International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2022)
J.-M. Normand and H. Uchiyama (Editors)

Evaluating Techniques to Share Hand Gestures for Remote Collaboration using Top-Down Projection in a Virtual Environment

Theophilus Teo[†], Kuniharu Sakudara, Masaaki Fukuoka, and Maki Sugimoto

Keio University, Japan

Abstract

Sharing hand gestures in a remote collaboration offers natural and expressive communication between collaborators. Proposed techniques allow sharing dependent (attached to something) or independent (no attachment) hand gestures in an immersive remote collaboration. However, there are research gaps for sharing hand gestures using different techniques and how it impacts user behaviour and performance. In this paper, we propose an evaluation study to compare sharing dependent and independent hand gestures. We developed a prototype, supporting three techniques of sharing hand gestures: Attached to Local, Attached to Object, and Independent Hands. Also, we use top-down projection, an easy-to-setup method to share a local user's environment with a remote user. We compared the three techniques and found that independent hands help a remote user guide a local user in an object interaction task quicker than hands attached to the local user. It also gives clearer instruction than dependent hands despite limited depth perception caused by top-down projection. A similar trend is also found in remote users' preferences.

CCS Concepts

• ***Human-centered computing*** → ***Systems and tools for interaction design;*** *Interaction design;*

1. Introduction

Remote collaboration techniques can enable a local worker to receive helps from a remote expert from different locations. The advancement of Augmented Reality (AR) and Virtual Reality (VR) technologies has offered enhancements to remote collaboration. A remote worker wears a VR Head Mounted Display (HMD) immersed in an environment shared from a local worker. Inside the environment, the remote worker can share expressive communications [AL11a] with a local worker through hand gestures using trackers [TLBA18] [CM16] or VR controllers [cit19]. This process allows a remote worker to feel as if his/her hands were extended to the shared environment [TVS*19]. It also enhances collaborative performance and user experience [TLBA18] [HBAK18a] [PDE*17]. Proposed techniques for sharing hand gestures in a remote collaboration categorize into dependent or independent hands. Dependent hands share hand gestures by attaching the hands to a user or an object. On the opposite, independent hands are shared by visualising the hands in the local worker's environment as an independent unit. However, it is uncommon to see both techniques co-exist because of the complications that result in a reduced system usability, which often leads to additional learning time and a cumbersome interface design [TLL*19].

Past works demonstrate sharing dependent hands in remote collaboration using standard camera(s) [KR15] [LZF17] or 360-degree camera [TLBA18] [LTKB17]. Hands attached to a local worker often overlaid on the screen, thus, do not require positional data. However, hands attached to an object must rely on these

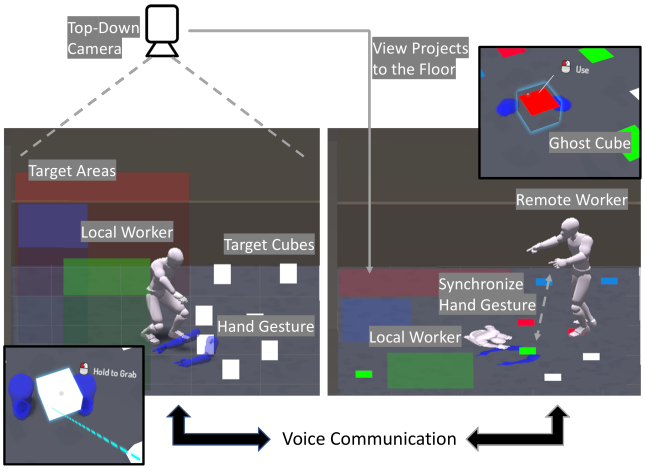

Figure 1: *System overview featuring a remote worker sharing object-dependent hand gestures to a local worker in a virtual environment.*

[†] e-mail: theop.teo@imlab.ics.keio.ac.jp

DOI: 10.2312/egve.20221272

data to work properly. As a result, it is not a popular option in a portable setup with minimum tracking data. Likewise, independent hands require spatial data of an environment. These techniques are common in a 3D system. 3D remote collaboration allows a local worker to share their environment through 3D reconstruction, beforehand [TLL*19] or in real-time [AAT13a]. It offers the system with spatial data to allow for sharing independent hands in a shared environment. Despite a 3D environment enhances the performance of sharing hand gestures, the configuration and cost can scale up dramatically for a large-scale setup.

Sharing dependent or independent hands are useful for tasks requiring a specific way to express instructions while merging them offers benefits in solving mixed or complicated problems [TLL*19]. However, it is uncertain if shared by using different techniques on the same task can influence user experience and problem-solving techniques. Moreover, can we design a simple prototype that supports sharing both dependent and independent hand gestures? Thus, we focus on the research question of **RQ1**: whether different ways of sharing hand gestures by a remote expert can influence the performance of a local user in an immersive remote pick-and-put task. Also, **RQ2**: how could it influence the way a remote expert communicates (verbal and nonverbal) with the local user. In this paper, we propose a study to evaluate sharing dependent and independent hand gestures in a remote collaboration. The contribution of this paper is twofold. First, it is the first user study that compares and evaluates sharing user-dependent, object-dependent, and independent hand gestures in the context of a collaborative object interaction task. Second, it explores the benefits and implications of a remote collaboration system using top-down projection to share a local worker's environment.

2. Related Work

2.1. Sharing Environment via Top-Down Perspective

Exploring a virtual environment from different perspectives can influence a user's behavior [UPS*10] and performance [UWTW19]. Ujkani et al. [UWTW19] reported perspectives that were closer to the user offer better precision whereas farther perspectives are suitable for understanding the overview of an environment. Furthermore, a top-down perspective can influence performance depending on the viewing distance and task conditions. Tang and Minneman [TM90] demonstrated sharing the view in a top-down perspective can be used for a collaborative drawing task. Also, Pinhanez and Pingali [PP04] proposed sharing by overlaying visual cues on a top-down perspective offers benefits to make expressive communications through hand gestures [AL11b] or visual annotations [AAT13b]. Recent works present a variation of sharing a top-down perspective via an overhead 360 panorama camera for quick travel [IUHI19] and mid-air collaboration [SOWM21]. These works indicate sharing a top-down perspective offers benefits for tasks that do not rely on depth information.

2.2. Sharing Dependent and Independent Hand in Remote Collaboration

Collaboration involves multiple peoples work together to solve one or several tasks. Two workers may (synchronously or asynchronously) work (closely or loosely) on (shared or individual) tasks from (same or different) location [TTP*06] [LCBLL16] [LCKM22]. That being said, a remote collaboration comprises collaborators of distinct locations to work together, which limits how peoples can interact or communicate in many ways. Sharing hand gestures in remote collaboration were studied for decades, exploring its benefits [AL11a] and usability [WA07]. Lee et al. [LTKB17] demonstrated sharing a live panorama view and user-dependent hand gestures helped a collaborator to understand their partner's focus in remote collaboration. Likewise, Gao et al. [GBLB16] found sharing user-dependent hand gestures with an oriented view can help remote collaborators to feel spatially and mentally connected. Separate experiments also indicate a lower mental effort [HBAK18b], as well as improvements in performance on the remote worker [LZF17] on unfamiliar tasks. When applied to robotics, Saraiji et al. [SSM*18] developed wearable supernumerary robotic limbs controlled by a remote user. The system allows a local worker to make direct, enforced, and induced communication with a remote worker to solve different collaborative tasks. Takizawa et al. [TVS*19] found that remote-controlling supernumerary robotic limbs attached to the local worker's body also improve user embodiment of the remote worker.

On the opposite, sharing independent hands are common in 3D remote collaboration. A remote worker can share 3D hand gestures as if they exist independently in the shared environment. Works demonstrated sharing independent hand gestures for collaborative tabletop tasks [TAH12] [WBB*21] and room-scale tasks [TLL*19] [BSYB20] [PLLB17]. In a user study evaluating the effects of sharing eye gaze and hand gestures in remote collaboration, Bai et al. [BSYB20] suggested sharing independent hands can induce a stronger sense of copresence among collaborators in an object interaction task. Later, Wang et al. [WBB*21] extended sharing hand gestures for user training in an industry scenario as it can improve user experience and a reduce completion time in an assembly task. Wood et al. [WTF*16] propose a technique to guide a local worker using 3D hand gestures through a monitor display. Their technique indicates enhancements in selection accuracy while promoting mutual understanding and engagement. Besides sharing hand gestures on the monitor, Kim et al. [KJP*20] demonstrate attaching hand gestures to points of interest in a 3D environment to reduce communication and spatial misunderstanding. Their work features a technique using object-dependent hand gestures for object selection tasks. At last, Teo et al. [TLL*19] proposed a hybrid prototype that supports switching between user-dependent and independent hand gestures. While their work compared 360 and 3D remote collaboration, it demonstrates techniques to solve collaborative tasks by mixing user-dependent and independent hand gestures.

3. Prototype Overview

3.1. Virtual World for Remote Collaboration

Figure 1 illustrates an overview of a remote collaboration prototype developed in a virtual social platform. Users wear a VR HMD and connect to a world in the platform to see each other in a 3D generic avatar. These avatars implemented inverse kinematics to support motion animations as users roam in the world via VR HMD and VR controllers. We built two separate rooms in the world for users

to perform remote collaboration. On one side, a local room consists of interactive objects and an overhead virtual camera for sharing the environment with the remote user. The virtual camera is positioned 3.075m from the ground and uses orthographic projection with a size of 2.1. On the other side, a remote room is an empty room with a floor (4m x 4m) that renders the preview of the local room's environment through the virtual camera (1048p x 1048p). Users can communicate verbally using a mic and speaker despite their positions in the world.

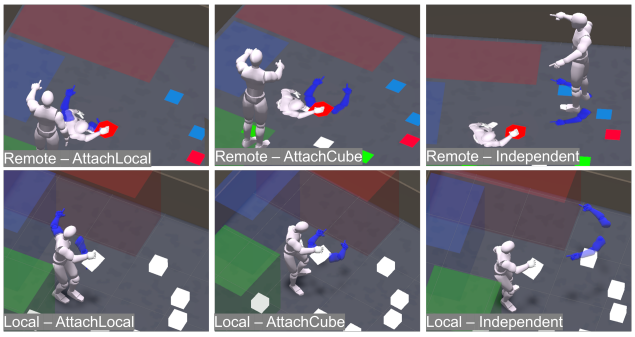

Figure 2: *Three conditions of the experiment from a local user's perspective. (a) Attach to Local, (b) Independent Hand, (c): Attach to Object*

3.2. Dependent and Independent Hand Gestures

The world integrates three known techniques of sharing hand gestures: Attach to Local [TLBA18], Attach to Object [KJP*20], and Remote Independent [BSYB20]. A remote user shares hand gestures by attaching virtual arms to a local user, an object, or independently in the local room. We use an animated 3D blue arm model with 50% opacity for the virtual arms. The virtual arms synchronize with the remote user's arms using rotation information. Before this, users' hand gestures were managed by matching the figure postures against their VR controllers. This allows users to make hand gestures in the world for nonverbal communication.

The virtual arms require specific position information to share hand gestures. In user-dependent hands, the virtual arms' position synchronizes with the local user. Likewise, virtual arms can be attached to an object in object-dependent hands using the corresponding information. However, this requires the remote user to specify a target object. To achieve this, the remote user needs to select an object in the local room through the display on the floor. We use a transparent ghost cube (See Fig.1, top right) to detect user selection and inform the corresponding object in the local room to provide the position data to the virtual arms. For independent hands, the remote user controls the virtual arms' movements in 6 degrees of freedom (DOF) using their position data while it appears in the local room. Yet, synchronizing virtual arms this way will render the virtual arms on the remote user's avatar. Hence, we add an offset to the position data, so virtual arms appear in the local room as if the remote user co-exists in the room. Figure 2 illustrates an overview of three techniques.

3.3. Implementation

We developed the virtual world using the Unity (2019.4.31f1) game engine running on VRChat (v2022.1.2), a social VR Platform on a Windows 10 PC (Intel Core i7-9750H CPU, 2.60GHz, 16GB RAM, NVIDIA GeForce RTX 2080 GPU). The local user on the PC side wears an HTC Vive Eye Pro (VR HMD) with 1440 x 1600 pixels per eye with a 90Hz refresh rate and 110-degree field of view. A remote user connects and joins the virtual world on an Oculus Quest 1 (Android 10 Standalone VR HMD, Octa-core Kryo 280 CPU, 4 x 2.45 GHz and 4 x 1.9 GHz, 4GB RAM, Adreno 540 GPU with 1440 x 1600 pixels per eye with a 72Hz refresh rate and 90-degree field of view). Vive lighthouse tracking was used to provide tracking for the VR HMD and the VR Controllers to allow for VR locomotion, body, and hand gestures. However, an internal tracker is used on the Oculus Quest 1 to achieve similar results. Users can communicate in the virtual world through VRChat Player Audio supports.

4. User Study

A user study was conducted to compare and evaluate the effects of sharing dependent and independent hand gestures in a remote collaboration system that shares an environment with limited depth information. We designed object interaction tasks to collect data and feedback for evaluation.

4.1. Task Environment and Design

Multiple rooms were constructed in the virtual world for different activities. A user who enters the world spawns in a waiting room for a quick briefing. The waiting room is an enclosed room with walls on all sides. One of the walls has simple texts, describing the task procedure and basic controls. Next, the user performs collaborative task activities in the task room, which is broken into local and remote rooms according to the user's role (See Fig.1). Finally, the user answers questionnaires in a survey room. The survey room is also an enclosed room with walls and simple texts for questions. In addition, seven cuboids are placed and horizontally aligned near the center area of the room. Each cuboid represents a response for a Likert scale, indicated by a floating text description (E.g.: Leftmost cuboid: Strongly Disagree / Rightmost cuboid: Strongly Agree). The user makes a response by pointing and clicking a cuboid. Doing so updates the text on the wall to the next question, repeating the procedure until all questions were asked. Despite different activities, all rooms are of equal size (4m x 4m) with a special cuboid to send users to the proceeding room. The proceeding room depends on the users' role and task progressions (See Fig.1). However, users must return to the center area of the room before traveling to the next room because the cuboid always spawn the user at the center area. Hence, we employed an actor in the experiment to notify the participant for their reposition to prevent spatial offset that may lead to physical hazards such as walking to walls.

For task design, the local room consists of interactive task cubes and task zones. The local user needs to restore the room by moving task cubes (25cm x 25cm x 25cm) to the task zones, as indicated by the color of the cubes and zones (E.g.: blue cubes into the blue zone etc...). In addition, we introduced two white cubes to the room as

noise cubes. On the opposite side, the remote user communicates with the local user from the remote room via verbal and nonverbal communications. To create a remote-expert scenario, the local user sees all task cubes in white colors, whereas the remote user sees the task cubes through the shared environment on the floor in original colors. As a result, the remote user collaborates with the local user to move colored task cubes to the respective task zones, one at a time.

In the task session, 8 cubes (2 red, 2 green, 2 blue, and 2 white) and 18 cubes (4 red, 4 green, 4 blue, and 2 white) are used for the training and main sessions. Despite the session, there are always three zones (red, green, and blue) of different sizes but a fixed volume ($100cm^3$). We designed six task sets (exclude training) using different orientations of the cubes and zones to reduce the bias effect.

For the physical setup, an actor was employed in the study to take part as a local user since the experiment focused on exploring the remote user's experience and performance when sharing hand gestures by manipulating them. The actor remotely joins as a local user with whom a participant needs to collaborate in the virtual world. Despite the role, both users enter the virtual world while wearing a VR HMD and standing in an empty, open area with 4m x 4m physical space.

4.2. Task Conditions and Restrictions

The experiment is a within-subject study that consists of three conditions (1) AttachLocal: "User-Dependent Hand", (2) AttachObj: "Object-Dependent Hand", and (3) IndHand: "Independent hand". Participants attempt all conditions in a single session as remote users. For the purpose of the evaluation, participants were encouraged to use hand gestures as the main technique to give instructions and verbal to aid the gestural cue. Examples include making a pointing gesture while saying "put this there...", doing a waving gesture while saying "over here...", or pointing in a direction while saying "this one...". This prevents participants to complete the task without trying hand gestures since the task can be solved by telling the color, such as "this is a red cube..." or "this goes to the green zone...".

4.3. Procedures

An experiment session with a participant started with the researcher briefing the study information to the participant through a video call. Participants attended the experiment in a different room as if connecting from a remote location. Following this, the participants signed a digital consent form once they agreed to participate. The participant was then asked to stand in the center area of the room and enter the virtual world while wearing an Oculus Quest. In the virtual world, the researcher "actor" invited the participants to familiarize themselves by moving around while explaining the task objectives. Then, the actor clicked the cuboid and a training session was held for the participant to learn the tasks and system features. The participant collaborated with the actor to move task cubes (two task cubes for every hand gesture technique) to the corresponding task zones.

After the training, the participant returns to the center area in the room and the actor clicks the cuboid to warp back to the waiting room. The participant is invited to take a five-minute break or continue without one. Once ready, the actor informs the participant for repositioning and warp to the activity rooms. The actor notifies the participant of the first condition and begins the collaborative task. The task is completed after all 16 task cubes are placed in the correct task zones. This triggers the long cuboid to spawn. The actor notifies the participant of the repositioning, then sends the participant to the survey room while the actor returns to the waiting room. The participant answers all questionnaires in the survey room and clicks the long cuboid at the end to return to the waiting room. The participant then proceeds into the second task after an optional five-minute break. This process was repeated with the second condition (the order of conditions was counterbalanced between participants). After completing all conditions, the participant was asked to complete a post-experiment questionnaire on a web browser. Overall, the experiment took 60 min on average for each participant.

4.4. Measurements

We collected objective measures and subjective feedback in each condition. The objective measures were task completion time and user gaze focus data. For subjective feedback, we collected questionnaire responses. This includes custom-developed questions accessing participants' sense of agency and sense of body ownership when using the different hand gestures techniques, Networked Mind Measure of Social Presence Questionnaire (SoPQ) [HB04], and another two custom-developed questionnaires accessing participant's collaborative experience and task difficulty. At the end of the experiment, we used a post-experiment question to access participants' user preferences between the hand gestures techniques, and also qualitative feedback on reasons for their preferences and other subjective comments.

5. Results

18 participants from the local campus (15 male, 3 female) were recruited for the study, with their ages ranging from 22 to 29 (M=23.7, SD=1.87). Half of the participants reported that they used VR at least a few times a month. Two participants from another half reported zero VR experience and the remaining participants with a yearly VR experience. This section reports the quantitative and objective measurements data, alongside qualitative feedback. The results were statistically analyzed using $\alpha = .05$ unless noted otherwise.

5.1. Task Completion Time

Figure 3 illustrates the average task completion time under different conditions. A Friedman test indicated a significant difference in the task completion time ($X^2(2)=21.778$, $p<.0005$). Post hoc analysis using Wilcoxon Signed Rank tests with Bonferroni correction (alpha=.0167) revealed that IndHand took significantly lesser time than AttachLocal (Z=-3.724, $p=<.0005$) and AttachObj (Z=-2.418, $p=.016$) in completing the task. Yet, no significant difference was found between AttachLocal and AttachObj (Z=-2.027, $p=.043$).

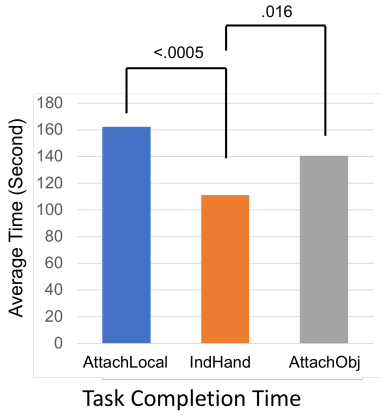

Figure 3: *Average task completion time (*: significant effect)*

5.2. User Gaze Focus

Figure 4 illustrates user gaze focus based on the duration (seconds) that the participants (N=17) had gazed upon the task-related objects during the task session. One participant's data was not recorded due to systematic error. The data was formatted into the percentage using the task completion time. A Friedman test found that participants' attention on the local user is statistically significant between the conditions (X^2(2)=16.353, p=<.0005). Post hoc analysis using Wilcoxon Signed Rank test with Bonferroni correction (alpha=.0167) has also indicated participants were paying significantly longer attention to the actor on AttachLocal when compared against AttachObj (Z=-3.574, p<.0005) and IndHand (Z=-2.627, p=.009). However, no significant differences were found for the other task-related objects.

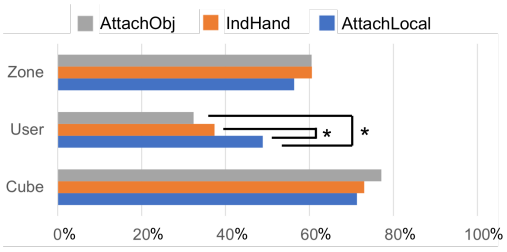

Figure 4: *Participant's gaze focus (%) on task related objects (Zone: Task zones, User: Local User/Actor, Cube: Task cubes; *: significant effect)*

5.3. Sense of Agency and Body Ownership

Participants' effects on hand ownership and sense of agency were measured using custom questions. Questions are accessed based on a 7-point Likert scale of (1: fully disagree-7: fully agree). Figure 5 illustrates individual questions, participants' ratings by the condition, and analyzed results. Significant differences are found using Friedman test (See columns 6 and 7, Figure 5).

Post hoc tests using Wilcoxon Signed Rank test with Bonferroni correction (alpha=.0167) indicate participants felt that their agency to control the virtual arms (Q1) were significantly stronger on IndHand when compared against AttachLocal (Z=-3.448, p=.001) and also AttachObj (Z=-2.541, p=.011). Participants also felt significantly constrained at controlling the arms with their movements (Q2) on AttachLocal, compared to IndHand (Z=-3.370, p=.001) and AttachObj (Z=-2.631, p=.009). Participants also felt significantly easier to control the location of the virtual arm at their own will (Q3) with IndHand, compared to AttachLocal (Z=-3.428, p=.001) and AttachObj (Z=-2.881, p=.004). However, participants only felt the virtual arms were synchronized with their movements (Q4) on IndHand, compared to AttachLocal (Z=-2.974, p=.003). In body ownership for having the virtual arms as part of the body (Q5), the effect is significant on IndHand, against the AttachLocal (Z=-3.207, p=.001). Surprisingly, participants only felt the presence of their bodies in the local room (Q6) on IndHand, when compared against AttachObj (Z=-3.230, p=.001) but not AttachLocal (Z=-1.044, p=.0296). In terms of embodiment, participants in AttachObj significantly thought that they became the cube (Q7) on AttachObj than the IndHand (Z=-2.738, p=.006). In addition, participants significantly felt that the virtual arms belong to someone else on AttachLocal than the IndHand (Z=-2.550, p=.011).

5.4. Collaborative Experience and Task Difficulty

Q11 and Q12 in Figure 5 report participants' ratings towards the collaborative experience and task difficulty. Friedman test indicates a significant difference on Q17, following a post hoc test using Wilcoxon Signed Rank test with Bonferroni correction (alpha=.0167) that reported that the task is significantly difficult on AttachLocal than IndHand (Z=-3.442, p=.001) and AttachObj (Z=-3.224, p=.001).

5.5. Social Presence

Figure 6 illustrates participants' average social presence rating. The SoPQ [HB04] included three sub-scales: Co-Presence (CP), Attentional Allocation (AA), and Perceived Message Understanding (PMU). The questionnaire consists of eighteen rating items on a 7-point Likert scale (1: strongly disagree- 7: strongly agree). The significant difference was not detected using Friedman test on the main scale (X^2(2)=1.400, p=.497). However, all conditions reported a high average social presence score (AttachLocal: M=5.559, SD=0.877; IndHand: M=5.898, SD=0.626; AttachObj: M=5.926, SD=0.527). This suggested significant results when tested against the medium score (S=4) using Wilcoxon Signed Rank test (AttachLocal: Z=-3.725, p<.0005; IndHand: Z=-3.729, p<.0005; AttachObj: Z=-3.725, p<.0005). A similar trend is also found when using Friedman test on the subscales, reporting no significant difference when comparing the conditions but for comparing the medium score.

5.6. User Preferences

Participants were asked to rank (1: best-3: worse) all the three conditions they have experienced based on their preference at the end of the experiment. The majority of participants ranked the IndHand as the best (61%) and AttachLocal as the worst (83%) (See Figure

7). A Friedman test indicated that there was a significant difference in ranking between the three conditions ($X^2(2)$ =14.778, p=.001). Post hoc tests using Wilcoxon Signed Rank tests with Bonferroni correction (alpha=.0167) showed that participants preferred IndHand significantly more than AttachLocal (Z=-2.741, p=.006) and AttachObj (Z=-2.946, p=.003).

Q1-Q12: Fully Disagree … Fully Agree		Mean	MD	$\chi^2(2)$	p
Agency and Body Ownership					
Q1: I felt like I could control the virtual arms as if it was my own body.	AttachLocal	4.28	4.0	20.933	**<0.0005**
	IndHand	6.11	6.5		
	AttachObj	5.22	5.0		
Q2: I felt like I controlled the left and right virtual arm with the movements of my left and right hands.	AttachLocal	5.06	5.0	17.167	**<0.0005**
	IndHand	6.39	7.0		
	AttachObj	6.00	6.0		
Q3: I felt like I could control the location of the virtual arm at my own will.	AttachLocal	3.50	3.0	22.727	**<0.0005**
	IndHand	6.44	6.5		
	AttachObj	4.89	5.5		
Q4: The movements of the virtual arms were caused by my movements.	AttachLocal	5.06	5.0	21.000	**<0.0005**
	IndHand	6.44	7.0		
	AttachObj	5.56	6.0		
Q5: I felt as if the virtual arms were part of my body.	AttachLocal	3.17	3.0	15.527	**<0.0005**
	IndHand	4.89	5.5		
	AttachObj	3.78	3.0		
Q6: I felt as if my body was located on the place where I saw the virtual arms.	AttachLocal	3.11	3.0	14.780	**0.001**
	IndHand	4.67	4.0		
	AttachObj	2.61	2.0		
Q7: It seemed as if my body became the cube.	AttachLocal	1.67	1.5	6.500	**0.039**
	IndHand	1.78	1.5		
	AttachObj	2.89	2.5		
Q8: It seemed as if my body became the local user.	AttachLocal	3.28	3.0	5.216	0.074
	IndHand	2.44	2.0		
	AttachObj	2.44	2.0		
Q9: It seemed as if my body became the room.	AttachLocal	2.11	2.0	0.438	0.804
	IndHand	2.06	2.0		
	AttachObj	2.33	2.0		
Q10: I felt as if the virtual arms I saw on the floor were someone else.	AttachLocal	3.33	3.0	0.478	**0.024**
	IndHand	2.17	2.0		
	AttachObj	2.56	2.5		
Experience and Difficulty					
Q11: I felt collaborative with my partner.	AttachLocal	6.06	6.0	1.024	0.599
	IndHand	6.22	6.0		
	AttachObj	6.11	6.0		
Q12: Overall, the task was easy.	AttachLocal	4.33	5.0	24.471	**<0.0005**
	IndHand	6.11	6.0		
	AttachObj	5.78	6.0		

Figure 5: *Subjective ratings on the user embodiment, collaborative experience, and task difficulty (**bold**: significant effect, **MD**: Median)*

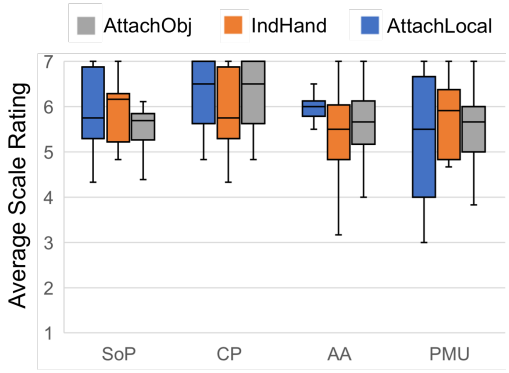

Figure 6: *Results of Social Presence questionnaire (SoP: Social Presence, CP: Co-Presence, AA: Attentional Allocation, PMU: Perceived Message Understanding).*

5.7. User Behavior and Subjective Comments

Participants employed unique behavior and technique for each condition. From observations, participants used minimal verbal communication on IndHand and AttachObj but a greater amount on AttachLocal. Nine participants were not confident with using hand

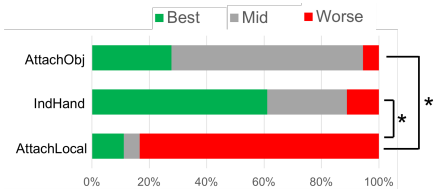

Figure 7: *User ranking preference (*: significant effect)*

gestures because of indirect perspective. Participants mentioned (P12) *"I must give instruction based on the local user's perspective... it was not possible to point at a cube that is behind another cube"*, (P5) *"... there was confusion when pointing at a cube or zone that is behind another cube or zone... it felt the least efficient"* and (P4) *"the visual cue highly depend on my partner's location"*. Also, participants highlighted a better focus is essential for AttachLocal. Some commented (P16) *"AttachLocal imposed me to watch the local user constantly to adjust the arm direction according to the local user's position"* and (P2) *"since my partner would move and I had to find the local user again before pointing at the cube I wanted to refer to"*. However, one participant commented (P11) *"AttachLocal made me focus on my partner the best and enabled better cooperation"*. Some participants preferred AttachLocal because they thought it (P17) *"was very interesting for collaboration"* with (P15) *"a better ease of use"*.

On the opposite, participants enjoyed IndHand because of high DOFs. Participants commented IndHand leads to natural and efficient communication. They mentioned, (P13) *"I can give instruction while moving the hand position freely"* and (P2) *"I could just supervise everything and walk to the location of the cube I wanted to show"*, thus (P12) *"easiest to operate and give instructions to the local user"*. It also helped (P5) *"pointing at cubes and zones was fast and without confusion"*, which makes the instruction (P14) *"easier for the local user to understand my thought"*. In another observation, a participant adapted a strategy by pointing while standing in a spot that is adjacent to the cube and the zone. It allows participant to instruct the actor without single verbal communication. In AttachObj, a few participants (N=4) thought it was fun as it allows them to (P1) *"take control of a cube and have arms grow out of it"* or (P13) *"to navigate the local user by pointing hand gestures from the perspective of a cube"*. Hence, it (P6) *"feels more collaborative with the local user"*. Furthermore, one participant appended that AttachObj (P4) *"solves the issue of the cube localization"* because (P2) *"it was easier to say which cube I was referring to as I could simply select it on AttachObj"*. Participants do not feel a similar constraint on AttachObj as the AttachLocal despite being a dependent cue. They commented the AttachObj was (P8) *"... a bit more difficult but it is still easy to point a direction even if I could not move the rest of my body"*.

6. Discussion

6.1. Dependent and Independent Hand Gestures in Remote Collaboration

Overall, the results suggest sharing an environment with limited depth perception benefits independent hand gestures in some ways.

This includes a shorter task completion time, better agency, body ownership, and reduced task difficulty in an object interaction task. Participants felt using IndHand is easier than AttachLocal and AttachObj because of mental rotation when using verbal and non-verbal instructions in AttachLocal and AttachObj. Furthermore, it is also harder for participants to give instruction in AttachLocal and AttachObj while the local user or the task object is in motion. This result is different from a study [TLL*19] that compares 360-panoramic and 3D immersive remote collaboration where hand gestures and visual pointers were shared using a similar technique as user-dependent and independent cues. In the study, participants had significantly longer time and larger difficulty using independent cues than the user-dependent cues for object searching tasks. However, the authors discussed that the poor 3D reconstructed environment in the dependent visual cue may have biased the user-dependent visual cue.

Besides performance, custom questions for agency and body ownership indicate sharing hand gestures in different DOFs can influence user experience. Participants felt stronger at controlling the virtual arms as if it was their own body in IndHands than AttachObj and AttachLocal. Likewise, they felt stronger as if the virtual arms belonged to someone else in AttachLocal than in other conditions. This suggests a relationship between user embodiment (agency and body ownership) and the amount of DOF when sharing visual cues that represent a part of the human body in remote collaboration.

In terms of body presence in the shared environment, participants reported a strong effect in the shared environment in IndHand than AttachObj. We suspect a non-humanoid body can negatively impact participants' sense of body presence. Yet, we need further study to verify if a similar effect exists in sharing user-dependent hands. In specific dependent-hand questions, participants thought their body became the cube in AttachObj when compared against IndHand. However, this is not the case for AttachLocal when asked if they felt their body became the local user. We postulate that the significant gaze focus on the local user between AttachLocal-AttachObj and AttachLocal-IndHand, as well as a greater amount of verbal communication, plays an important role to the result. For example, participants paid strong attention to the local user, hence is harder to feel that their bodies became the local user.

6.2. Remote Collaboration using Top-Down Projection

We proposed a technique to share a local user's environment using top-down projection. Despite the limited depth perception, the study result implies that sharing the environment in VR maintains a strong sense of social presence as indicated by the significant and above-average scores. Likewise, participants felt collaborative with their partners using the setup, but it is unclear if the hand-sharing techniques would cause an effect.

In task solving, participants who employ a strategy to solve the task by colors sometimes struggle to communicate with the local user in AttachLocal. This is because they had a stricter requirement of asking the local user to pick a cube with a specific color. However, it is difficult in a scenario when the cube is behind another cube or in a group. Supposedly, one can point by raising the arm at different angles to indicate distinct selections. Yet, the inability to see those angles from a top-down perspective limits it. Because of this, participants in IndHand adapted a strategy by walking to a spot where they can clearly and synchronously point at a cube and the task zone for giving instructions. While we expect a similar issue in AttachObj, the location of the task zone allows participants to point without worrying the ambiguity. As a result, more than half of the participants (N=11) preferred using IndHand to share hand gestures. Therefore, it is suitable to share an environment using top-down projection for tasks without strict requirements on precision [UWTW19].

7. Conclusion and Future Work

We presented an evaluation study to explore the impact of sharing hand gestures in user-dependent (AttachLocal), object-dependent (AttachObj), and independent (IndHand) settings. We designed a prototype that supports sharing a local user's environment using top-down projection. In a study based on an object interaction task, participants felt a strong sense of agency and performed better using IndHand. Participants also preferred IndHand over AttachObj and AttachLocal because of clearer instruction despite limited depth perception by the top-down perspective. Surprisingly, participants thought the virtual arms they manipulated belonged to someone else in AttachLocal. While we suspect the significant gaze attention on the local user could cause an influence, we need a further study to pinpoint the reason.

Despite the result, our task design could have influenced the outcome in some ways. First, the task restricted our findings to a remote-expert collaboration. Second, a verbal restriction to speak specific keywords could manipulate participants' methodology to solve the task. Although they mentioned sharing hand gestures caused verbal communication to be unnecessary, we suspect it was partially and passively enforced by the restriction. Therefore, we note that there is a need for further investigation to compare different hand gesture techniques on different view perspectives and tasks. For example, in an equal-role collaboration task or a larger room with a different layout. On AttachObj, we found interesting patterns as participants felt that their bodies became the cube but not when attached to the local user. A further evaluation to understand the reason can contribute and create extensions to embodiment and remote collaboration. Furthermore, we will add spatial audio to our future prototype for a realistic experience.

Overall, we expect our work will contribute to bridging robotics and remote collaboration. For example, a remote expert controls a pair of robotic limbs to guide a local user as if the limbs are naturally part of the remote expert's body. Depending on the task requirements, the robotic limbs can be worn as a backpack by the local user (AttachLocal), installed on an object (AttachObj), or a teleoperated robot (IndHand). A minimal design by sharing a local user's environment through top-down projection simplifies the setup while maintaining an effective collaboration.

Acknowledgement

This work was supported by JST ERATO Grant Number JPMJER1701, Japan.

References

[AAT13a] ADCOCK M., ANDERSON S., THOMAS B.: Remotefusion: Real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (New York, NY, USA, 2013), VRCAI '13, Association for Computing Machinery, p. 235–242. URL: https://doi.org/10.1145/2534329.2534331, doi:10.1145/2534329.2534331. 2

[AAT13b] ADCOCK M., ANDERSON S., THOMAS B.: Remotefusion: Real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (New York, NY, USA, 2013), VRCAI '13, Association for Computing Machinery, p. 235–242. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/2534329.2534331, doi:10.1145/2534329.2534331. 2

[AL11a] ALEM L., LI J.: A Study of Gestures in a Video-Mediated Collaborative Assembly Task. *Advances in Human-Computer Interaction 2011* (2011), 987830. URL: https://doi.org/10.1155/2011/987830, doi:10.1155/2011/987830. 1, 2

[AL11b] ALEM L., LI J.: A Study of Gestures in a Video-Mediated Collaborative Assembly Task. *Advances in Human-Computer Interaction 2011* (2011), 987830. URL: https://doi.org/10.1155/2011/987830, doi:10.1155/2011/987830. 2

[BSYB20] BAI H., SASIKUMAR P., YANG J., BILLINGHURST M.: *A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing*. Association for Computing Machinery, New York, NY, USA, 2020, p. 1–13. URL: https://doi.org/10.1145/3313831.3376550. 2, 3

[cit19] Asymmetric Interface: User Interface of Asymmetric Virtual Reality for New Presence and Experience. *Symmetry 12*, 1 (dec 2019), 53. URL: https://www.mdpi.com/2073-8994/12/1/53, doi:10.3390/sym12010053. 1

[CM16] CLARK A., MOODLEY D.: A system for a hand gesture-manipulated virtual reality environment. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists* (New York, NY, USA, 2016), SAICSIT '16, Association for Computing Machinery. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/2987491.2987511, doi:10.1145/2987491.2987511. 1

[GBLB16] GAO L., BAI H., LEE G., BILLINGHURST M.: An oriented point-cloud view for mr remote collaboration. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications* (New York, NY, USA, 2016), SA '16, Association for Computing Machinery. URL: https://doi.org/10.1145/2999508.2999531, doi:10.1145/2999508.2999531. 2

[HB04] HARMS C., BIOCCA F.: Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. *Seventh Annual International Workshop: Presence*, 2004 (2004), 246–251. 4, 5

[HBAK18a] HUANG W., BILLINGHURST M., ALEM L., KIM S.: Handsintouch: Sharing gestures in remote collaboration. OzCHI '18, Association for Computing Machinery, p. 396–400. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/3292147.3292177, doi:10.1145/3292147.3292177. 1

[HBAK18b] HUANG W., BILLINGHURST M., ALEM L., KIM S.: Handsintouch: Sharing gestures in remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (New York, NY, USA, 2018), OzCHI '18, Association for Computing Machinery, p. 396–400. URL: https://doi.org/10.1145/3292147.3292177, doi:10.1145/3292147.3292177. 2

[IUHI19] IZUMIHARA A., URIU D., HIYAMA A., INAMI M.: Exleap: Minimal and highly available telepresence system creating leaping experience. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), pp. 1321–1322. doi:10.1109/VR.2019.8798064. 2

[KJP*20] KIM S., JING A., PARK H., LEE G. A., HUANG W., BILLINGHURST M.: Hand-in-air (hia) and hand-on-target (hot) style gesture cues for mixed reality collaboration. *IEEE Access 8* (2020), 224145–224161. doi:10.1109/ACCESS.2020.3043783. 2, 3

[KR15] KASAHARA S., REKIMOTO J.: Jackin head: Immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2015), VRST '15, Association for Computing Machinery, p. 217–225. URL: https://doi.org/10.1145/2821592.2821608, doi:10.1145/2821592.2821608. 1

[LCBLL16] LIU C., CHAPUIS O., BEAUDOUIN-LAFON M., LECOLINET E.: Shared interaction on a wall-sized display in a data manipulation task. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), CHI '16, Association for Computing Machinery, p. 2075–2086. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/2858036.2858039, doi:10.1145/2858036.2858039. 2

[LCKM22] LIU H., CHOI M., KAO D., MOUSAS C.: Synthesizing game levels for collaborative gameplay in a shared virtual environment. *ACM Trans. Interact. Intell. Syst.* (aug 2022). Just Accepted. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/3558773, doi:10.1145/3558773. 2

[LTKB17] LEE G. A., TEO T., KIM S., BILLINGHURST M.: Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics amp; Interactive Applications* (New York, NY, USA, 2017), SA '17, Association for Computing Machinery. URL: https://doi.org/10.1145/3132787.3139203, doi:10.1145/3132787.3139203. 1, 2

[LZF17] LE K.-D., ZHU K., FJELD M.: Mirrortablet: Exploring a low-cost mobile system for capturing unmediated hand gestures in remote collaboration. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (New York, NY, USA, 2017), MUM '17, Association for Computing Machinery, p. 79–89. URL: https://doi.org/10.1145/3152832.3152838, doi:10.1145/3152832.3152838. 1, 2

[PDE*17] PIUMSOMBOON T., DAY A., ENS B., LEE Y., LEE G., BILLINGHURST M.: Exploring enhancements for remote mixed reality collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics amp; Interactive Applications* (New York, NY, USA, 2017), SA '17, Association for Computing Machinery. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/3132787.3139200, doi:10.1145/3132787.3139200. 1

[PLLB17] PIUMSOMBOON T., LEE Y., LEE G., BILLINGHURST M.: Covar: A collaborative virtual and augmented reality system for remote collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies* (New York, NY, USA, 2017), SA '17, Association for Computing Machinery. URL: https://doi.org/10.1145/3132818.3132822, doi:10.1145/3132818.3132822. 2

[PP04] PINHANEZ C., PINGALI G.: Projector-camera systems for telepresence. In *Proceedings of the 2004 ACM SIGMM Workshop on Effective Telepresence* (New York, NY, USA, 2004), ETP '04, Association for Computing Machinery, p. 63–66. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/1026776.1026795, doi:10.1145/1026776.1026795. 2

[SOWM21] SABET M., ORAND M., W. MCDONALD D.: Designing telepresence drones to support synchronous, mid-air remote collaboration: An exploratory study. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2021), CHI '21, Association for Computing Machinery. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/3411764.3445041, doi:10.1145/3411764.3445041. 2

[SSM*18] SARAIJI M. Y., SASAKI T., MATSUMURA R., MINAMIZAWA K., INAMI M.: Fusion: full body surrogacy for collaborative communication. pp. 1–2. doi:10.1145/3214907.3214912. 2

[TAH12] TECCHIA F., ALEM L., HUANG W.: 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* (New York, NY, USA, 2012), VRCAI '12, Association for Computing Machinery, p. 323–328. URL: https://doi.org/10.1145/2407516.2407590, doi:10.1145/2407516.2407590. 2

[TLBA18] TEO T., LEE G. A., BILLINGHURST M., ADCOCK M.: 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* (New York, NY, USA, 2018), OzCHI '18, Association for Computing Machinery, p. 406–410. URL: https://doi.org/10.1145/3292147.3292200, doi:10.1145/3292147.3292200. 1, 3

[TLL*19] TEO T., LAWRENCE L., LEE G. A., BILLINGHURST M., ADCOCK M.: Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, p. 1–14. URL: https://doi.org/10.1145/3290605.3300431, doi:10.1145/3290605.3300431. 1, 2, 7

[TM90] TANG J. C., MINNEMAN S. L.: Videodraw: A video interface for collaborative drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1990), CHI '90, Association for Computing Machinery, p. 313–320. URL: https://doi.org/10.1145/97243.97302, doi:10.1145/97243.97302. 2

[TTP*06] TANG A., TORY M., PO B., NEUMANN P., CARPENDALE S.: Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI '06, Association for Computing Machinery, p. 1181–1190. URL: https://doi-org.kras.lib.keio.ac.jp/10.1145/1124772.1124950, doi:10.1145/1124772.1124950. 2

[TVS*19] TAKIZAWA R., VERHULST A., SEABORN K., FUKUOKA M., HIYAMA A., KITAZAKI M., INAMI M., SUGIMOTO M.: Exploring perspective dependency in a shared body with virtual supernumerary robotic arms. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2019), pp. 25–257. doi:10.1109/AIVR46125.2019.00014. 1, 2

[UPS*10] USTINOVA K., PERKINS J., SZOSTAKOWSKI L., TAMKEI L., LEONARD W.: Effect of viewing angle on arm reaching while standing in a virtual environment: Potential for virtual rehabilitation. *Acta Psychologica 133*, 2 (2010), 180–190. URL: https://www.sciencedirect.com/science/article/pii/S0001691809001656, doi:https://doi.org/10.1016/j.actpsy.2009.11.006. 2

[UWTW19] UJKANI A., WILLMS J., TURGUT L., WOLF K.: The effect of camera perspectives on locomotion accuracy in virtual reality. In *Proceedings of Mensch Und Computer 2019* (New York, NY, USA, 2019), MuC'19, Association for Computing Machinery, p. 835–838. URL: https://doi.org/10.1145/3340764.3344918, doi:10.1145/3340764.3344918. 2, 7

[WA07] WICKEY A., ALEM L.: Analysis of hand gestures in remote collaboration: Some design recommendations. In *Proceedings of the 19th Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces* (New York, NY, USA, 2007), OZCHI '07, Association for Computing Machinery, p. 87–93. URL: https://doi.org/10.1145/1324892.1324909, doi:10.1145/1324892.1324909. 2

[WBB*21] WANG P., BAI X., BILLINGHURST M., ZHANG S., WEI S., XU G., HE W., ZHANG X., ZHANG J.: 3dgam: using 3d gesture and cad models for training on mixed reality remote collaboration. *Multimed Tools Appl 80* (2021), 31059–31084. doi:https://doi.org/10.1007/s11042-020-09731-7. 2

[WTF*16] WOOD E., TAYLOR J., FOGARTY J., FITZGIBBON A., SHOTTON J.: Shadowhands: High-fidelity remote hand gesture visualization using a hand tracker. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces* (New York, NY, USA, 2016), ISS '16, Association for Computing Machinery, p. 77–84. URL: https://doi.org/10.1145/2992154.2992169, doi:10.1145/2992154.2992169. 2