnSight, Retrieved July 26, 2020 from https://opslab.jpl.nasa.gov/, 2015.

[41] S.-T. Noh, H.-S. Yeo, and W. Woo. An HMD-based Mixed Reality System for Avatar-Mediated Remote Collaboration with Bare-hand

Interaction. In M. Imura, P. Figueroa, and B. Mohler, eds., *ICAT-EGVE 2015 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2015. doi: 10.2312/egve.20151311

[42] M. Norman, G. A. Lee, R. T. Smith, and M. Billingurst. The impact of remote user's role in a mixed reality mixed presence system. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–9, 2019.

[43] K. L. Nowak and F. Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5):481–494, 2003.

[44] N. Ogawa, T. Narumi, and M. Hirose. Virtual hand realism affects object size perception in body-based scaling. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 519–528. IEEE, 2019.

[45] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. A. Chou, S. Mennicken, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pp. 741–754. ACM, New York, NY, USA, 2016. doi: 10.1145/2984511.2984517

[46] Y. Pan and A. Steed. The impact of self-avatars on trust and collaboration in shared virtual environments. *PloS one*, 12(12):e0189078, 2017.

[47] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pp. 1716–1725. ACM, New York, NY, USA, 2016. doi: 10.1145/2818048.2819965

[48] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G. Lee, and M. Billinghurst. Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, p. 16. ACM, 2017.

[49] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. CoVAR: Mixed-Platform Remote Collaborative Augmented and Virtual Realities System with Shared Collaboration Cues. *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2017*, pp. 218–219, 2017. doi: 10.1109/ISMAR-Adjunct.2017.72

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

[51] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 46:1–46:13. ACM, New York, NY, USA, 2018. doi: 10.1145/3173574.3173620

[52] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–17, 2019.

[53] N. Ranieri, J.-C. Bazin, T. Martin, P.-Y. Laffont, T. Popa, M. Gross, et al. An immersive bidirectional system for life-size 3d communication. In *Proceedings of the 29th International Conference on Computer Animation and Social Agents*, pp. 89–96. ACM, 2016.

[54] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 215–222. IEEE, 2018.

[55] D. Roth, K. Waldow, M. E. Latoschik, A. Fuhrmann, and G. Bente. Socially immersive avatar-based communication. In *Virtual Reality (VR), 2017 IEEE*, pp. 259–260. IEEE, 2017.

[56] V. Schwind, P. Knierim, L. Chuang, and N. Henze. " where's pinky?" the effects of a reduced number of fingers in virtual reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pp. 507–515, 2017.

[57] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. These are not my hands!: Effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1577–1582. ACM, 2017.

[58] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 289. ACM, 2018.

[59] Spatial Systems. Spatial, Retrieved July 26, 2020 from https://spatial.io/, 2018.

[60] M. Sra, A. Mottelson, and P. Maes. Your place and mine: Designing a shared vr experience for remotely located users. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*, pp. 85–97. ACM, 2018.

[61] A. Tang, F. Biocca, and L. Lim. Comparing differences in presence during social interaction in augmented reality versus virtual reality environments: An exploratory study. *Proceedings of PRESENCE*, pp. 204–208, 2004.

[62] F. Tecchia, L. Alem, and W. Huang. 3d helping hands: a gesture based mr system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 323–328. ACM, 2012.

[63] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–14, 2019.

[64] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pp. 406–410. ACM, 2018.

[65] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. Investigating the use of different visual cues to improve social presence within a 360 mixed reality remote collaboration. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 1–9, 2019.

[66] S. Thanyadit, P. Punpongsanon, and T.-C. Pong. Observar: Visualization system for observing virtual reality users using augmented reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 258–268. IEEE, 2019.

[67] H. Wauck, G. Lucas, A. Shapiro, A. Feng, J. Boberg, and J. Gratch. Analyzing the effect of avatar self-similarity on men and women in a search and rescue game. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 485. ACM, 2018.

[68] C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. Social presence and cooperation in large-scale multi-user virtual reality-the relevance of social interdependence for location-based environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 207–214. IEEE, 2018.

[69] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[70] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 143–146. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963

[71] E. Wood, J. Taylor, J. Fogarty, A. Fitzgibbon, and J. Shotton. Shadowhands: High-fidelity remote hand gesture visualization using a hand tracker. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, pp. 77–84. ACM, 2016.

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

[73] J. Yu, S. Noh, Y. Jang, G. Park, and W. Woo. A hand-based collaboration framework in egocentric coexistence reality. In *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pp. 545–548. IEEE, 2015.

[74] K. Zibrek, E. Kokkinara, and R. Mcdonnell. The effect of realistic appearance of virtual characters in immersive environments - does the

character's personality play a role? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1681–1690, April 2018. doi: 10.1109/TVCG.2018.2794638

[75] F. R. H. Zijlstra. Efficiency in work behaviour: A design approach for modern tools. 1995.