ReMotion: Supporting Remote Collaboration in Open Space with Automatic Robotic Embodiment

Mose Sakashita, Ruidong Zhang, Xiaoyi Li, Hyunju Kim, Michael Russo,
Cheng Zhang, Malte F. Jung, François Guimbretière
Cornell University
USA

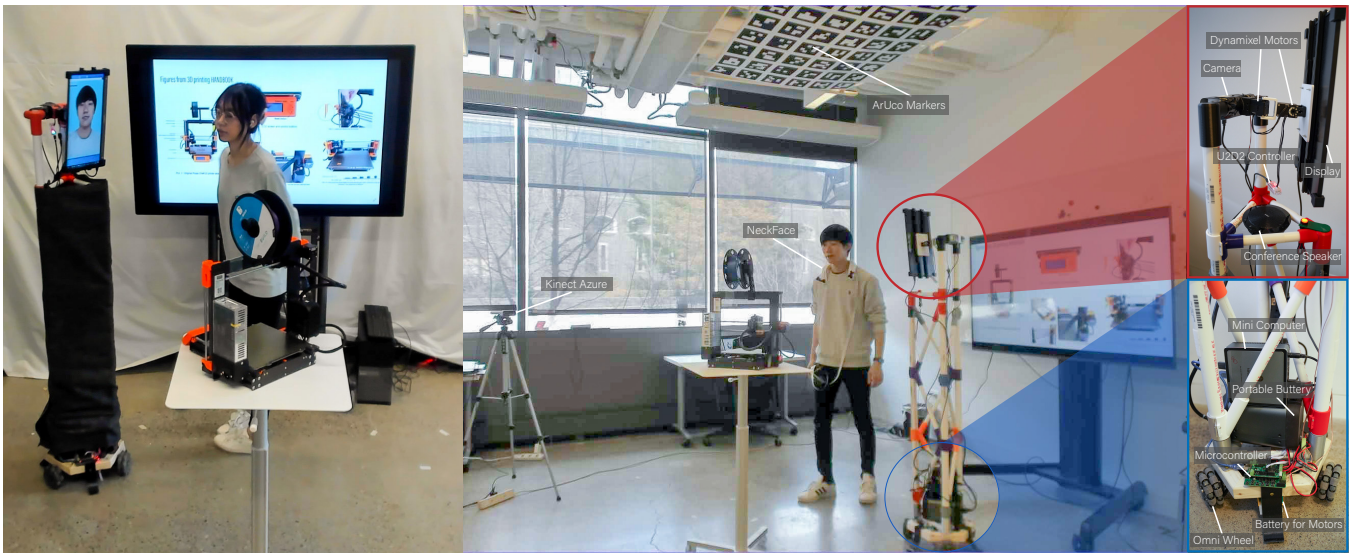

Figure 1: Two collaborators are reviewing a physical artifact remotely by using the ReMotion system while sharing information on a digital whiteboard. We show both sites (Left and Center). At each site, the system includes (Center): a Kinect Azure to capture body motions, a NeckFace system worn by the local user to capture head and facial expressions, the embodiment robot, and ArUco markers board used by the robot for position feedback. The ReMotion robotic proxy (Right) uses an omnidirectional platform for movement flexibility and an articulated display to render head orientation and facial expressions.

ABSTRACT

Design activities, such as brainstorming or critique, often take place in open spaces combining whiteboards and tables to present artefacts. In co-located settings, peripheral awareness enables participants to understand each other's locus of attention with ease. However, these spatial cues are mostly lost while using videoconferencing tools. Telepresence robots could bring back a sense of presence, but controlling them is distracting. To address this problem, we present *ReMotion*, a fully automatic robotic proxy designed to explore a new way of supporting non-collocated open-space design activities. *ReMotion* combines a commodity body tracker (Kinect) to capture a user's location and orientation over a wide area with a minimally invasive wearable system (NeckFace) to capture facial expressions. Due to its omnidirectional platform, *ReMotion* embodiment can render a wide range of body movements. A formative evaluation indicated that our system enhances the sharing of attention and the sense of co-presence enabling seamless movement-in-space during a design review task.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Robotic Embodiment; Telepresence Robot; Remote Collaboration;

ACM Reference Format:
Mose Sakashita, Ruidong Zhang, Xiaoyi Li, Hyunju Kim, Michael Russo, Cheng Zhang, Malte F. Jung, François Guimbretière. 2023. ReMotion: Supporting Remote Collaboration in Open Space with Automatic Robotic Embodiment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3544548.3580699

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580699

1 INTRODUCTION

Design activities, such as brainstorming or critique, often take place in open space settings around physical artifacts and whiteboards. In such settings, designers move from one locus of attention to the next. In front of the board, they may move from side to side to look at different ideas presented on the board. They may also move away from the board and switch their attention to a table to review and discuss a physical model [30]. Because of this, understanding one another's body position and pose plays a crucial role in the group dynamic and joint understanding of the current design process [17]. Relative distances and orientation also help participants to manage coupling, whether they are working together or alone [17, 30, 65]. Head orientation is often used together with body movements to help people more accurately estimate partners' attention [43] or to send signals of agreement, disagreement, or back channelling [39]. In addition, pointing at surrounding objects in open-space is important to share the understanding of deictic reference [9, 17].

In co-located design practices, awareness of one another's attention and location is easily achieved due to the peripheral awareness through which participants can quickly gather information about the workspace and their co-worker [18, 59]. Once collaborators are geographically distributed, establishing mutual understanding between individuals quickly becomes challenging [3, 61].

Video-based telepresence systems for remote collaboration have been proposed to better afford mutual awareness among collaborators, but many of them have focused on more traditional meeting settings where participants are seated around a table [45, 46, 49]. While there have been remote collaboration systems that allow collaborators to move around in front of the board [2, 23, 28, 76], they are not designed to support room-scale workspaces where there are other task areas besides a board in a physical space.

Mobile telepresence robots can be useful to support remote interactions in room-scale workspaces during instructional or social settings [34, 35, 44], but the cognitive load required for such controls can be significantly distracting to a remote user when working on hands-on tasks [1, 48, 54, 55, 59]. While automatic control systems for telepresence robots have been previously explored and can improve peripheral awareness in a hands-on task, they only support around-table tasks [1, 48, 59]. Thus, these robots are not suited for open-space tasks where collaborators must engage in design activities while moving around in space.

In this paper, we present *ReMotion*, a novel system designed to support remote design in open space and to enable seamless movement-in-space through a robotic embodiment that automatically replicates the body and head movement of the remote collaborator as shown in Fig. 1. *ReMotion* eliminates typical control interfaces by constantly tracking users' body position and orientation. The captured body position and orientation are rendered through an omnidirectional platform to accommodate a wide range of motion, such as shuffling side to side. At the same time, head orientation and facial expressions are tracked using NeckFace [12], an unobtrusive wearable tracking system, and rendered through an articulated display similar to that of table-based systems [1, 48, 59]. Combining these two, *ReMotion* allows the remote user to move freely across a large area and provides a physical rendering of their locus of attention in the local setting. When used in a symmetrical setting, *ReMotion* can enable seamless end-to-end interactions in a setting such as design review or training sessions, allowing non-co-located collaborators to rely on peripheral awareness to understand the each other's intention in a way similar to a co-located setting. We believe that this novel approach can make moving-in-space interactions seamless and reduce the previously-reported discrepancy in the sense of presence and engagement between remote and local users [7, 54, 64].

We conducted a formative evaluation study to examine if *ReMotion* can assist remote collaborators in establishing a shared understanding of spatial relationships [9] for open space activity, as they would in face-to-face communication, by simulating human movement-in-space. We address a shortcoming of our system discovered during the evaluation through followed up implementation such as augmenting the face tracking system with an IMU to achieve a stable head animation. We conclude by presenting implication derived from the development and evaluation of *ReMotion*.

2 RELATED WORK

2.1 Frameworks for Remote Collaboration

Several frameworks have been presented to guide researchers in designing systems or tools for supporting remote collaboration. In this section, we introduce frameworks and observation studies that we use to identify the key elements to be shared across dif ferent locations for conducting open space activity. Gutwin et al. emphasizes the importance of attaining *workspace awareness*, that is awareness of how other people are interacting with the shared space [17]. Buxton introduced shared space as person space (verbal and facial cues), task space (where work appears), and reference space (body language to refer to the work), highlighting the importance of seamlessly integrating them to deliver a natural flow in distributed collaboration [9, 10]. Vertegaal identified "relative position" and "head orientation" as two of the requirements to more effectively assist joint attention in multiparty collaboration [71]. The integration of body and head orientation is used as information to perceive and estimate one another's attention [43]. The observations of whiteboard sessions conducted by Ju et al. revealed that collaborators switch formations, changing their position between a whiteboard and table [30]. The analysis indicated that not only the dynamic transitions among different areas but also movements around a board reflect the status of designers. For example, a collaborator steps forward to comment or initiate drawing and steps back to analyze and evaluate. Similar findings were reported with respect to the use of space at a table [65]. These studies highlight the importance of collaborators' body and head movements in relation to both objects around them and to other collaborators in space , on which we design and build our robotic proxy for open-space collaboration (Section 3).

2.2 Video-based Systems

A number of videoconferencing tools render a person space to enhance awareness of remote users by projecting life-sized images [45, 46, 49, 73]. VideoWhiteboard [66] attempts to merge the person and task spaces by projecting a shadow of the remote participant over a shared task space. Also, ClearBoard [28] extends this idea on

overlaying shared task space over a person's space for whiteboard interactions. Other whiteboard-type systems have been proposed to allow collaborators to move around in front of the board [2, 23, 76]. More recent work explores approaches that use depth-cameras and projection techniques to support larger-scale interactions. For example, Room2Room [51] projects a life-sized person on furniture in a room-scale environment. Beck et al. presented a group-to-group telepresence [5] that projects people captured by depth cameras on a large screen to assist collaboration in virtual environments. Buxton introduced a design principle of preserving spatial context to design a videoconferening system [9]. For example, Hydra [61] renders both a person and task space in such a way that maintains spatial relations using a camera and monitor pair. The limitation of these systems is that only movements under the camera image can be observed. Our system takes a similar approach but extends also to support room-scale interactions, where there are some other task areas (e.g., tables or shelves) besides a monitor. In this context, the robot's mobility enhances the peripheral awareness that helps construct a shared physical frame of reference.

2.3 Robotic Embodiment

2.3.1 Telepresence Robots. Since the introduction of Paulos and Canny's Personal Roving Presence [50], similar mobile robotic systems have been explored as means to collaborate remotely in a wide space [4, 58]. A laser pointer is sometimes attached to provide a pointing feature [34, 50]. These mobile platforms have been tested for use in a conference [44] or in a workspace [35] and are a practical means of physically rendering a person space to a shared space. Other robotic proxies take the form of an articulated display in which a face is projected on its display and rotates based on where a remote user is looking [26, 74]. While both types of robotic embodiment confer the benefits of physical representation on a sense of presence and peripheral awareness [59], one of the main drawbacks of these systems in supporting design activities is the relatively high cognitive load required to control them [48, 54, 59, 70].

2.3.2 Auto Kinectic Displays. Several systems addressed the problem of the cognitive load by providing automatic control for articulated displays [1, 48, 63]. MeBot [1] is a robotic proxy that has an articulated display and arms for head and hand gestures operated via head movements and joysticks. Similarly, Sirkin et al. explored implicit control design by giving a remote operator a panoramic view of the environment and mapping the head movement to the robot [63]. However, these interfaces still require a user to be seated in front of a laptop. MMSpace [48] supports face-to-face conversations by automatically controlling the animated monitor to convey who is talking to whom. RemoteCoDe [59] further expands this type of auto control interface to support hands-on design tasks that require various task areas by affording peripheral awareness through the robot's movements. These two systems are examples of systems that share frames of reference in shared spaces using spatial context [9]. However, due to the lack of mobility, these kinectic displays only support relatively small setups where collaborators are seated around a table.

2.3.3 Auto Mobile Proxies. Mobile telepresence robots are promising for facilitating collaboration in open spaces. Previous research on mobile robots has suggested systems that afford semi-automatic navigation [38, 57]. These systems can simplify navigation by only requiring users to specify a destination. Other work has explored immersive interfaces using a head-mounted display (HMD) [22, 29, 33, 75]. For example, VROOM-ing [29] has a 360 camera where a VR user observes the remote space and uses joysticks to control Beam. Eye-gaze has been used to allow those who have motor disabilities to have control in VR [75]. However, for both types of control systems, users still must play an active role either by interacting with a GUI or by monitoring a remote view. This could constrain a remote user's access to their local space and freedom to move around. Our work attempts to redesign a typical mobile robotic proxy to support wider open-space activity, allowing both remote and local collaborators to walk around without direct involvement in controlling the robot.

2.4 Mixed Reality Collaboration

MR systems have great potential for remote collaboration as they can overlay the virtual presence of remote collaborators over physical objects in a see-through headset [52, 53, 68]. More recent VR based solutions such as Holoportation [47] support room-scale interactions by rendering an entire space that is captured with multiple depth cameras. While these solutions could be a possible approach to support open-space collaboration by rendering all types of spaces [9], peripheral awareness could be influenced by the limited field of view that many HMDs have. We also note that some people prefer not to wear a headset. In those situations, an alternative way to render the presence of a remote person would be beneficial.

3 REMOTION DESIGN

Typical interactions during design critique, training sessions, or brainstorming sessions encompass a complex combination of verbal and non-verbal communications. Collaborators often stand side by side near a whiteboard to discuss the information displayed on the board. They change their body position to take a closer look

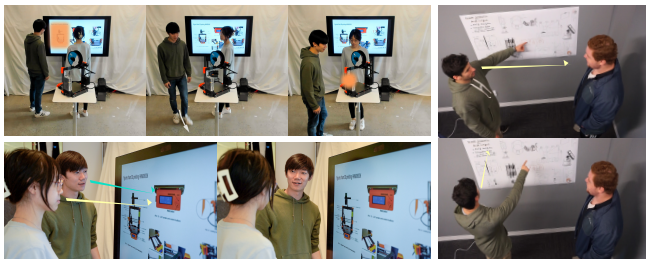

Figure 2: Typical interactions during a design review session. Two collaborators are changing their body location to switch attention between the board and table (Top Left) and are also using their head rotation to look at specific parts of the board or initiate face-to-face conversations (Bottom Left). Pointing and head direction can be combined to contrast or relate different design aspects (Right).

at different areas of the board or to shift their focus to discuss various elements of the design with their teammates. In addition, their work space is not typically limited to the whiteboard but is extended to wider areas of the space. For example, the space may have a table where models or even samples of different materials are presented (Fig. 2). Collaborators move around the space from one area to another depending on the focus of interest to achieve their design goal [30]. In such a setting, collaborators rely extensively on body cues to enable them to understand the focus of others in the room, coordinate actions, establish joint attention [14], and manage coupling [17].

This peripheral awareness is not only conveyed through the location and orientation of a collaborator's body; the direction people's heads are facing is also vital to afford accurate estimation of the focus of each person's attention [43]. For example, in front of the board, users may rotate their heads to indicate whether they are looking at the board or at their partner to initiate face-to-face conversations, as shown in Fig. 2. While looking at a model on a table, turning one's head towards a nearby display may indicate that one is looking for more information to cross-validate a possible idea. Being able to see facial expressions can also benefit understanding the emotional state or intent of the other person. Pointing gestures facilitate understanding of deictic reference in space [17] when explaining different parts of the design board, for example (Fig. 2). Although we have focused on gross motor movements thus far, we should acknowledge the importance of gaze to convey detailed attention and implicit intention or mental state [15]. This has been well established by other works [9, 10, 21, 28], and given our focus on free movement in open space, we decided to not include this aspect during our design process.

The key to success in designing a remote system for complex open-space workspace scenarios is to liberate users from the complexity of having to gauge where their collaborators are in the shared space and to allow them to understand each other implicitly, as if they were co-located. This can be achieved by rendering task, person, and reference spaces according to spatial context and relieving users from the cognitive burden of coordinating their workspaces explicitly [9].

From this analysis we believe that a system supporting remote collaboration in open space should:

- track the users in open space including their body and head movements, facial expressions, and pointing gestures;
- be able to reproduce body movements automatically that include complex human motions seen in typical interactions (e.g., lateral movements);
- be able to offer accurate rendering of head and facial animations in conjunction with body movements;

In this paper, we present *ReMotion*, a system supporting remote collaboration in open space through a mobile robotic embodiment as shown in Fig. 1. The system is capable of tracking a collaborator's movement using a kinect sensor and rendering their movement using an omnidirectional platform well-suited to mimic complex human motion. To track head movement relative to the body as well as facial expressions, we use NeckFace [12], an unobtrusive wearable face tracking system. The information gathered by the NeckFace system is rendered on an articulated display as a front face shot. The combination of both tracking systems frees the remote user from the workload induced by controlling a remote avatar. At the same time, the flexibility of our embodiment platform permits the support of a wide variety of natural interactions, enabling collaborators to seamlessly move and change focus of attention in space. To enable users to be implicitly aware of shared space, we designed a symmetric system where both remote and local people have similar configurations, so that both parties can share the same frame of reference in space with spatial consistency and understand each other's locus of attention through embodied movements. This symmetric setup is a common practice for collaboration system research as it allows for easy observation of how both collaborators use shared space [23, 45, 46, 48, 66]. While this setup can constrain the flexibility of work environment and limit the range of scenarios it can support, the proposed prototype can still be useful for training people to use a certain machine, demonstrating an artifact to a client, or conducting design reviews. We briefly discuss how our novel approach can be extended further for asymmetric scenarios in Section 6.3.

3.1 Rendering Body Movements

3.1.1 Tracking Body Movements. In design activities, the size of a work space varies depending on the design task collaborators are working on. Even though simple assembly tasks can be done within a small area around a table, brainstorming or design review tasks usually require a significantly larger work area.

Among the many tracking systems available, we decided to adopt the Microsoft Kinect Azure system as it is readily available, offers a body tracking SDK [40] to get the local position and orientation of the user's body, and does not require users to wear tracking devices. Furthermore, several Kinect Azures can be easily synchronized to cover more space as needed. *SPINE_CHEST* joint provided by the Kinect body tracking SDK was used as a body position.

3.1.2 Mobile Proxy for Supporting Wide Range of Movements. As noted before, human movements in collaborative settings are often

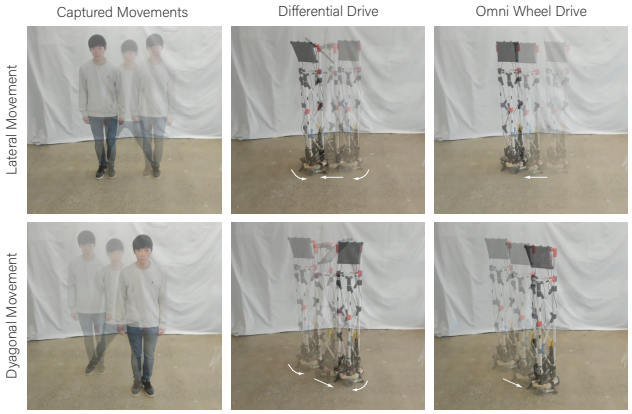

Figure 3: Comparison of omnidirectional drive and differential drive controls in replaying lateral and diagonal movements. Note that differential drive was simulated using omnidirectional wheels by adding similar constraints.

complex. For example, people do lateral shuffling to shift their focus from one area of the design board to another. They also turn their body from a board to a table and, at almost the same time, begin moving toward the table.

In our early prototypes, we explored a differential wheeled robot, as differential drive systems are used for typical telepresence robots such as Beam [4] and Double [58]. Unfortunately, this setting limits the reproduction of human movements such as shuffling. Due to these kinetic constraints, the robot may deviate significantly from its ideal trajectory, creating inconsistency in position and rotation between the remote participant and the robot as shown in Fig. 3 center column. This may also create disparity in timing. These inconsistencies are problematic in collaborative tasks, since a robot would be incapable of rendering the state of the user's location and locus of attention in a timely manner. In contrast to two-wheeled robots, omnidirectional wheels enable moving a mobile robot to any direction regardless of the orientation of the robot [11]. The right column of Fig. 3 demonstrates how our omnidirecitonal platform is able to tracking human movements in realtime with limited deviation. This exploration led us to adapt this type of drive system into our mobile platform.

One of the drawbacks of the omnidirectional platform is that it is often subject to drift. Taking scalability into account, we decided to use an inside-out system in which the camera is mounted on top of the robot observing an array of ArUco markers attached to the ceiling (see Fig. 1). With this inside-out setting, extending the tracking area only requires printing a larger set of markers. Using this tracking information, we implemented a simple PID feedback loop to be sure that the robot quickly catches up to the position and orientation inputs provided by the user tracking data.

3.1.3 Mapping Movements between the User and Robot. In mapping process, the system assumes that both locations have the same size of working area as well as configurations of furniture (e.g., table or monitor). As the position of the Kinect and the height of the ceiling vary in each room, the system uses a homography transform to map coordinates between the two spaces. When a user walks outside the boundary, the target position snaps to the closest position on the boundary. This approach prevents the robot from moving beyond the space in which its movements can be tracked.

To simplify calibration, we marked four corners of the working boundary of our system. During the calibration process, and for each corner, we record: 1) the position provided by the Kinect for this user and 2) the corresponding position provided by the robot tracking system. We compute the matrix that can map from the ArUco tracking coordinates to that of the Kinect to obtain both positions in the same coordinate system. We then calculate the orientation difference and the vector from the robot's current position to the user's position, which is used to control the omni robot.

3.2 Rendering Head Movements and Facial Expressions

While the Kinect body tracking SDK [40] provides head orientation, we discovered that this information is often quite noisy. Moreover, our system also must be capable of generating a front shot of the face to be rendered on an articulated display attached to the proxy

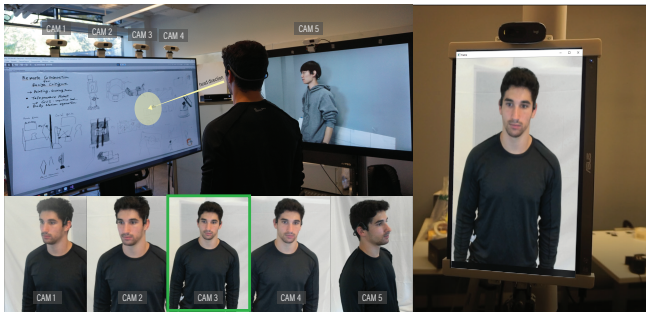

Figure 4: We used multiple servo-controlled cameras in our initial approach for rendering a front face, but then discarded them in our final design.

in order to avoid users misinterpreting their partner's head rotation as the robot moves [31, 59]. Our first approach to create a front shot was to use multiple servo-controlled cameras, each camera tracking the position of the user's head (Fig. 4). Depending on the orientation of the head, the system selected the feed closest to a front shot. This approach was ineffective because the frequent transitions among different cameras were noticeable and distracting even after applying some fade-in-out transition effects. Scaling up was also dif cult since the number of cameras required for open-space interactions can be extremely large to cover every task area. Secondly, we considered using an iPhone TrueDepth camera placed on a chest holder. This tracking method has been used for multi-monitor interactions [72] as well as a telepresence robot [59]. However, having a body mount camera can be bulky and highly distracting for conducting design tasks while walking around.

Instead we decided to use an avatar which reflects the realtime facial expressions of the remote collaborator. For this, we put to use the compact NeckFace system [12]. *NeckFace* predicts facial expressions as well as head rotation using two IR cameras worn on a shoulder pad (Fig. 5 Left), which allows hands-free capture suited for design activity. The NeckFace is able to provide roll, pitch, and yaw rotations as well as 52 blendshapes that the iOS face tracking SDK relies on to depict complex facial expressions (Fig. 5 Right). The NeckFace prototype works reliably under controlled conditions (e.g., lighting) and provides updates at approx. 13 FPS [12]. While our current rendering system is far from creating video-like rendering, high quality photorealistic avatars such as those generated through recent NeRF based approaches [16, 37] can be combined

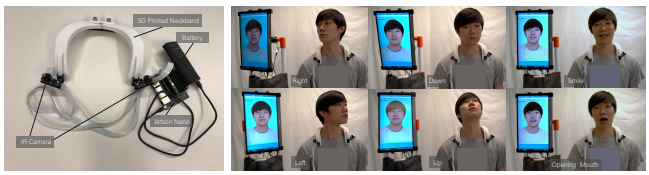

Figure 5: The NeckFace system. We show the shoulder pad used to hold the cameras and the illumination sources as well as the computing module (Left). Several examples of facial poses and expressions and how they are rendered on our articulated display (Right).

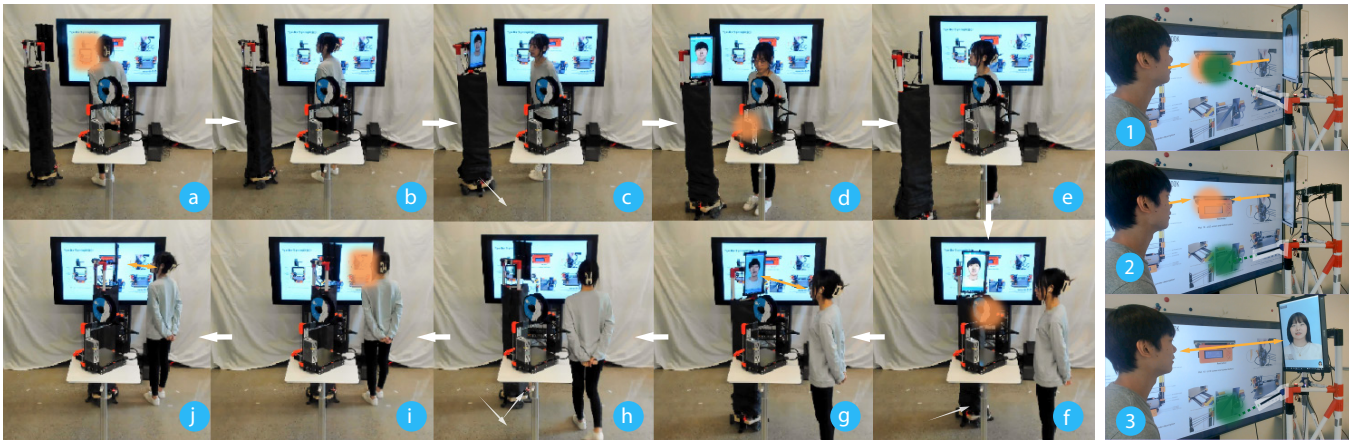

Figure 6: Demonstrations of ReMotion system in action. The system enables seamless moving-in-space interactions through the mobile robot between the design board and artifact (a-f). The pointing arm and articulated display facilitate smooth discussion around the board (1-3).

into our system in future design. Note that we did not include this in our study due to its focus and the sensitivity of the NeckFace system (See Section 4.1).

3.2.1 Kinetic Display for Rendering Head. Our system uses an articulated display similar to RemoteCoDe [59], that is, a 2-DOF kinetic robot with a monitor which acts as a remote user's head. The pan and tilt rotations are provided by the neckband system and used to actuate the articulated monitor in two axes. A face avatar is generated along with 52 blendshapes through a single image using a service called AvatarSDK [60]. We then render the face of the avatar from the front to obtain a front shot to display on the articulated monitor. Fig. 5 shows examples of mapping. The height of a robot bears significance as it has effects on the dynamics of communications [56], but for this first prototype, the height of the display is fixed, set at approx. 170cm.

Having an articulated display on top of the robot can create some oscillations as the robot moves quickly when tracking a user. This oscillation could affect the user's perception of the robot and could also impact the accuracy of robot tracking. To minimize the oscillation of the robot, we moved the center of mass of the robot as close as possible to the base. The triangular truss structure built with PCV pipes also helped the robot to be more stiff and stable (Fig. 1).

3.3 Visualization of Hand Pointing

We explored several means to support pointing. We started with a system similar to that of RemoteCoDe [59] by showing a task camera stream on an iPad controlling a remote pointer. This proved very difficult to use in practice for complex 3D objects. We also considered placing a laser pointer on the robot itself, but allowing the visualization of pointing to remain stable is difficult on a moving platform because a wrongly displayed laser point can be highly misleading even if there is a small offset from an actual position. Further, hand tracking in space is challenging in our target setting.

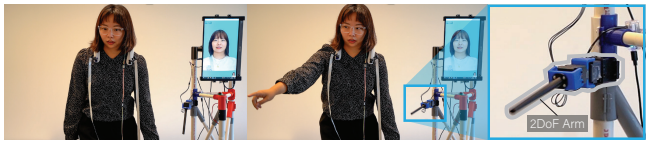

Figure 7: The system tracks the user's arm direction and visualizes hand pointing direction (Left). The pointing is rendered through the 2 DoF pointing arm attached to the mobile robot (Right).

Our final solution was to add a small 2-DOF arm to the side of the mobile platform, acting as one of the arms as shown in Fig. 7. To prevent the robotic arm from appearing too humanlike and causing a sense of unease in viewers, we designed the arm with low DoFs, while still allowing it to perform the essential task of indicating general pointing directions. We use the skeleton captured by Kinect to track the pointing motion. The pan angle was approximated using *HAND_RIGHT*, *SHOULDER_RIGHT*, and *NECK* joints, and the tilt angle was using *HAND_RIGHT*, *NECK*, and *SPINE_BASE* provided by the body tracking SDK. This part of the design was not included during the study in Section 4 as it was not implemented at the time.

3.4 ReMotion in Action

Combining all the features discussed above, *ReMotion* is able to render visual cues of where a remote collaborator is in the space and what the collaborator is paying attention to via a robotic embodiment. Fig. 6 shows a typical flow of interactions using *ReMotion*. A remote and a local collaborator are reviewing materials that present the key features of a 3D printer. They first review the information on the shared digital board and discuss it (a,b). The movements of the robot clearly highlight the intention of the remote user. As the robot departs from the display toward the table, the local user sees that the focus is changing (c,d,e). While discussing the printer, the robot body is more or less static, but the movement of the display

clarifies when the remote user is focusing on the display and when he is looking for a face to face interaction (g,f). Finally, the remote user moves to another part of the display, where the discussion continues (h,i,j). When the discussion happens around the display, the remote collaborator points at a specific part of the board (1), changes their pointing direction to make a reference to a different part of the design (2), and establishes a face-to-face discussion (3) through the combination of the pointer arm and articulated display.

3.5 Implementation

3.5.1 Software. We used the Unity engine to implement the *ReMotion* system, with the addition of python scripts to interface with our robotic embodiment. Mirror Networking [41] for Unity helped us with distributed processing and communicating between different client applications. We deployed a server application through which all the client applications exchange information such as movements, robot commands, and streaming images of an avatar. We used Zoom [27] for audio transmission.

3.5.2 Hardware. As shown in Fig. 1, the mobile platform includes the omnidirectional wheel system powered by its own battery; the articulated display with Dynamixel motors and controller; a portable battery powering the rest of the system; a portable Windows computer for rendering a face avatar on the display, managing motor controllers, and for processing the tracking information provided by the camera on top of the robot. The robot also includes an emergency button to stop the omnidirectional wheels and a speaker to share the voice of the remote partner.

4 FORMATIVE EVALUATION

The goal of our evaluation is to investigate how our automatic embodiment prototype af interactions in open space activity by embodying seamless movements in space. We seek to examine if our prototype can help users create a physical frame of reference in a shared open space [9] and help them understand their collaborators' cues by the movements of the robot. To achieve this goal, we measure the amount of shared attention among collaborators within the workspace, observe users' movements and behavior during the task, and collect qualitative feedback to understand participants' perceptions of interacting with a collaborator through the embodied robot.

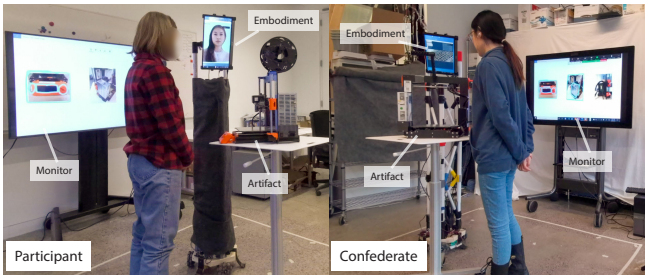

Figure 8: The ReMotion setup used for our user study showing the two symmetrical settings. There is a table used to place the artifact and a display monitor used to present three aspects of the artifact.

4.1 Experimental Setup

Our goal of designing the experimental setup and task was to recreate a studio-like setting in which participants have several areas of interest in an open space. In a typical design studio, physical objects are scattered around the space. There may be a board to put information on and a table used to present physical artifacts. To simulate similar interactions described in Section 3.4, we created the setting presented in Fig. 8. This setting included a table and a large monitor. We picked two 3D printers (Prusa MK3 or Mini+) as physical artifacts to work with, as these offer different aspects to discuss and view from various angles. Our initial intent was to test the system with a pair of participants. This proved impractical given the length of time that both data collection and training require.

To maintain consistency in interactions between participants, the participants interacted with a remote confederate who was trained to describe the key parts of the physical artifact on a table. Having a trained confederate work with participants is a common practice for remote collaboration studies [33, 54, 56, 63]. The core of the confederate training was to ensure they consistently followed the study protocol for each participant (e.g., annotations on a digital board) and be mindful of our robot's limited speed.

We then tested using NeckFace only on a confederate. We discovered that NeckFace was sensitive to lighting conditions for detecting head rotations. Pilots showed that erroneous rendering of the remote participants' head through the display was very distracting and could invalidate our data on the use of space due to its large movement. Given the focus of our preliminary evaluation on body location and orientation in space and its impact on attention sharing, we decided to disable the NeckFace feature. We revisit and explore alternative head tracking later in the paper (see Section 6.1).

To compensate for the missing feature, the confederate tried to align the body and head angles such that participants could still read their attention direction during the task.

4.2 Task

In the task, participants completed a session in which the confederate used information on the board to explain three main aspects of a 3D printer (either Prusa MK3 or Prusa Mini+), drawing some annotations using the touch display. Participants were then asked to find specific parts on the actual printer in their room. They were allowed to move around the space as they wished during the session. Explanations of each 3D printer were created in order to ensure that participants had an opportunity to locate certain components on the printer each time the confederate explained the component.

We used a within-subjects design with the following two conditions:

(1) ReMotion – Participants interacted with a robot mirroring the movements of the confederate. The participant was also embodied through a robot in the confederate's location. In each location, the user's position and rotation was tracked and rendered by the robot in the other space. This symmetrical setup was configured to understand if two people can mutually understand the position and attention of their partner through our robotic embodiment. The pointing arm was not included in this evaluation as it had not been implemented by the test.

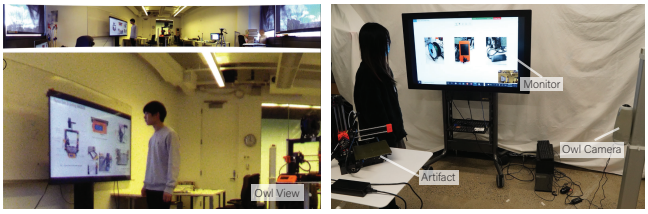

Figure 9: A typical view of Owl system that provides a wide angle and zoomed images of the room (Left) and setup for the Owl condition (Right).

(2) Meeting Owl Pro – Participants used a videoconferencing tool called *Meeting Owl Pro*. We chose this system as a reference because it is one of the latest available videoconferencing tools. It provides a wide angle view as well as a presenter view that automatically tracks and zooms in on the user's location Fig. 9. This feature is well adapted to people moving around in an open space. Both the participant and the confederate had one Owl device in their room to share their view with one another, which was displayed on a large monitor in their room directly below the presentation pictures. We placed the Owl device on the opposite side of the study location as there was a lighting issue.

4.3 Procedure

After participants completed an informed consent form, they were introduced to the first condition. The order of the conditions was counterbalanced. They received the explanation that they would be interacting with a remote partner in another room in realtime and that their partner had a similar setting as the participant's experiment room (the table and display). They were told not to step beyond the square boundary (150cm x 105cm) as this was where the system could track.

For each condition, participants had 1-2 minutes of a practice session with a confederate where they learned how they could expect the system to behave during the task. In the practice session for both conditions, we asked participants to look at a table or desk and move to each area. Through the practice, in the Owl condition, participants observed that the Owl system tracked their position and displayed both zoom-in views of them and an overview of the entire space. In the ReMotion condition, participants observed that the ReMotion robot mimicked the movements of their remote partner's body in the local space and those of themselves in the remote space. We turned on a remote camera view so that participants could see how the robot in the other room responded, based on their movements. We turned off the camera view after the practice.

The main task involving the printer evaluation was completed twice, once for each condition. Immediately after each condition, participants completed the corresponding evaluation questionnaires. The order of conditions and the type of printer (MK3 vs. Mini+) evaluated in each condition was counterbalanced across participants.

After completing both conditions, participants filled out a post-study questionnaire. The experimenter then conducted a few minutes of an in-person interview with the participant to follow up on their answers on the questionnaire and gather additional qualitative feedback on the system.

4.4 Participants

We recruited 13 participants (7 female, 5 male, 1 preferred not to specify). Their ages ranged from 18 to 34. Despite the partial counter-balancing due to the uneven number, we included all the participants in our analysis as the results of statistical significance were the same with or without the last participant except for the preference measure (see Section 5.4). Participants received $20 for their participation in the hour-long study.

4.5 Measures

Several objective and subjective measures were used to evaluate the interactions during the collaboration task.

4.5.1 Presence. We used a similar method as in the study by Rae et al. [54] to measure the sense of presence. For this measure, participants drew one circle for themselves and another circle for a remote partner on a printed map to indicate whether they felt that they and their remote partner were working in the same space or in a different space (see Fig. 10). Our hypothesis was that the ReMotion condition would increase participants' feelings of being together in the same environment.

We also used items from the widely accepted Networked Minds Measure of Social Presence [8] and adapted them to fit our task scenario. This measure assessed social presence in terms of co-presence, behavioral interdependence, and psychological involvement.

4.5.2 Joint Attention. Establishing joint attention or shared attention efficiently benefits design collaboration. To measure the amount of joint attention during the interaction, we initially were planning to use the data provided by the system, but unfortunately an error in the logging system prevented us from doing so. Instead, we conducted a video analysis of the recordings during the tasks. We measured the duration in seconds of when both the confederate and the participant attend to the same area (either the display or table area). We then calculated the percentage of time when both the participant and the confederate had the same focus. In the video coding process, a researcher coded all the sessions as a main coder and a hired coder recoded half the videos randomly selected from the recordings. We then assessed inter-coder reliability using intra-class correlation coefficient (ICC). The single ICC measure indicated good reliability (.84, 95% Confidence Interval (CI) [.54, .95]) [32].

As a subjective measure, we also asked to what extent the participants felt they shared the same attention with the remote collaborator in each condition.

4.5.3 Switch of Attention. Through video analysis, we also measured how many times the participants changed their attention between the display and table areas during the task. Our hypothesis is that ReMotion offers peripheral awareness of the other collaborator, as reported for a kinectic display robot [59], and therefore, participants do not find it necessary to switch their focus away from the task as frequently as in the Owl condition in order to know the state of the other collaborator. The single ICC measure indicated excellent reliability (.93, 95% CI [0.77, 0.98]) [32].

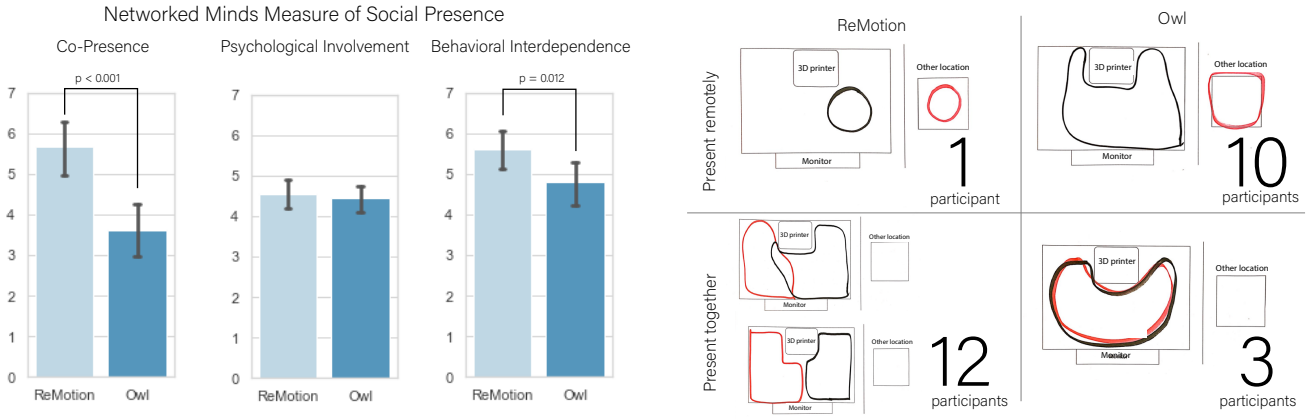

Figure 10: Results of Networked Minds Measures of Social Presence (left). Typical set of drawings we collected to represent where participants felt the different actors were located (Right). We cluster similar drawings and show the number of instances in each cluster. The vast majority of participants placed both actors in the local setting when using the ReMotion system.

4.5.4 Positioning and Distance. We recorded the participants' and confederate's positioning of their bodies in the room using Kinect to understand how differently participants used their workspace when using ReMotion versus Owl. One participant's data was not recorded properly and was omitted. Using the recorded movements, we then calculated the distance between the two collaborators as it was able to help us understand how participants shared the space. We also conducted a video analysis to observe how many participants reacted to the change of the confederate's positions in ReMotion condition.

4.5.5 Other Measures. We also asked participants about their general preference between the two systems for remote collaboration (binary). We assessed subjective ratings of enjoyment of the interactions. Finally, we used NASA's Situation Awareness Rating Technique (SART) [67] to measure the quality of shared information and used NASA's Task Load Index (TLX) questionnaire [20] to compare the two conditions in terms of cognitive load required for understanding one's partner's location and attention.

5 RESULTS

5.1 Presence

We start by reporting the results from the Networked Minds Measure of Social Presence including Co-Presence, Psychological Engagement and Behavioral Interdependence (Fig. 10). We found that in the ReMotion condition participants reported a heightened sense of co-presence and behavioral interdependence ($t(12) = 8.05$, $p < 0.001$, and $t(12) = 2.98$, $p = 0.012$, respectively). However, we found no significant difference in psychological involvement ($t(12) = .84$, $p = .42$).

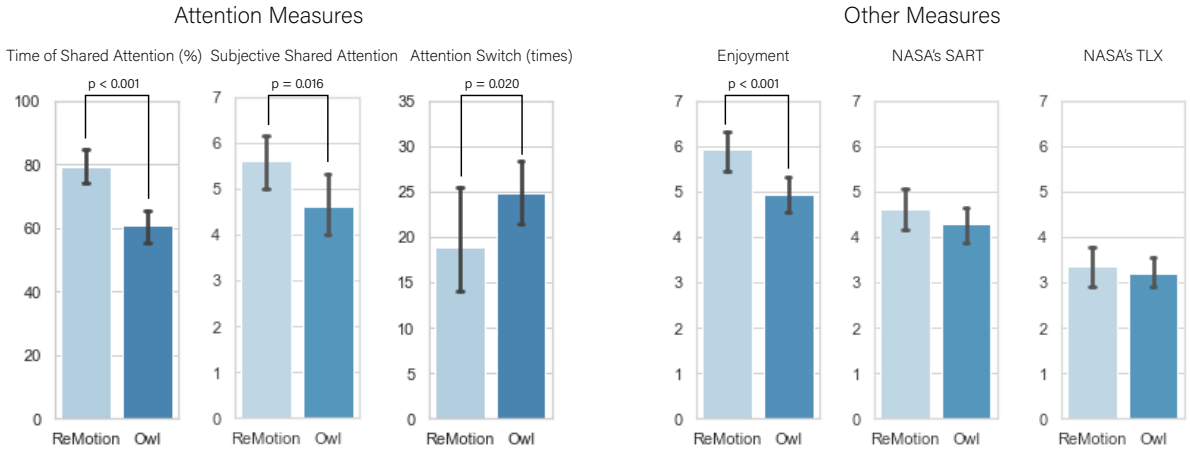

Figure 11: Results of attention measures: time of shared attention, subjective rating on shared attention, and attention switches (Left). Results of other measures: enjoyment, NASA SART, and NASA TLX (Right).

The strong finding regarding co-presence was further re-enforced by the drawing performed by the users to reflect their perception of presence. As shown in Fig. 10, 12 out of 13 participants drew both actors in the same room for the ReMotion condition but only 3 out of 13 did so for the Owl condition. This difference is statistically significant ($\tilde{\chi}^2(1, n = 13) = 6.54, p = 0.011$).

5.2 Joint Attention

We found that in the ReMotion condition the percentage of time when participants and confederate shared the same attention during the task was significantly higher ($t(11) = 7.60, p < 0.001$) as shown Fig. 11.

Participants also reported that they felt the sense of sharing the same attention with their confederate more strongly when using the ReMotion system than when using the Owl system ($t(12) = 2.92, p = 0.016$).

We found that participants switched their attention significantly less in the ReMotion condition than in the Owl condition ($t(11) = 2.72, p = 0.020$) as shown Fig. 11. This implies that participants could easily understand the confederate's intention through peripheral awareness without the need to look at the monitor, which is aligned with the known effect [59].

5.3 Positioning and Distance

Next we turn to the analysis of the relative position of the two actors. We found this analysis noteworthy as the distance maintained by the two actors reflects their perception of the interactions from a proxemics point of view.

We show heat maps of where the participants positioned their body in the two conditions in Fig. 12. In the ReMotion condition, the heat maps indicate that the participants were located mostly within one half of the workspace, reserving the other half of the space for the confederate, mirroring the findings of Fig. 10. In the Owl condition, participants occupied most of the space around the table, irrespective of the space occupied by the confederate.

The finding that participants were keeping their distance from their confederate is re-enforced by observing the distribution graph of instantaneous distance between the two actors shown on Fig. 12. The statistical results showed that participants distanced themselves from the confederate significantly more when using ReMotion than when using Owl ($t(11) = 6.55, p < 0.001$). Using a Levene test on all samples captured during our study, we found there is a significant difference in variances between the sample captured with ReMotion and with Owl ($F(9073,9303)=1715.38, p < 0.001$). In the ReMotion condition, we noticed that the participants were never closer than 41cm to the measured position of the confederate. We acknowledge that it is possible that the participants were concerned about their safety, but none of them mentioned this nor used the robot kill switch, though one participant mentioned that some interaction felt uncomfortably close (see Section 5.5). Through video analysis, we also found that all the participants changed their body position following the movements of the robotic proxy at least once during the task.

Together these results seem to indicate that ReMotion was able to elicit a strong sense of co-presence and enable the ability of sharing attention and space through the seamless movements of the proxy. It also implies that the participants reserved some space for the other collaborator during the task, as similarly seen in in-person interactions [19].

5.4 Other Measures

We conclude by presenting the final measures we captured during our experiment (see Fig. 11). Ten out of 13 participants preferred the ReMotion system to the Owl system ($\tilde{\chi}^2(1, n = 13) = 3.77, p = 0.052$), and, overall, participants enjoyed interacting with the ReMotion system more than interacting with the Owl system ($t(12) = 5.10, p$

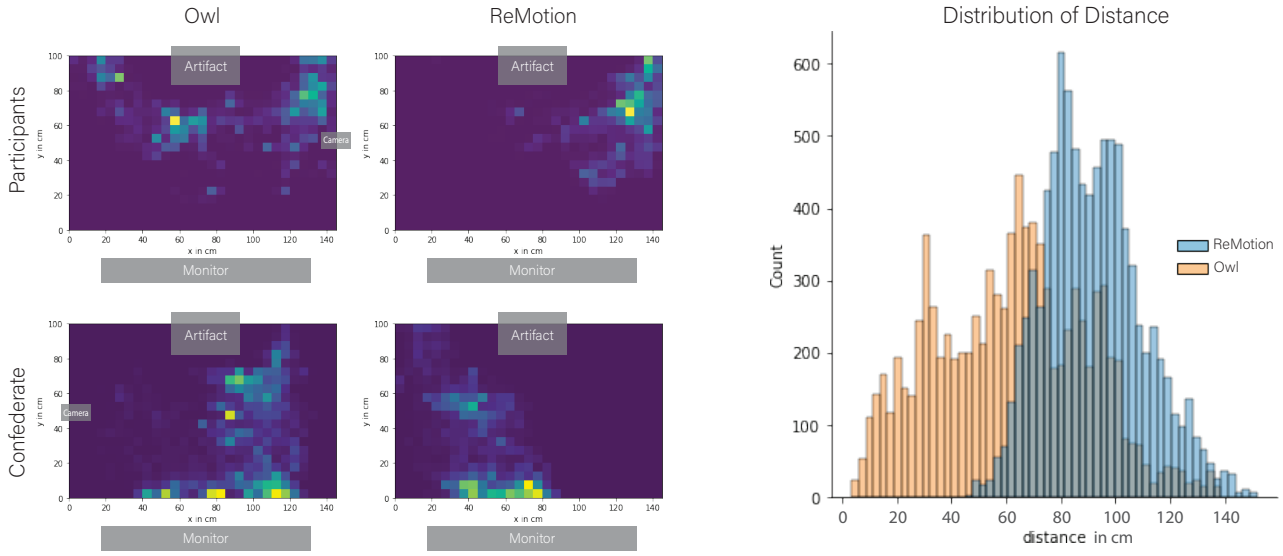

Figure 12: Heat maps of participants' and confederate' movements during the task in each condition (Left). Note that the room setup was mirrored for the Owl to optimize lighting. Distribution of distance in two conditions (Right).

< 0.001). However, these results are to be interpreted with caution, given the threat of demand characteristic. Neither the SART nor the TLX showed any significant results ($t(12) = 1.64$, $p = 0.13$ and $t(12) = 0.84$, $p = 0.42$, respectively).

5.5 Qualitative Feedback

The qualitative feedback from our survey supported our results on the presence measures. Here are some quotes from participants: *"I think it really brings me a sense of co-locating with each other inside the room (P8)."*, *"(I was) feeling actually being in the same room with my partner (P11)."*, and *"The interaction was engaging because my partner was with me in the same room (and) that allowed me to stay concentrated throughout (P12)."*.

Conversely, the results of our study indicate that the system does not have much of an impact in terms of their psychological involvement. Several participants commented on the inability to see an animated face on the display, for example, saying *"I was unable to see her facial expression to gain a better understanding of how she was receiving me (P2)."* Some also mentioned that being able to see their confederate's faces would have been better, *"Being able to see her face, or knowing that she could see mine, would have enhanced our synergy and helped us communicate better (P2)."* It is important to remember that technical difficulties prevented us from using NeckFace during our test. This part of the system would have provided a rendering of the face. More studies will be needed to investigate the real impact of the full system, including face rendering, in practice.

Participants also mentioned that the robot was supporting the peripheral awareness of their collaborator. Here are some quotes on gauging attention through the movements of the robot: *"... I felt that when my partner's robot was facing me she was also facing me and giving me her attention. When my partner turned away from me, I understood that as she was giving her presentation and paying attention to teaching me about the 3D printer (P4)."*, *"I looked at the display and the robot. When the robot moved or turned, I could see how the location and attention of my partner changed (P7)."* During the interview, four participants mentioned the usefulness of the peripheral vision in the context of knowing the other person's status. For example, P2 said *"I could actually focus on the screen cause she is still in my peripheral view. Kind of like a real person waking up to screen and walking back to the table... I didn't have to look more than one place at once. It was kind of all in my field of vision."*. Several people also mentioned the noise from the robot as an indicator of the partner's moving or changing attention, although few participants mentioned it as a distracting factor *"I think there are some noises when the robot moves, which are distracting a little (P1)."* In addition, one participant mentioned that they felt too close to the robot at times, saying *"There might have been times where I felt a bit uncomfortable with the proximity of the robot (i.e. it got a little too close at times), causing me to take a step back so it/she had more space (P2)."*

6 LIMITATIONS AND FUTURE WORK

6.1 Tracking System and Gaze Direction

In a studio room, there are scenarios where the Kinect sensor accidentally detects those who are not involved in a design activity.

Figure 13: Wearable glasses device with an IMU sensor for tracking head orientation (Left). Demonstration of the articulated display following the user's right, left, up, and down head movements (Right).

The system labels people detected by the sensor and maintains its focus on the user who is first selected as an "active" user. This allows the system to track the same user regardless of whether there are other people in the frame. If the system loses the tracking of the user, the robot stops until a user is re-selected. We also faced some issues with occlusion for skeleton tracking. For this, multiple Kinect devices can be combined to expand the workspace or deal with occlusion issues.

We also plan to explore an alternative wearable-based full body pose tracking system similar to a chest-band [25] or a wrist-mounted device [24, 36]. Adopting these wearable-based tracking systems would potentially improve the tracking performance under complex scenarios, would further enhance the interaction experience of our system.

Our system could also benefit from tracking and rendering eye movement to convey implicit gaze direction and social cues. Wearable eye tracking systems are becoming more common [69], but it is important to note that, for conveying accurate gaze direction through eye rendering, alternative ways to render a collaborator's face may be needed, such as face-shaped display [42] or 3D display [62], as a flat articulated display is known to cause Mona Lisa effect [31].

We noted during the study preparation that visualizing erroneous head rotations through a kinectic display has the potential to be distracting and misleading. To explore a solution to address this problem, we attached a 9 DoF IMU sensor (BNO055) to a pair of glasses that tracks the absolute head orientation, which can be combined with NeckFace used for facial tracking (Fig. 13 Left). The detected rotations were used for controlling the pan and tilt direction of the display. We used the body orientation captured with Kinect to obtain the relative head orientation to the body. Our test shows that the tracking is very reliable and it works well with NeckFace (Fig. 13 Right).

6.2 Enabling Physical Manipulation

The focus of our work is set primarily on enabling seamless movement-in-space. Thus, object manipulation was not implemented as a part of the system. There are many design activities that do not require object manipulation. In a design review session, for example, students present their projects on a design board. In this particular case, a shared digital board (as demonstrated in our system) will act as the shared task space. Architects often present their artefact without letting their clients manipulate the model. Also, in training scenarios, an expert can train beginners on how to use tools such as a 3D printer without moving it as demonstrated during our study.

In these cases, assuming symmetry between the two locations is a viable solution. Clearly object manipulation could be important, but we concluded that it was best left as future work given our focus. Some previous work in robotics or HRI communities has explored means to manipulate physical objects [6], and the integration of object manipulation to our approach could benefit remote collaboration.

6.3 Adapting to Asymmetric Scenarios

Although both remote and local collaborators can access physical representation of one another through a robot, our system assumes that each location has an identical setup in its layout and size. Practical settings often have different configurations of layout for physical objects (e.g., design board, tables, chairs) and space size. One possible solution to this is to remap human movements in one space to the mobile robot in the other space in such a way that the context is maintained, similarly to the approach adopted in RemoteCoDe [59] although mobility makes it more complicated. Alternatively, a VR interface could be a viable way to enable movement-in-space asymmetrically as the remote VR user could use limited space to move from one task area to another while being immersed in a remote environment. Looking further into the future, we envision a system in which the robot has enough autonomy to remap automatically and dynamically the intended movements of the remote user into acceptable and collision-free movement in the local setting.

6.4 Height of a Robot

The adjustable height of a robotic embodiment helps to convey whether a user is "sitting" or "standing" [13]. The study by Ju et al. demonstrated that some collaborators sit around a table to examine an artifact or listen to a speaker at a whiteboard, while others go to a whiteboard to elaborate or explore ideas [30]. Designers switch the status of sitting and standing when changing their focus area or picking up a tool. During the design process, we also noticed that people lean forward to take a close look at an object even if they are not seated. Enabling height variation would be a useful addition to help designers better coordinate their actions in collaborative design. A solution to support this could be to place the display on a linear translation table so that the height can be adapted as needed.

7 CONCLUSIONS

We propose a novel approach of enabling seamless movement-in-space interactions in open space through the design and implementation of *ReMotion*, designed to support remote open-space activities. *ReMotion* tracks body, head, and face motions to control automatically a mobile robotic proxy and reproduce the intentions of a remote collaborator in shared space. Our omnidirectional mobile platform has the ability to reproduce complex human movements. Our preliminary study showed that *ReMotion* can enhance the sense of presence through the movements of the robot and facilitate the sharing of attention among collaborators by affording peripheral awareness. It also revealed that the movements of a collaborator affect the participants' usage of the shared space.

ACKNOWLEDGMENTS

This research was supported by NSF Award IIS-1563705 and IIS-1925100, and the Nakajima Foundation. We would like to thank Tanzeem Choudhury for lending us Beam+ and Wil Thomason for helping us test Beam+ with rosbeam [1] during the early prototyping. We are thankful to itSeez3D for providing free license to use AvatarSDK [60] for research purposes. We would like to thank Rachel Lynne Witzig and Corinna E. Loeckenhoff for providing feedback to polish the paper writing.

REFERENCES

[1] S. O. Adalgeirsson and C. Breazeal. 2010. MeBot: A robotic platform for socially embodied telepresence. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 15–22. https://doi.org/10.1109/HRI.2010.5453272

[2] Ignacio Avellino, Cédric Fleury, Wendy E. Mackay, and Michel Beaudouin-Lafon. 2017. CamRay: Camera Arrays Support Remote Collaboration on Wall-Sized Displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6718–6729. https://doi.org/10.1145/3025453.3025604

[3] Michael Baker, Tia Hansen, Richard Joiner, and David Traum. 1999. The Role Of Grounding In Collaborative Learning Tasks.

[4] Beam. 2023. Suitable Technologies Inc. Retrieved Feb 6, 2023 from https://suitabletech.com

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

[6] Michael Beetz, Freek Stulp, Piotr Esden-Tempski, Andreas Fedrizzi, Ulrich Klank, Ingo Kresse, Alexis Maldonado, and Federico Ruiz-Ugalde. 2010. Generality and legibility in mobile manipulation. *Autonomous Robots* 28 (2010), 21–44.

[7] Jacob T. Biehl, Daniel Avrahami, and Anthony Dunnigan. 2015. Not Really There: Understanding Embodied Communication Affordances in Team Perception and Participation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. ACM, New York, NY, USA, 1567–1575. https://doi.org/10.1145/2675133.2675220

[8] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*. 1–9.

[9] Bill Buxton. 2009. *Mediaspace – Meaningspace – Meetingspace*. Springer London, London, 217–231. https://doi.org/10.1007/978-1-84882-483-6_13

[10] William A. S. Buxton. 1992. Telepresence: Integrating Shared Task and Person Spaces. In *Proceedings of the Conference on Graphics Interface '92* (Vancouver, British Columbia, Canada). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 123–129.

[11] G. Campion, G. Bastin, and B. Dandrea-Novel. 1996. Structural properties and classification of kinematic and dynamic models of wheeled mobile robots. *IEEE Transactions on Robotics and Automation* 12, 1 (1996), 47–62. https://doi.org/10.1109/70.481750

[12] Tuochao Chen, Yaxuan Li, Songyun Tao, Hyunchul Lim, Mose Sakashita, Ruidong Zhang, Francois Guimbretiere, and Cheng Zhang. 2021. NeckFace: Continuously Tracking Full Facial Expressions on Neck-mounted Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5 (06 2021), 1–31. https://doi.org/10.1145/3463511

[13] Munjal Desai, Katherine M. Tsui, Holly A. Yanco, and Chris Uhlik. 2011. Essential features of telepresence robots. In *2011 IEEE Conference on Technologies for Practical Robot Applications*. 15–20. https://doi.org/10.1109/TEPRA.2011.5753474

[14] N.J. Emery. 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews* 24, 6 (2000), 581–604. https://doi.org/10.1016/S0149-7634(00)00025-7

[15] Chris D Frith and Uta Frith. 2008. Implicit and explicit processes in social cognition. *Neuron* 60, 3 (2008), 503–510.

[16] Guy Gafni, Justus Thies, Michael Zollhofer, and Matthias Niessner. 2021. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8649–8658.

[17] C. Gutwin and S. Greenberg. 2004. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW)* 11 (2004), 411–446.

[18] Carl Gutwin, Saul Greenberg, and Mark Roseman. 1996. Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation. In *People and Computers XI*, Martina Angela Sasse, R. Jim Cunningham, and Russel L. Winder (Eds.). Springer London, London, 281–298.

[1] https://github.com/people-robots/rosbeam

[19] Edmund T Hall and Edward Twitchell Hall. 1966. *The hidden dimension*. Vol. 609. Anchor.

[20] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload* 1, 3 (1988), 139–183.

[21] Zhenyi He, Keru Wang, Brandon Yushan Feng, Ruofei Du, and Ken Perlin. 2021. GazeChat: Enhancing Virtual Conferences with Gaze-Aware 3D Photos. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 769–782. https://doi.org/10.1145/3472749.3474785

[22] Yasamin Heshmat, Brennan Jones, Xiaoxuan Xiong, Carman Neustaedter, Anthony Tang, Bernhard E. Riecke, and Lillian Yang. 2018. Geocaching with a Beam: Shared Outdoor Activities through a Telepresence Robot with 360 Degree Viewing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173933

[23] Keita Higuchi, Yinpeng Chen, Philip A. Chou, Zhengyou Zhang, and Zicheng Liu. 2015. ImmerseBoard: Immersive Telepresence Experience Using a Digital Whiteboard. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 2383–2392. https://doi.org/10.1145/2702123.2702160

[24] Ryosuke Hori, Ryo Hachiuma, Hideo Saito, Mariko Isogawa, and Dan Mikami. 2021. Silhouette-Based Synthetic Data Generation For 3D Human Pose Estimation With A Single Wrist-Mounted 360° Camera. In *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1304–1308.

[25] Dong-Hyun Hwang, Kohei Aso, Ye Yuan, Kris Kitani, and Hideki Koike. 2020. Monoeye: Multimodal human motion capture system using a single ultra-wide fisheye camera. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 98–111.

[26] Xandex Inc. 2023. KUBI Telepresence Robot. Retrieved Feb 6, 2023 from https://kubiconnect.com/

[27] Zoom Video Communications Inc. 2023. One platform to connect. Retrieved Feb 6, 2023 from https://zoom.us

[28] Hiroshi Ishii and Minoru Kobayashi. 1992. ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 525–532. https://doi.org/10.1145/142750.142977

[29] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. 2021. Belonging There: VROOM-Ing into the Uncanny Valley of XR Telepresence. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 59 (apr 2021), 31 pages. https://doi.org/10.1145/3449133

[30] Wendy Ju, Lawrence Neeley, Terry Winograd, and Larry Leifer. 2006. *Thinking with Erasable Ink: Ad-hoc Whiteboard Use in Collaborative Design*. Technical Report.

[31] Ikkaku Kawaguchi, Hideaki Kuzuoka, and Yusuke Suzuki. 2015. Study on Gaze Direction Perception of Face Image Displayed on Rotatable Flat Display. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 1729–1737. https://doi.org/10.1145/2702123.2702369

[32] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15, 2 (2016), 155–163.

[33] Sven Kratz and Fred Rabelo Ferriera. 2016. Immersed remotely: Evaluating the use of Head Mounted Devices for remote collaboration in robotic telepresence. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 638–645. https://doi.org/10.1109/ROMAN.2016.7745185

[34] Hideaki Kuzuoka, Shinya Oyama, Keiichi Yamazaki, Kenji Suzuki, and Mamoru Mitsuishi. 2000. GestureMan: A Mobile Robot That Embodies a Remote Instructor's Actions. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 155–162. https://doi.org/10.1145/358916.358986

[35] Min Kyung Lee and Leila Takayama. 2011. "Now, I Have a Body": Uses and Social Norms for Mobile Remote Presence in the Workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 33–42. https://doi.org/10.1145/1978942.1978950

[36] Hyunchul Lim, Yaxuan Li, Matthew Dressa, Fang Hu, Jae Hoon Kim, Ruidong Zhang, and Cheng Zhang. 2022. BodyTrak: Inferring Full-Body Poses from Body Silhouettes Using a Miniature Camera on a Wristband. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 154 (sep 2022), 21 pages. https://doi.org/10.1145/3552312

[37] Lingjie Liu, Marc Habermann, Viktor Rudnev, Kripasindhu Sarkar, Jiatao Gu, and Christian Theobalt. 2021. Neural Actor: Neural Free-View Synthesis of Human Actors with Pose Control. *ACM Trans. Graph.* 40, 6, Article 219 (dec 2021), 16 pages. https://doi.org/10.1145/3478513.3480528

[38] Douglas G. Macharet and Dinei A. Florencio. 2012. A collaborative control system for telepresence robots. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 5105–5111. https://doi.org/10.1109/IROS.2012.6385705

[39] Evelyn Z. McClave. 2000. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics* 32, 7 (2000), 855–878. https://doi.org/10.1016/S0378-2166(99)00079-X

[40] Microsoft. 2023. Azure Kinect DK - develop AI models: Microsoft Azure. Retrieved Feb 6, 2023 from https://azure.microsoft.com/en-us/services/kinect-dk/

[41] Mirror. 2023. Mirror Networking – Open Source Networking for Unity. Retrieved Feb 6, 2023 from https://mirror-networking.com/

[42] Kana Misawa, Yoshio Ishiguro, and Jun Rekimoto. 2012. LiveMask: A Telepresence Surrogate System with a Face-Shaped Screen for Supporting Nonverbal Communication. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (Capri Island, Italy) *(AVI '12)*. Association for Computing Machinery, New York, NY, USA, 394–397. https://doi.org/10.1145/2254556.2254632

[43] Pieter Moors, Filip Germeys, Iwona Pomianowska, and Karl Verfaillie. 2015. Perceiving where another person is looking: the integration of head and body information in estimating another person's gaze. *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.00909

[44] Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 418–431. https://doi.org/10.1145/2818048.2819922

[45] David Nguyen, John Canny, and John Canny. 2005. MultiView: Spatially Faithful Group Video Conferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 799–808. https://doi.org/10.1145/1054972.1055084

[46] Ken-Ichi Okada, Fumihiko Maeda, Yusuke Ichikawaa, and Yutaka Matsushita. 1994. Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (Chapel Hill, North Carolina, USA) *(CSCW '94)*. ACM, New York, NY, USA, 385–393. https://doi.org/10.1145/192844.193054

[47] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. ACM, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517

[48] K. Otsuka. 2016. MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. In *2016 IEEE Virtual Reality (VR)*. 19–28. https://doi.org/10.1109/VR.2016.7504684

[49] Ye Pan and Anthony Steed. 2014. A Gaze-preserving Situated Multiview Telepresence System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 2173–2176. https://doi.org/10.1145/2556288.2557320

[50] Eric Paulos and John Canny. 1998. PRoP: Personal Roving Presence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 296–303. https://doi.org/10.1145/274644.274686

[51] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew Wilson. 2016. 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* (San Francisco, California, USA) *(CSCW '16)*. ACM, New York, NY, USA, 1716–1725. https://doi.org/10.1145/2818048.2819965

[52] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 46, 13 pages. https://doi.org/10.1145/3173574.3173620

[53] Thammathip Piumsomboon, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. CoVAR: A Collaborative Virtual and Augmented Reality System for Remote Collaboration. In *SIGGRAPH Asia 2017 Emerging Technologies* (Bangkok, Thailand) *(SA '17)*. Association for Computing Machinery, New York, NY, USA, Article 3, 2 pages. https://doi.org/10.1145/3132818.3132822

[54] Irene Rae, Bilge Mutlu, and Leila Takayama. 2014. Bodies in Motion: Mobility, Presence, and Task Awareness in Telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 2153–2162. https://doi.org/10.1145/2556288.2557047

[55] Irene Rae, Leila Takayama, and Bilge Mutlu. 2012. *One of the Gang: Supporting in-Group Behavior for Embodied Mediated Communication*. Association for Computing Machinery, New York, NY, USA, 3091–3100. https://doi-org.proxy.library.cornell.edu/10.1145/2207676.2208723

[56] Irene Rae, Leila Takayama, and Bilge Mutlu. 2013. The Influence of Height in Robot-mediated Communication. In *Proceedings of the 8th ACM/IEEE International*

Conference on Human-robot Interaction (Tokyo, Japan) *(HRI '13)*. IEEE Press, Piscataway, NJ, USA, 1–8. http://dl.acm.org/citation.cfm?id=2447556.2447558

[57] Lorenzo Riano, Christopher Burbridge, and TM McGinnity. 2011. A Study of Enhanced Robot Autonomy in Telepresence. In *Unknown Host Publication*. AICS, Ireland. Proceedings of Artificial Intelligence and Cognitive Systems, AICS ; Conference date: 01-01-2011.

[58] Double Robotics. 2023. Telepresence Robot for the Hybrid Office. Retrieved Feb 6, 2023 from https://www.doublerobotics.com/

[59] Mose Sakashita, E. Andy Ricci, Jatin Arora, and François Guimbretière. 2022. RemoteCoDe: Robotic Embodiment for Enhancing Peripheral Awareness in Remote Collaboration Tasks. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 63 (apr 2022), 22 pages. https://doi.org/10.1145/3512910

[60] Avatar SDK. 2023. Realistic avatars for games, VR and AR. Lifelike avatars for the metaverse. Retrieved Feb 6, 2023 from https://avatarsdk.com/

[61] Abigail J. Sellen. 1992. Speech Patterns in Video-mediated Conversations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. ACM, New York, NY, USA, 49–59. https://doi.org/10.1145/142750.142756

[62] Stephen Siemonsma and Tyler Bell. 2022. HoloKinect: Holographic 3D Video Conferencing. *Sensors* 22, 21 (2022), 8118.

[63] David Sirkin, Gina Venolia, John Tang, George Robertson, Taemie Kim, Kori Inkpen, Mara Sedlins, Bongshin Lee, and Mike Sinclair. 2011. Motion and Attention in a Kinetic Videoconferencing Proxy. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part I* (Lisbon, Portugal) *(INTERACT'11)*. Springer-Verlag, Berlin, Heidelberg, 162–180. http://dl.acm.org/citation.cfm?id=2042053.2042074

[64] Brett Stoll, Samantha Reig, Lucy He, Ian Kaplan, Malte F. Jung, and Susan R. Fussell. 2018. Wait, Can You Move the Robot? Examining Telepresence Robot Use in Collaborative Teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 14–22. https://doi.org/10.1145/3171221.3171243

[65] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. 2006. Collaborative Coupling over Tabletop Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1181–1190. https://doi.org/10.1145/1124772.1124950

[66] John C. Tang and Scott Minneman. 1991. VideoWhiteboard: Video Shadows to Support Remote Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New Orleans, Louisiana, USA) *(CHI '91)*. ACM, New York, NY, USA, 315–322. https://doi.org/10.1145/108844.108932

[67] Richard M Taylor. 2017. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Situational awareness*. Routledge, 111–128.

[68] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174. https://doi.org/10.1145/3332165.3347872

[69] Tobii. 2022. Global leader in eye tracking for over 20 years. Retrieved Feb 6, 2023 from https://www.tobii.com/

[70] Katherine M. Tsui and Holly A. Yanco. 2013. Design Challenges and Guidelines for Social Interaction Using Mobile Telepresence Robots. *Reviews of Human Factors and Ergonomics* 9, 1 (2013), 227–301. https://doi.org/10.1177/1557234X13502462 arXiv:https://doi.org/10.1177/1557234X13502462

[71] Roel Vertegaal. 1999. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 294–301. https://doi.org/10.1145/302979.303065

[72] Simon Voelker, Sebastian Hueber, Christian Holz, Christian Remy, and Nicolai Marquardt. 2020. GazeConduits: Calibration-Free Cross-Device Collaboration through Gaze and Touch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376578

[73] Wei-Chao Wen, H. Towles, L. Nyland, G. Welch, and H. Fuchs. 2000. Toward a compelling sensation of telepresence: demonstrating a portal to a distant (static) office. In *Proceedings Visualization 2000. VIS 2000 (Cat. No.00CH37145)*. 327–333. https://doi.org/10.1109/VISUAL.2000.885712

[74] Nicole Yankelovich, Nigel Simpson, Jonathan Kaplan, and Joe Provino. 2007. Porta-person: Telepresence for the Connected Conference Room. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA, USA) *(CHI EA '07)*. ACM, New York, NY, USA, 2789–2794. https://doi.org/10.1145/1240866.1241080

[75] Guangtao Zhang, John Paulin Hansen, Katsumi Minakata, Alexandre Alapetite, and Zhongyu Wang. 2019. Eye-Gaze-Controlled Telepresence Robots for People with Motor Disabilities. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 574–575. https://doi.org/10.1109/HRI.2019.8673093

[76] Jakob Zillner, Christoph Rhemann, Shahram Izadi, and Michael Haller. 2014. 3D-Board: A Whole-Body Remote Collaborative Whiteboard. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 471–479. https://doi.org/10.1145/2642918.2647393