Hands or Controllers? How Input Devices and Audio Impact Collaborative Virtual Reality

Alex Adkins
adkins4@clemson.edu
Clemson University
Clemson, South Carolina, USA

Ryan Canales
rcanale@clemson.edu
Clemson University
Clemson, South Carolina, USA

Sophie Jörg
sophie.joerg@uni-bamberg.de
University of Bamberg
Bamberg, Germany
Clemson University
Clemson, South Carolina, USA

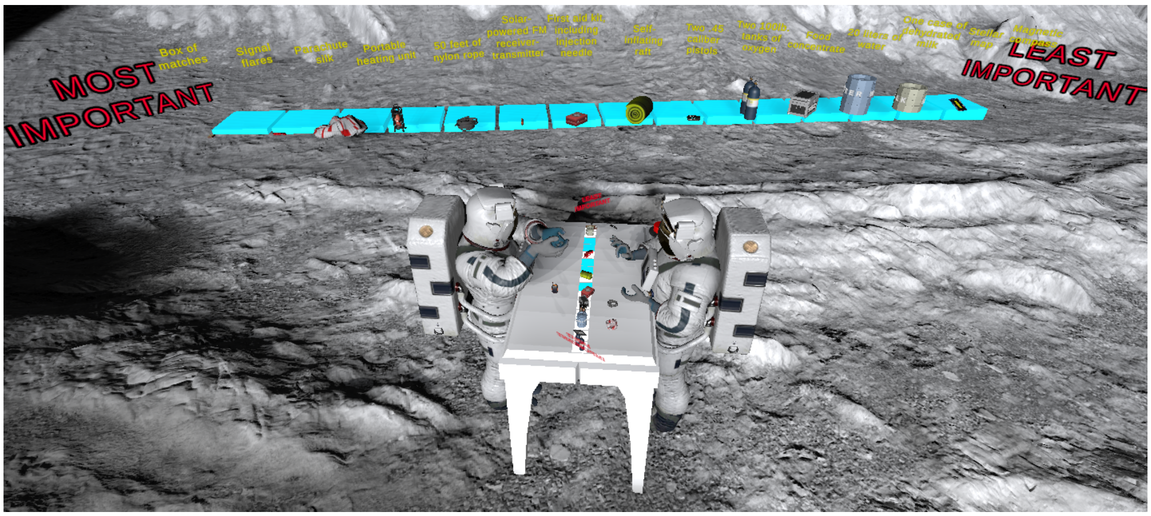

Figure 1: Two participants with embodied avatars in our remote collaboration scenario. They are negotiating which items are most or least important to their survival in NASA's Survival on the Moon task, using either VR controllers or hand tracking with or without verbal communication.

Abstract

Advancing virtual reality technologies are enabling real-time virtual-face to virtual-face communication. Hand tracking systems that are integrated into Head-Mounted Displays (HMD) enable users to directly interact with their environments and with each other using their hands as opposed to using controllers. Due to the novelties of these technologies our understanding of how they impact our interactions is limited. In this paper, we investigate the consequences of using different interaction control systems, hand tracking or controllers, when interacting with others in a virtual environment. We design and implement NASA's *Survival on the Moon* teamwork evaluation exercise in virtual reality (VR) and test for effects with and without allowing verbal communication. We evaluate social presence, perceived comprehension, team cohesion, group synergy, task workload, as well as task performance and duration. Our findings reveal that audio communication significantly enhances social presence, perceived comprehension, and team cohesion, but it also increases effort workload and negatively impacts group synergy. The choice of interaction control systems has limited impact on various aspects of virtual collaboration in this scenario, although participants using hand tracking reported lower effort workload, while participants using controllers reported lower mental workload in the absence of audio.

CCS Concepts

• **Human-centered computing → Interaction devices**; *Empirical studies in collaborative and social computing*; • **Computing methodologies → Virtual reality**.

Keywords

Communication, collaboration, gestures, avatars

ACM Reference Format:
Alex Adkins, Ryan Canales, and Sophie Jörg. 2024. Hands or Controllers? How Input Devices and Audio Impact Collaborative Virtual Reality. In *30th ACM Symposium on Virtual Reality Software and Technology (VRST '24), October 09–11, 2024, Trier, Germany.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3641825.3687718

VRST '24, October 09–11, 2024, Trier, Germany

ACM ISBN 979-8-4007-0535-9/24/10
https://doi.org/10.1145/3641825.3687718

1 Introduction

Remote collaboration has significantly increased in recent years as more workplaces allow for hybrid and remote work. Many people regularly use video conferencing applications (Zoom, Facetime, MS Teams, etc.) for meetings, task collaboration, and even to socialize. However, challenges persist with the technology that limit effectiveness of communication such as issues achieving correct eye contact [16], difficulties pointing to things [46], and significant behavioral differences, particularly regarding gestures [3]. Embodied virtual reality (VR) and shared immersive virtual environments could provide solutions to some of these issues. In these environments, users are represented as virtual avatars that enable users to exhibit their own movements to some extent in the virtual space. Embodied VR shows promise of being an effective form of remote collaboration, enabling natural interactions and social presence levels similar to face-to-face conversation [91]. Current consumer VR devices perform head and hand movement tracking via tracked controllers, and several even track the hands and finger motions without controllers. Hand tracking is particularly important when it comes to communication in VR as it can enable natural communicative gestures that might not be possible using controllers alone. While previous research has evaluated the usage of hand tracking compared to controllers when interacting with objects in VR [61], the influence of detailed hand tracking for remote collaboration in VR has yet to be studied.

In this paper, we study the influence of detailed tracked hand and finger motions compared to controller driven hand movement with several button-mapped gestures in a real-time collaborative task involving communicating with another person as well as interacting with objects. In a between-subjects study, dyads of participants embody avatars in a shared virtual environment, solving NASA's *Survival on the Moon* [36] teamwork evaluation exercise individually then collaboratively. Dyads experienced one of two control modes: Hand Tracking and VR Controllers. To further isolate the effects of hand tracking on communication, we manipulate verbal communication by either enabling or disabling the transmission of participants' voices. We measure participants' levels of social presence, perceived comprehension, team cohesiveness, group synergy, task workload, task performance, and duration.

Our findings reveal that audio communication is essential for a robust VR experience, finding significant positive impact on social presence, perceived comprehension, and team cohesion. However, an added audio channel may have surprising drawbacks in a negotiation task. The added discussion enabled by the audio channel had participants working longer and harder and ultimately with worse score improvement (synergy) from the individual phase. We find little in the way of results for interaction controls which may imply that, in this type of setting and task, hand tracking and VR controllers are equivalent for many aspects of communication, though participants using the VR controllers did report working harder to achieve their level of performance. Participants reported higher mental workload when using hand tracking with no audio than with VR controllers and no audio, which might hold implications of trying to gesturally communicate without audio.

2 Related Work

Claude Shannon considered communication as a form of information processing [87], where a "transmitter" sends a message through a channel to a "receiver." However, the information transmitted picks up "noise" along the way, thus degrading the quality of communication. If we consider VR as a channel or system of communication, discussed by Biocca and Levy as early as 1995 [14], then we must consider that using this channel may alter the communication that is traditionally transmitted face-to-face. Though still a long way from matching the richness of face-to-face communication, a growing body of research suggests VR is a promising channel for natural remote collaboration and communication [2].

2.1 Communication and Collaboration in VR

Several studies have compared collaboration in virtual reality to other types of remote collaboration and found that VR collaboration has similarities to face-to-face collaboration, especially when embodied avatars are used. For a collaborative task involving planning furniture placement, Smith & Neff [91] found comparable social presence between embodied VR and face-to-face interactions even with low-quality avatars. Abdullah *et. al.* [3] compared videoconferencing (e.g. Zoom) to shared VR spaces with high fidelity movement tracking for remote collaboration and found that users' behavior in the shared VR space more closely matched behavior observed during face-to-face interactions, whereas videoconferencing led to an increased use of self-adaptor gestures, which are associated with increased anxiety [101], and decreased deictic gestures (e.g. pointing). They also found evidence that more effort was required to maintain social connection in the videoconferencing condition.

Whole-body avatars have been shown to improve communication, presence, and usability in VR, and using them is somewhat comparable to face-to-face interaction [26, 75, 91, 99]. Avatar expressiveness, embodiment, and naturalness are important factors for collaboration in VR [9]. Wu *et al.* [106] compared a highly expressive VR system with detailed facial expressions and hand and body tracking to a low expressive system with limited motion tracking, facial expressions, and controllers for the hands in a charades task. They found that participants interacting with the highly expressive avatar felt higher social presence and attraction, and could better perform a charades task. Freeman *et al.* [30] highlighted the importance of avatar embodiment, realism, and naturalness in facilitating virtual collaboration with others based on a series of in-depth interviews with social VR users.

Collaboration in virtual reality continues to be an active research field, with new approaches and applications [29, 54, 108]. However, the significance of the non-verbal components, particularly hand and finger motion, remains to be investigated.

2.2 Non-verbal Communication and Hand Motions

In a group setting, both verbal and non-verbal communication serve important roles. Verbal communication is often used to define and order group tasks, determine how tasks will be completed, and to talk about the group itself. Non-verbal communication, which comprises between 70-93% of communication at any given time [69], contributes to the quality of relationships between group members and serves to add meaning and emphasis to interactions [39].

While communication between individuals within VR can occur over text, voice chat, icons, and cues [57, 60], an advantage of VR for communication is that it enables expressive non-verbal interactions such as facial expressions [45, 93], proxemics [17, 18, 41], and gestures [15, 65]. Hand gestures in particular are an important part of communication and an essential part of face-to-face dialogue. In fact, individuals gesture even when their conversational partner cannot see them, such as when talking on the phone [103]. Furthermore, research suggests that gesturing helps to facilitate thinking [33], with studies finding that the rate of gesture increases when speakers are describing a scene from memory or when individuals reason about problems [7, 24]. We also use gestures to convey information to a listener. Gestures can substitute missing vocabulary [10], help explain social structures [27], change the observer's perception of the individual [76, 100], or indicate deception [22, 25].

Because finger motions can be difficult to capture in real-time, they are rarely considered in VR communication experiments [26], even though research suggests they may influence comprehension and perceived personality [48]. Of particular importance are gestures [32, 47, 68], but without hand tracking, gestures are constrained to what users can choose via interface or pantomime with controllers, often without meaningful control [93]. Although there has been evidence that the lack of detailed finger motion reduces social presence and influences the perceived content of a conversation when watching virtual characters on a screen [5], only recently are detailed gestures entering into immersive VR spaces, with few works examining how they might matter in group VR settings.

2.3 Hand Tracking and Controllers in VR

Handheld controllers remain the standard for interactions in virtual environments in many current applications [31]. However, with hand tracking technology improving over recent years, there has been a growing interest in research evaluating hand tracking for interaction with virtual objects [8, 61]. Research suggests that more natural interaction techniques, enabled by hand tracking, can increase user enjoyment, presence, and embodiment even if performance might be reduced [4, 62, 67, 74, 90]. In the absence of haptic feedback when grasping, visual or auditory feedback can increase performance and is generally preferred by users over not having any feedback even in cases where it does not improve efficiency [21, 55, 81, 98].

A few studies have directly compared hand tracking and VR controllers [4, 37, 50, 62, 64, 96, 97]. However, all of them focus on an individual's interaction with objects, none of them investigate effects during communication or in a collaborative setting as we do in our study. In these studies, overall results point toward reaching higher performance with controllers [4, 37, 50, 64], whereas hand interaction is often preferred [50, 62, 97] and can result in higher perceived ownership or embodiment [4, 62, 96].

In this work, we compare interactions via controllers, which provide a fixed set of hand poses mapped to buttons, to interactions allowing detailed gestures via tracked hand and finger motions in a collaborative setting. We specifically want to investigate how detailed finger motion, or the lack thereof, might influence communication in remote collaboration tasks in shared virtual spaces.

Table 1: The two independent variables and resulting experimental conditions.

		Interaction Controls	
		Hand Tracking	VR Controller
Audio	Audio	HandsA	ControllersA
Presence	No Audio	HandsNA	ControllersNA

2.4 Evaluating Experiences and Communication in VR

A major metric for evaluating experiences in VR is the sense of presence, which is the "sense of being" within a virtual space [88]. Presence can be subdivided into three types: self presence, social presence, and physical presence [43, 58, 85]. Self presence, the feeling of embodying your virtual avatar, social presence, the feeling that others also exist within the virtual space with you, and physical presence, the sense of the environment being an actual space around you, all play a critical role in inducing presence within VR [11, 13, 19, 58] and are often used as measures for VR research experiences. In this study, we measure social presence, which can influence satisfaction, enjoyment, attraction, and trust [42, 59], and is often associated with positive communication outcomes [78]. Previous work shows higher reported self and social presence and enjoyment with intuitive controls over buttons and gamepads [6], though indirect button input was still preferred for some actions [79].

The quality of communication in a group setting can influence team cohesion. Team cohesion, also called team cohesiveness, is the team members' desire to remain in the group and their willingness to work to accomplish the team's goals [51]. Team cohesiveness is a popular metric in determining team quality and is a predictor of team performance and success [23, 28, 44, 102]. It has been used to investigate the effects of computer mediated communication on teamwork and group effectiveness. Torro *et al.* [94] found that team cohesion was limited in social virtual reality, but that developing VR technologies will likely soon mitigate these effects. Liszio *et al.* [63] found that interactive social entities increase team cohesion in a virtual environment. VR is also being proposed as a way to study team cohesion and mental health for long-term space missions [83].

To assess participants' workload we employ NASA's Task Load Index (TLX) [40], which is commonly used in VR research. Previous work has indicated that VR [34, 35] and hand tracking [37] increase mental task load, but these findings have not been investigated in a collaborative environment.

3 Experimental Design

This study uses a 2x2 between-subjects design to evaluate the effects of the Interaction Control conditions, Hand Tracking and VR Controller, and the Audio Presence conditions, Audio and No Audio, (see Table 1) on social presence, perceived comprehension, team cohesion, group synergy, and task workload. As behavioral measures, we furthermore evaluate task performance and duration.

3.1 Hypotheses

Our hypotheses are that more detailed hand motions as provided in the hand tracking condition will increase social presence [5], perceived comprehension [5], team cohesion, and task workload [37] as compared to the VR Controller condition. Effects might be more pronounced in the absence of audio. We do not expect an effect of the type of interaction controls on group synergy as previous works have not found that performance improves between mediums [3, 53, 92] for intellective tasks as defined by McGrath [66]. However, Potter and Balthazard [80] did find that virtual teams produced less group synergy when compared to face-to-face teams.

For the influence of audio, we expect that the possibility to talk will increase our measures except for task workload. So we hypothesize that the presence of audio will increase social presence [84], perceived comprehension, team cohesion, and group synergy compared to the No Audio condition, and that it will reduce the perceived task workload due to the added simple means of communication.

3.2 Participants

There were a total of 30 participants (15 dyads) for each condition combination outlined in Table 1, for a total of 120 participants. 29% reported themselves as female, 70% as male, and 1% as other. Participants' ages ranged from 18 to 62 ($mean = 24, std = 6.7$). Participants were recruited locally via flyers, hand-outs, e-mail, Reddit, and word of mouth, with most being university students. Participants were recruited individually. They were asked to sign up for time slots and were not hindered from signing up for the same time slot as someone they knew. The experiment was announced as a group task. Conditions alternated and were assigned based on the participants' identification number, which were assigned sequentially. The study was approved by Clemson University's Institutional Review Board (IRB).

3.3 Experimental Conditions

In this section we present the technical details and implementation of our conditions.

3.3.1 Interaction Control Conditions. All participants wear a Meta Quest 2 HMD. Participants in the VR Controller condition use the Meta Quest 2 Touch controllers to interact within the VR scene. Participants in the hand tracking condition use their own hands tracked by Meta Quest 2's markerless optical tracking [72]. While the technology has limitations such as a small capture area and further techniques exist [49, 104], we have designed the environment and task of our experiment to encourage good quality tracking, placing all objects within easy reach to promote motions where the hands stay in the area with the highest tracking accuracy.

Both Interaction Control conditions use the native Meta Quest 2 tracking technologies, implemented in Unity 2020.3.34f1 with Oculus' Interaction SDK version 40.0 [95] and Oculus' Legacy OVRPlugin [71]. The Mirror Networking for Unity package (version 66.0.9) manages the transmissions between a nearby desktop computer (the server) and the participants' HMDs. Each player saw their own hands as represented by the Oculus Interaction SDK package, virtual hands with motions that either closely mimic their own hand movements for the Hand Tracking condition, or that are selected from a series of common hand poses for the VR Controller condition. The hand poses in the VR Controller condition are determined by the buttons being pressed or touched on the controller. Sensors exist for the thumb button, index finger trigger, and middle/ring finger trigger, creating a total of 8 poses for the virtual hands (see video). In the Hand Tracking condition participants can grab items by pinching the item with their thumb and index finger. Participants in the VR Controller condition press the thumb button and the index finger trigger to grasp. The virtual hand makes a similar pinching pose as in the Hand Tracking condition, see Figure 2. For all participants, audio feedback indicates whether they have grasped or released an object as audio feedback is preferred to not having audio feedback when grasping [20].

Figure 2: The virtual Astronaut hands. The "relaxed" (top) and "grasping" (bottom) virtual hands as controlled by Hand Tracking (left) and VR Controllers (right). In the Hand Tracking condition the virtual hand follows the users' detailed finger motions, in the VR Controller condition the poses are pre-defined.

3.3.2 Audio Presence Conditions. All audio is emitted from the built-in Meta Quest 2 HMD speakers. Vocal communication for the Audio condition is transmitted using the Dissonance Voice Chat for Mirror Networking Unity package (version 8.1.0). For dyads within the Audio condition, the HMD's internal microphones activate and the participants can communicate vocally after they have entered the multi-player environment. Dyads participating in the No Audio condition must conduct all of their communication non-verbally.

3.4 The Astronaut Avatar

The virtual avatars within the shared virtual spaces are identical for each participant. The full-body avatar is represented as an astronaut with an opaque, reflective helmet on. Previous works show higher usability when utilizing a full-body avatar [99], so inverse kinematics steered the avatar's arms based on the hands' positioning. Additionally, the astronaut's helmet rotates within a limit according to the participant's headset rotation. The avatars are standing, but only their torsos and up are visible above the virtual table unless participants lean far out of the operating space. These settings are present in both the single and multi-player task phases.

Previous work has found that more realistic hand representations create a higher sense of body ownership [61], but that a mismatch between the gender of the participant and their virtual hand can cause reduced presence [86], and that a race mismatch can alter body ownership and behavior [52]. The astronaut avatar is neutral and appropriate for all participants and it fits the task setting (described in Section 3.5).

The participants' virtual astronaut hands are each rendered twice, once locally for their own viewing, and again on the server for their partner to see. Both sets of virtual hands display the same motions at all times and are scaled to match the participants' own hand size. Due to inverse kinematics constraints, the avatar hands can pull away from the arms if the player reaches far, as the avatar's torso does not move. Thus, all items were placed close to the players to limit this behavior.

3.5 Scenario: NASA's Survival on the Moon Task

Participants are placed into a virtual scene designed around NASA's Survival on the Moon task (see Figure 1). In this scenario, participants have crash-landed on the moon 200 miles from their rendezvous point with a mother ship. Their survival depends on reaching the mother ship. Participants must rank 15 items scavenged from their ship (e.g., a box of matches, tanks of oxygen, flare gun, stellar map, etc.) in order of importance to their survival of the 200-mile trip. The virtual environment is set on the moon, with miniature virtual models of the items to be ranked placed on a table in front of them in the same order for all participants, visible in Figure 3. The mini items' descriptions appear when the item is being grasped. Real-size models of the items are visible in the near distance with their descriptions always hovering above them (see Figure 1). The table has 15 slots for items, ordered by importance. The miniature items snap into place and the slot changes color to indicate it is occupied.

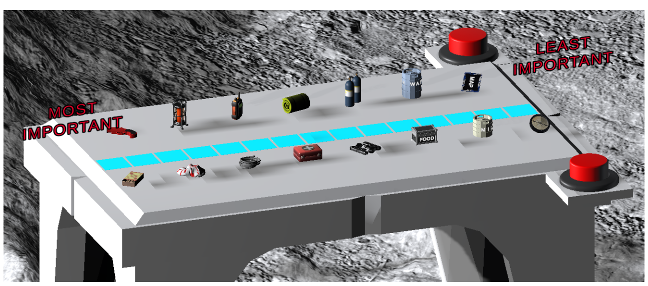

Figure 3: The table and operating space where participants place the items in order of importance in the multi-player phase. Here, in the multi-player phase, the items are distributed to either side of the table. Participants are randomly placed on either side of the table and remain on that side for both phases.

3.6 Procedure

Two participants are welcomed concurrently by two researchers into separate laboratory spaces. They are asked to complete and sign consent and cybersickness forms. Participants must consent to allow their data to be used in the study and pass all cybersickness checks to be qualified to continue. If there are no disqualifiers among either participant they fill out a demographics questionnaire on a nearby desktop computer. The researcher measures the participants' interpupillary distance (IPD) using the GlassesOn mobile app [1], then presents the scenario description and instructions on a piece of paper. While participants review the scenario, the researchers adjust the HMD to their IPD. Once participants have reviewed and confirmed their understanding of the scenario, the researchers instruct them on how to grasp and manipulate the virtual items, as described in Section 3.3.1. Finally, the participants are shown to their respective VR station and are aided in putting on and adjusting the HMD. Participants remain seated for the duration of the scenario. They are guided on how to open the Moon application and how to input their participant number. Once they have done so they are now present within the virtual moon environment in the single-player phase. The items appear on the same side of the table as the participant. Each participant first ranks the items individually by placing them onto the ranked spaces within individual scenes without seeing or communicating with their partner. Participants are directed to rank the items based on their textual names, as in the original NASA scenario, and not based on their visual appearance.

Once participants have ranked the items individually, they press a virtual button and are then placed into the multi-player environment. This environment is identical to the single-player environment, except that the virtual items do not appear until both players have entered the scene. Players would wait an average of 213 seconds (std = 189s, median = 161s, min = 6s, max = 850s) for their partner to join them in the multi-player environment. When both participants are in the multi-player phase, they can see each other's astronaut avatar and, if in the Audio condition, speak with one another through the HMD's microphone and speakers. They are then asked to collaboratively negotiate and discuss which scavenged items are the most important to their survival. They rank the items accordingly by picking up the items and placing them into ranked item spaces. Both participants can interact with all scavenged items, though only one hand may grasp one item at a time. If items leave the operating space above the table they reappear in their original positions so that an item may not be lost. There are no time restrictions on either phase of the task.

After completing the VR phase of the experiment, participants are asked to fill out the post-experiment questionnaire (see Section 3.7 and Table 2) on the nearby desktop computers. Finally, each participant is invited to a debrief together in an adjacent laboratory space. They are told their individual and group scores as compared to expert rankings, and awarded a $5 incentive card.

3.7 Measures

In this study, we evaluate social presence, perceived comprehension, team cohesion, group synergy, task workload, as well as task performance and duration. The detailed survey questions for the subjective measures are listed in Table 2.

3.7.1 Social Presence & Perceived Comprehension. Social Presence is measured with Nowak and Biocca's [77] measure of social presence, originally sourced from Short *et al* [89].

Biocca *et al*'s Networked Minds Social Presence Inventory [12, 13] includes a measure of perceived comprehension that we use for this study. Accurate comprehension implies effective communication and will likely be affected by Interaction Controls. Both of these measures use 7-pt Likert scales.

3.7.2 *Team Cohesion.*

Our measurement of team cohesion is based on Michalisin *et al.* [73]. They define a 5-pt Likert team cohesion assessment (see Table 2) that focuses on good working relationships, high contribution levels, and shared commitment to completing the group task.

3.7.3 *Performance & Group Synergy.*

NASA's Survival on the Moon task provides expert item rankings, with a rationale for each item. These rankings act as an answer key and are provided to participants during the post-experiment debrief. Participants' individual and group rankings are scored compared to NASA's solution. A lower score indicates better task performance in a range from 0 to 112. Participants' ranking scores are compared to produce weak and strong group synergy scores, which are measures of the group's gain in performance compared to the individuals' [56, 70].

3.7.4 *Task Workload.*

Participants' workload is measured using NASA's Task Load Index (TLX) [40]. Harris *et al.* [38] recently developed and validated an updated version of the index designed for VR. All questions are asked on a 1-10 rating or as a choice between two factors.

4 Results

Data for all measures were found to be significantly non-normal using Shapiro-Wilk normality test, and thus no parametric assumptions are met. Due to this, we use a robust two-way independent ANOVA on trimmed means (t2way in R) to conduct analysis on all metrics [105]. The corresponding post-hoc test for a two-way ANOVA based on trimmed means (mcp2atm in R) was performed for interaction effects. Test statistics, p-values, means, and standard deviations are provided in Table 2.

4.1 Social Presence, Perceived Comprehension, Team Cohesion

A significant main effect of Audio Presence was found for Social Presence ($Q = 6.01, p < 0.05$), Perceived Comprehension ($Q = 31.14, p < 0.001$), and Team Cohesion ($Q = 4.96, p < 0.05$), visible in Figure 4. Participants who could speak with one another in the Audio condition reported higher social presence, perceived comprehension, and team cohesion as opposed to participants in the No Audio condition. No effects of Interaction Control were found for these measures.

4.2 Performance, Synergy, and Task Duration

Performance is measured as the accuracy of the item rankings when compared to NASA experts' rankings. The absolute differences of each item's ranking, compared to its official NASA ranking, are summed as a performance score. A lower score indicates better performance. Scores were recorded as individual scores ($mean = 40.95, std = 11.02$) and group scores ($mean = 38.4, std = 10.86$); there were no significant effects on performance of either the Interaction Control or Audio Presence. We did not expect a significant impact of either condition on performance; performance metrics are present to enable calculation of Group Synergy.

Group synergy was calculated as two factors: Weak Synergy and Strong Synergy, as in Meslec and Curşeu's [70] investigation of group synergy when using the Survival on the Moon Task (described in Section 3.5). Weak Synergy is computed as the difference between the group's score and the mean of the pair's individual scores. Strong Synergy is the difference between the group score and the best performing individual's. Lower synergy calculations indicate better synergy due to lower performance scores being the better scores.

A significant main effect of Audio Presence was present for Strong Synergy ($Q = 9.29, p < 0.01$). For this case, participants in the No Audio condition who could not speak to one another during the scenario's multi-player phase showed lower group Synergy than participants in the Audio condition (Figure 4).

A similar result of Audio Presence was found for Group Duration ($Q = 15.72, p < 0.001$). Groups who could verbally communicate with each another took a significantly longer amount of time to complete the task.

4.3 Task Workload

The adjusted task workloads were calculated by multiplying the task demand (TLX 1-6, see Table 2) by weights determined by each participant, which are the summed cumulative score derived from the 15 comparisons between demand dimensions.

Significant main effects of both independent variables, Interaction Control ($Q = 4.91, p < 0.05$) and Audio Presence ($Q = 7.12, p < 0.05$), were present for Effort Workload (TLX5: How hard did you have to work to accomplish your level of performance in the overall experience?). Participants who could not speak to one another (No Audio) reported a lower Effort Workload, while participants using the VR Controllers reported a higher Effort Workload, as visualized in Figure 4.

Additionally, a significant interaction effect ($Q = 6.22, p < 0.05$) was present between the Interaction Controls and Audio Presence for Mental Workload (TLX1: How mentally demanding was the overall experience?) as seen in Figure 5. Post-hoc testing revealed that VR Controllers resulted in significantly lower Mental Workload when compared to Hand Tracking in the No Audio condition ($\hat{\psi} = -10.89, p < 0.05$), but showed no significant difference on Mental Workload between Interaction Controls in the Audio condition.

4.4 Qualitative Feedback

As the last question of the post-experiment questionnaire, participants were asked to provide feedback on their experience. Two participants using VR Controllers without an Audio channel indicated that having an audio channel would "help" with "articulat[ing] ... certain reasonings for choices," while a third participant in the same condition expressed a desire for a "thumbs up sign" to help communication. No participants using Hand Tracking indicated similar desires.

5 Discussion and Limitations

The results confirm our hypotheses that the possibility to speak or not (Audio or No Audio) increases social presence, perceived

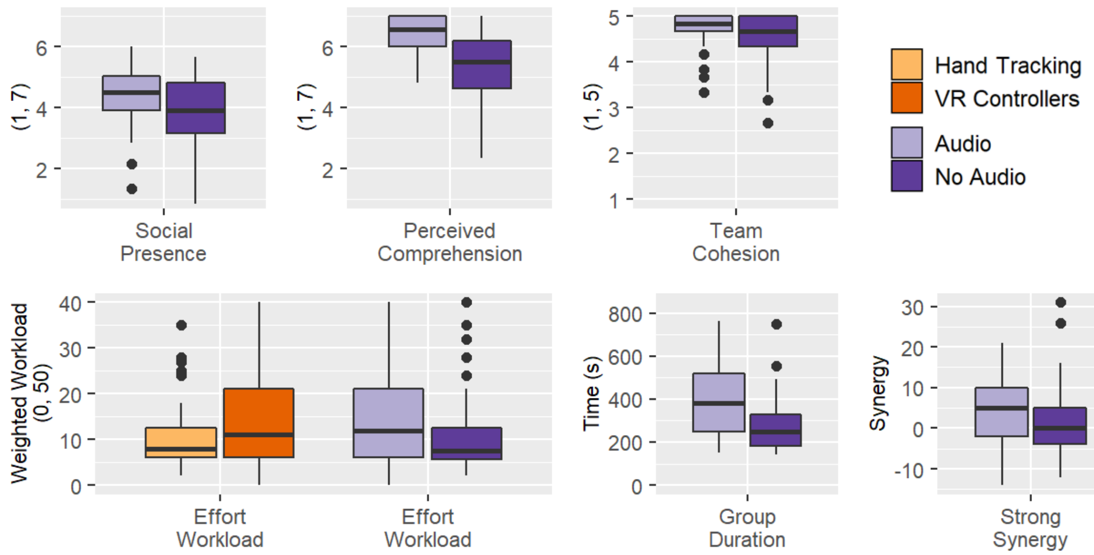

Figure 4: Significant results for Social Presence, Perceived Comprehension, Team Cohesion, Effort Workload (TLX5), Group Duration, and Strong Synergy. Lower synergy values indicate better group synergy.

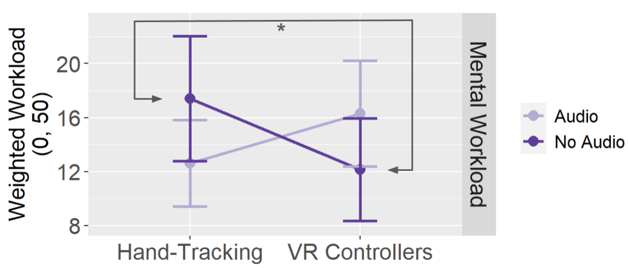

Figure 5: Significant Interaction Effect for Mental Workload (TLX1).

comprehension, and team cohesion. It is not surprising that being able to verbally communicate improves perceived comprehension. Being able to speak is closer to real-world, face-to-face experiences, and verbal communication itself is a rich communication channel. Our results also align with previous work that found that more face-to-face-like experiences can foster greater social presence and team cohesion [84].

Contrary to our expectations, the presence of audio was associated with an increase in perceived task workload, specifically Effort Workload ("How hard did you have to work to accomplish your level of performance in the overall experience?"). One explanation could be that the inclusion of audio allowed for extensive discussions, which is supported by the significantly longer group duration in the Audio condition. In the absence of voice communication, individuals might have been more inclined to accept their partner's choices, since communicating the reasoning behind their own choices may have been too difficult or impossible without using words. As one participant remarked, "without voice communication, it is somewhat difficult to articulate the intricacies behind certain reasonings for choices." However, in the Audio condition, participants may have engaged in more negotiation efforts, potentially resulting in an elevated sense of effort.

We did not find evidence that Audio improved group synergy. Surprisingly, results showed that the Audio condition was detrimental to strong synergy (the difference between the group score and the best-performing individual's score) compared to the No Audio condition. We postulate that the presence of the audio channel allowed for further negotiation and discussion between the participants, but it is unclear why the enhanced ability to reason and negotiate diminished the group performance.

We did not find evidence that Interaction Controls (Hand Tracking vs. Controllers) affected Social Presence, Perceived Comprehension, or Team Cohesion. Results from Adkins *et al.* [5] show a decrease in perceived comprehension and in social presence when hand motions are missing, but show no such effect when hand motions are merely less accurate. Based on this previous work we can assume that the hand motions generated by the controllers, while not being as accurate as tracked hand motions, are enough to create a similar level of perceived comprehension, social presence, and team cohesion in our scenario.

Interestingly, Hand Tracking resulted in lower perceived Effort Workload compared to VR controllers, contradicting our hypothesis. Previous work is not fully consistent when it comes to workload using hand tracking and controllers [37, 96]. One explanation would be that the effect depends on the exact task and implementation. In our study, participants communicated with one another, often using their hands to indicate and perform other gestures. Our Hand Tracking condition facilitated more natural gesturing, which may have led to lower perceived effort, since one does not need to press specific combinations of buttons to achieve certain gestures, as in the Controllers condition. Furthermore, the hand tracking on the

Table 2: Questionnaire and main results. Values from questions marked with (R) were reversed before analysis. Shorthands used are as follows: IC: Interaction Control, AP: Audio Presence, HT: Hand-Tracking, VC: VR Controllers, A: Audio, NA: No Audio

Measure	Question	Scale	Statistical Values	Results	Mean, Std.
Social Presence - Mean of SP1 - SP6 - 7pt Likert - Nowak & Biocca [77]	SP1. To what extent did you feel able to assess your partner's reactions to what you said?	Not able to assess reactions - Able to assess reactions	Audio Presence: $Q = 6.01, p < 0.05$	A > NA	A: (4.43, 0.9) NA: (3.92, 1.16)
	SP2. To what extent was this like a face-to-face meeting?	Not like face-to-face at all - A lot like face-to-face			
	SP3. To what extent was this like you were in the same room with the virtual character?	Not like being in the same room at all - A lot like being in the same room			
	SP4. To what extent did the virtual character seem "real?"	Not real at all - Very real			
	SP5. How likely is it that you would choose to use this system of interaction for a meeting in which you wanted to persuade others of something?	Not likely at all - Very likely			
	SP6. To what extent did you feel you could get to know someone that you met only through this system?	Not at all - Very well			
Perceived Comprehension - Mean of PC1 - PC6 - 7pt Likert - Biocca & Harms [12]	PC1. I was able to communicate my intentions clearly to my partner.	Strongly disagree - Strongly agree	Audio Presence: $Q = 31.14, p < 0.001$	A > NA	A: (6.47, 0.51) NA: (5.31, 1.22)
	PC2. My thoughts were clear to my partner.				
	PC3. I was able to understand what my partner meant.				
	PC4. My partner was able to communicate their intents clearly to me.				
	PC5. My partner's thoughts were clear to me.				
	PC6. My partner was able to understand what I meant.				
Team Cohesion - Mean of TC1 - TC6 - 5pt Likert - Michalisin *et al.* [73]	TC1. I enjoyed working with my teammates.	Strongly disagree - Strongly agree	Audio Presence: $Q = 4.96, p < 0.05$	A > NA	A: (4.73, 0.35) NA: (4.55, 0.49)
	TC2. I wish I were on a different team. (R)				
	TC3. The team worked well together.				
	TC4. Everyone contributed to the discussion.				
	TC5. The team wasted a lot of time. (R)				
	TC6. I trust that my teammates will do their fair share of the work.				
NASA's Task Load Index - 10pt Likert Weighted - Harris *et al.* [38]	TLX1. How mentally demanding was the overall experience?	Low - High (lower is better)	Interaction Effect (IC x AP): $Q = 6.22, p < 0.05$	HT,NA > VC,NA	HT,A: (12.63, 8.59) HT,NA: (17.4, 12.41) VC,A: (16.3, 10.48) VC,NA: (12.13, 10.2)
	TLX2. How physically demanding was the overall experience?				(1.32, 2.29)
	TLX3. How hurried or rushed was the pace of the overall experience?				(6.6, 6.43)
	TLX4. How successful were you in accomplishing what you were asked to do in the overall experience?				(30.72, 12.65)
	TLX5. How hard did you have to work to accomplish your level of performance in the overall experience?		Interaction Control: $Q = 4.91, p < 0.05$ Audio Presence: $Q = 7.12, p < 0.05$	HT < VC NA < A	HT: (10.83, 8.05) VC: (13.9, 10.22) A: (14.03, 9.77) NA: (10.7, 8.54)
	TLX6. How insecure, discouraged, irritated, stressed and annoyed were you with the overall experience?				(2.83, 5.11)
Additional Measures	Individual Score	0 (best) - 112 (worst)			(40.95, 11.02)
	Group Score				(38.4, 10.86)
	Weak Synergy	(lower is better)			(-2.55, 7.59)
	Strong Synergy		Audio Presence: $Q = 9.29, p < 0.01$	NA < A	A: (4.47, 8) NA: (2.13, 9.14)
	Individual Duration	Time (seconds)			(366.36, 218.4)
	Group Duration		Audio Presence: $Q = 15.72, p < 0.001$	NA < A	A: (394, 178.45) NA: (279.45, 135.7)

Meta Quest 2 uses modern hand tracking techniques without the need for additional hardware, such as gloves, thus reducing the extra effort that older hand tracking technologies may have induced. Other tasks, such as virtual grasping or precise object interaction, might be more difficult to perform via hand tracking, so using VR controllers could require less effort in those cases.

When examining Mental Workload ("How mentally demanding was the overall experience?") we found a noteworthy interaction effect. Participants reported lower Mental Workload when using VR Controllers compared to Hand Tracking, but only in the absence of audio communication. Due to the gestural freedom afforded by hand tracking, participants may have been more intentional in how they gestured in an attempt to effectively communicate with their partners when lacking an audio channel. Conversely, having limited gestural options in the VR Controller condition may be analogous to having fewer decisions to make (the scope of possible gestures is much more limited), and thus may have been why Mental Workload was lower when communication was limited to gesturing only. Although previous research suggests that gestures can serve as cognitive aids for verbal communication [33], it may have been unnatural and challenging for participants to gesture without the simultaneous ability to speak.

Several limitations should be noted in our study. The avatars used lacked realism and facial expressions, potentially impacting social presence and communication. While we animated the avatar's arm movement, hand movement, and head rotation to match that of the users, the rest of the body remained static. Perhaps the use of avatars that allow for more expressive movement would have led to slightly different results for social presence and communication performance, as previous research suggests [106]. Moreover, our task primarily involved manipulating virtual items, potentially diverting participants' focus from their partner's movements. Tasks demanding more deictic and iconic gestures might reveal differences between interaction control methods more prominently. Exploring diverse tasks with varying communication demands could offer a more comprehensive understanding of the effects of interaction controls on collaboration. Additionally, the physical motions of participants' gestures were not measured within this study. An investigation into the frequency and types of motions used when communicating during the scenario may have yielded informative results.

Another limitation is that our study was not gender balanced, with 70% of participants identifying as male, 29% as female, and 1% reported as other, and most of the participants were in their 20s. This imbalance might limit the generalizability of the results of our study, since research indicates that, for example, gender can affect negotiation and collaboration strategies in virtual environments [82, 107]. Finally, how well participants know each other might also influence results. We asked participants to report how well they knew the other participant on a scale from 1-5 ("not at all" to "very well"). 48.3% of participants reported not knowing their partner at all, whereas 15% somewhat agreed and 30% strongly agreed with the statement of knowing their teammate very well before the experiment.

While our study allowed participants to directly sign up for time slots and therefore did not control how pairs were formed, it would be interesting to purposefully form pairs based on gender and based on how well participants know each other to gain further insights on these variables. Of course, a generally broader population of participants, including a broader age distribution, the inclusion of members of the disability community, and a broader range of educational and cultural backgrounds would also be desirable and interesting to investigate.

6 Conclusion

In this study, we investigated the consequences of employing different Interaction Control systems, Hand Tracking and VR Controllers, in a collaborative virtual environment. We implemented NASA's Survival on the Moon teamwork evaluation exercise and examined the impact of these controls in the presence and absence of audio communication.

Our findings revealed limited differences in interaction controls, which might indicate that the hand motions enabled by controllers are sufficiently detailed for the type of collaborative task we tested. We did find a surprising decrease in perceived effort workload with Hand Tracking, contrary to our expectations. Not surprisingly, the presence of audio communication lead to significant positive effects, increasing social presence, perceived comprehension, and team cohesion. However, adding audio communication also increased perceived effort workload and negatively impacted group synergy and group task duration. A notable interaction effect emerged, where Hand Tracking without audio led to higher perceived mental workload compared to VR Controllers in similar conditions.

In conclusion, audio communication proves crucial for a robust collaborative VR experience. Our results also suggest that hand tracking and VR controllers are not significantly different in most aspects in a collaborative task that involves interaction with objects and communication with another person. Future work could explore the use of more realistic self-avatars, experiment with scenarios where the communicative aspects are more prominent, or examine further variables such as how the communicative partner is perceived or how gestures might differ between types of input devices.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number IIS-1652210. We would also like to thank Jackson Henry and Grace Lim for their assistance in running participants to gather data for this study.

References

[1] 6over6 Vision LTD. 2022. GlassesOn | Pupils & Lenses — App on Google Play. https://play.google.com/store/apps/details?id=com.sixoversix.copyglass&hl=en_US&gl=US. Accessed on 2022-07-11.

[2] Ali Abbas, Minji Choi, JoonOh Seo, Seung Hyun Cha, and Heng Li. 2019. Effectiveness of Immersive Virtual Reality-based Communication for Construction Projects. *KSCE Journal of Civil Engineering* 23, 12 (Dec. 2019), 4972–4983. https://doi.org/10.1007/s12205-019-0898-0

[3] Ahsan Abdullah, Jan Kolkmeier, Vivian Lo, and Michael Neff. 2021. Videoconference and Embodied VR: Communication Patterns Across Task and Medium. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 453:1–453:29. https://doi.org/10.1145/3479597

[4] Alex Adkins, Lorraine Lin, Aline Normoyle, Ryan Canales, Yuting Ye, and Sophie Jörg. 2021. Evaluating Grasping Visualizations and Control Modes in a VR Game. *ACM Transactions on Applied Perception* 18, 4 (Oct. 2021), 19:1–19:14. https://doi.org/10.1145/3486582

[5] Alex Adkins, Aline Normoyle, Lorraine Lin, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2023. How Important Are Detailed Hand Motions for Communication for a Virtual Character Through the Lens of Charades? *ACM Trans. Graph.* 42, 3 (2023), 27:1–27:16. https://doi.org/10.1145/3578575

[6] Monthir Ali and Rogelio E. Cardona-Rivera. 2020. Comparing Gamepad and Naturally-Mapped Controller Effects on Perceived Virtual Reality Experiences. In *ACM Symposium on Applied Perception 2020 (SAP '20).* 1–10. https://doi.org/10.1145/3385955.3407923

[7] Martha W. Alibali, Sotaro Kita, and Amanda J. Young. 2000. Gesture and the Process of Speech Production: We Think, Therefore We Gesture. 15, 6 (2000), 593–613. https://doi.org/10.1080/016909600750040571

[8] Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lécuyer. 2016. The Role of Interaction in Virtual Embodiment: Effects of the Virtual Hand Representation. In *2016 IEEE Virtual Reality (VR).* IEEE, 3–10.

[9] Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (May 2021), 2608–2617. https://doi.org/10.1109/TVCG.2021.3067783

[10] Janet Bavelas, Christine Kenwood, Trudy Johnson, and Bruce Phillips. 2002. An Experimental Study of When and How Speakers Use Gestures to Communicate. *Gesture* 2, 1 (Jan. 2002), 1–17. https://doi.org/10.1075/gest.2.1.02bav

[11] Frank Biocca. 1997. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication* 3, 2 (Sept. 1997). https://doi.org/10.1111/j.1083-6101.1997.tb00070.x

[12] Frank Biocca and Chad Harms. 2003. Guide to the Networked Minds Social Presence Inventory v. 1.2. http://cogprints.org/6743/.

[13] Frank Biocca, Chad Harms, and Judee Burgoon. 2003. Towards A More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence* 12 (Oct. 2003), 456–480. https://doi.org/10.1162/105474603322761270

[14] Frank Biocca and Mark R. Levy. 1995. Virtual Reality as a Communication System. In *Communication in the Age of Virtual Reality.* Lawrence Erlbaum

Associates, Inc, Hillsdale, NJ, US, 15–31.

[15] Frank Biocca and Mark R. Levy. 2013. *Communication in the Age of Virtual Reality.* Routledge.

[16] Leanne S. Bohannon, Andrew M. Herbert, Jeff B. Pelz, and Esa M. Rantanen. 2013. Eye contact and video-mediated communication: A review. *Displays* 34, 2 (2013), 177–185. https://doi.org/10.1016/j.displa.2012.10.009

[17] Andrea Bönsch, Sina Radke, Heiko Overath, Laura M. Asché, Jonathan Wendt, Tom Vierjahn, Ute Habel, and Torsten W. Kuhlen. 2018. Social VR: How Personal Space Is Affected by Virtual Agents' Emotions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 199–206. https://doi.org/10.1109/VR.2018.8446480

[18] Lauren E. Buck, John J. Rieser, Gayathri Narasimham, and Bobby Bodenheimer. 2019. Interpersonal Affordances and Social Dynamics in Collaborative Immersive Virtual Environments: Passing Together Through Apertures. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 2123–2133. https://doi.org/10.1109/TVCG.2019.2899232

[19] Saniye Tugba Bulu. 2012. Place Presence, Social Presence, Co-Presence, and Satisfaction in Virtual Worlds. *Computers & Education* 58, 1 (Jan. 2012), 154–161. https://doi.org/10.1016/j.compedu.2011.08.024

[20] Ryan Canales and Sophie Jörg. 2020. Performance Is Not Everything: Audio Feedback Preferred Over Visual Feedback for Grasping Task in Virtual Reality. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Virtual Event, SC, USA) *(MIG '20).* Association for Computing Machinery, New York, NY, USA, Article 4, 6 pages. https://doi.org/10.1145/3424636.3426897

[21] Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. Virtual Grasping Feedback and Virtual Hand Ownership. In *ACM Symposium on Applied Perception 2019.* 1–9. https://doi.org/10.1145/3343036.3343132

[22] Letizia Caso, Fridanna Maricchiolo, Marino Bonaiuto, Aldert Vrij, and Samantha Mann. 2006. The Impactof Deception and Suspicious on Different Hand Movements. *Journal of Nonverbal Behavior* 30, 1 (March 2006), 1–19. https://doi.org/10.1007/s10919-005-0001-z

[23] François Chiocchio and Hélène Essiembre. 2009. Cohesion and Performance: A Meta-Analytic Review of Disparities Between Project Teams, Production Teams, and Service Teams. *Small Group Research* 40, 4 (Aug. 2009), 382–420. https://doi.org/10.1177/1046496409335103

[24] Jan-Peter de Ruiter. 1998. *Gesture and Speech Production.* Number 6 in MPI Series in Psycholinguistics. Univ, Nijmegen.

[25] Bella M. DePaulo, James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper. 2003. Cues to Deception. *Psychological Bulletin* 129, 1 (2003), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

[26] Trevor J. Dodds, Betty J. Mohler, and Heinrich H. Bülthoff. 2011. Talk to the Virtual Hands: Self-Animated Avatars Improve Communication in Head-Mounted Display Virtual Environments. *PLoS ONE* 6, 10 (Oct. 2011), e25759. https://doi.org/10.1371/journal.pone.0025759

[27] N. J. Enfield. 2005. The Body as a Cognitive Artifact in Kinship Representations: Hand Gesture Diagrams by Speakers of Lao. *Current Anthropology* 46, 1 (Feb. 2005), 51–81. https://doi.org/10.1086/425661

[28] Charles R. Evans and Kenneth L. Dion. 1991. Group Cohesion and Performance: A Meta-Analysis. *Small Group Research* 22, 2 (1991), 175–186. https://doi.org/10.1177/1046496491222002

[29] C. G. Fidalgo, M. Sousa, D. Mendes, R. Dos Anjos, D. Medeiros, K. Singh, and J. Jorge. 2023. MAGIC: Manipulating Avatars and Gestures to Improve Remote Collaboration. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR).* IEEE Computer Society, Los Alamitos, CA, USA, 438–448. https://doi.org/10.1109/VR55154.2023.00059

[30] Guo Freeman, Dane Acena, Nathan J. McNeese, and Kelsea Schulenberg. 2022. Working Together Apart through Embodiment: Engaging in Everyday Collaborative Activities in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (Jan. 2022), 1–25. https://doi.org/10.1145/3492836

[31] Giorgos Ganias, Christos Lougiakis, Akrivi Katifori, Maria Roussou, Yannis Ioannidis, and Ioannis Panagiotis Ioannidis. 2023. Comparing Different Grasping Visualizations for Object Manipulation in VR using Controllers. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2369–2378. https://doi.org/10.1109/TVCG.2023.3247039

[32] Susan Goldin-Meadow. 1999. The Role of Gesture in Communication and Thinking. *Trends in Cognitive Sciences* 3, 11 (Nov. 1999), 419–429. https://doi.org/10.1016/S1364-6613(99)01397-2

[33] Susan Goldin-Meadow. 2005. *Hearing Gesture: How Our Hands Help Us Think.* Harvard University Press.

[34] Scott W Greenwald, Wiley Corning, Markus Funk, and Pattie Maes. 2018. Comparing Learning in Virtual Reality with Learning on a 2D Screen Using Electrostatics Activities. *Journal of Universal Computer Science* 24, 2 (2018), 220–245.

[35] Scott W. Greenwald, Zhangyuan Wang, Markus Funk, and Pattie Maes. 2017. Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality. In *Immersive Learning Research Network (Communications in Computer and Information Science)*, Dennis Beck, Colin Allison, Leonel Morgado, Johanna Pirker, Foaad Khosmood, Jonathon Richter, and Christian Gütl (Eds.). Springer International Publishing, Cham, 75–90. https://doi.org/10.1007/978-3-319-60633-0_7

[36] Jay Hall and W. H. Watson. 1970. The Effects of a Normative Intervention on Group Decision-Making Performance. *Human Relations* 23, 4 (Aug. 1970), 299–317. https://doi.org/10.1177/001872677002300404

[37] Asim Hameed, Andrew Perkis, and Sebastian Möller. 2021. Evaluating Hand-tracking Interaction for Performing Motor-tasks in VR Learning Environments. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX).* 219–224. https://doi.org/10.1109/QoMEX51781.2021.9465407

[38] David Harris, Mark Wilson, and Samuel Vine. 2020. Development and Validation of a Simulation Workload Measure: The Simulation Task Load Index (SIM-TLX). *Virtual Reality* 24, 4 (Dec. 2020), 557–566. https://doi.org/10.1007/s10055-019-00422-9

[39] Thomas E Harris and John C Sherblom. 2018. *Small Group and Team Communication.* Waveland Press.

[40] Sandra G. Hart. 1986. NASA Task Load Index (TLX).

[41] Béatrice S. Hasler and Doron A. Friedman. 2012. Sociocultural Conventions in Avatar-Mediated Nonverbal Communication: A Cross-Cultural Analysis of Virtual Proxemics. *Journal of Intercultural Communication Research* 41, 3 (Nov. 2012), 238–259. https://doi.org/10.1080/17475759.2012.728764

[42] Khaled Hassanein and Milena Head. 2007. Manipulating Perceived Social Presence through the Web Interface and Its Impact on Attitude towards Online Shopping. *International Journal of Human-Computer Studies* 65, 8 (Aug. 2007), 689–708. https://doi.org/10.1016/j.ijhcs.2006.11.018

[43] Carrie Heeter. 1992. Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments* 1, 2 (Jan. 1992), 262–271. https://doi.org/10.1162/pres.1992.1.2.262

[44] Michael A Hogg. 1992. The Social Psychology of Group Cohesiveness: From Attraction to Social Identity.

[45] Ruud Hortensius, Felix Hekele, and Emily S. Cross. 2018. The Perception of Emotion in Artificial Agents. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (Dec. 2018), 852–864. https://doi.org/10.1109/TCDS.2018.2826921

[46] Ellen A Isaacs and John C Tang. 1993. What video can and can't do for collaboration: a case study. In *Proceedings of the first ACM International Conference on Multimedia.* 199–206.

[47] Naomi Jacobs and Alan Garnham. 2007. The Role of Conversational Hand Gestures in a Narrative Task. *Journal of Memory and Language* 56, 2 (Feb. 2007), 291–303. https://doi.org/10.1016/j.jml.2006.07.011

[48] Sophie Jörg, Jessica Hodgins, and Carol O'Sullivan. 2010. The Perception of Finger Motions. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization - APGV '10.* ACM Press, Los Angeles, California, 129. https://doi.org/10.1145/1836248.1836273

[49] Sophie Jörg, Yuting Ye, Franziska Mueller, Michael Neff, and Victor Zordan. 2020. Virtual Hands in VR: Motion Capture, Synthesis, and Perception. In *SIGGRAPH Asia 2020 Courses.* ACM, Virtual Event, 1–32. https://doi.org/10.1145/3415263.3419155

[50] M-Carmen Juan, Julen Elexpuru, Paulo Dias, Beatriz Sousa Santos, and Paula Amorim. 2023. Immersive virtual reality for upper limb rehabilitation: comparing hand and controller interaction. *Virtual reality* 27, 2 (2023), 1157–1171.

[51] Joann Keyton and Jeff Springston. 1990. Redefining Cohesiveness in Groups. *Small Group Research* 21, 2 (1990), 234–254. https://doi.org/10.1177/1046496490212006

[52] Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. 2013. Drumming in Immersive Virtual Reality: The Body Shapes the Way We Play. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 597–605. https://doi.org/10.1109/TVCG.2013.29

[53] Daesang Kim, Dong-Joong Kim, and Woo-Hyung Whang. 2013. Cognitive Synergy in Multimedia Learning. *International Education Studies* 6, 4 (March 2013), p76. https://doi.org/10.5539/ies.v6n4p76

[54] D. Kodama, T. Mizuho, Y. Hatada, T. Narumi, and M. Hirose. 2023. Effects of Collaborative Training Using Virtual Co-embodiment on Motor Skill Learning. *IEEE Transactions on Visualization and Computer Graphics* 29, 05 (may 2023), 2304–2314. https://doi.org/10.1109/TVCG.2023.3247112

[55] Meng Chun Lam, Haslina Arshad, Anton Satria Prabuwono, Siok Yee Tan, and Seyed Mostafa Mousavi Kahaki. 2018. Interaction Techniques in Desktop Virtual Environment: The Study of Visual Feedback and Precise Manipulation Method. *Multimedia Tools and Applications* 77, 13 (2018), 16367–16398.

[56] James R. Larson. 2010. *In Search of Synergy in Small Group Performance.* Psychology Press.

[57] Alex Leavitt, Brian C. Keegan, and Joshua Clark. 2016. Ping to Win? Non-Verbal Communication and Team Performance in Competitive Online Multiplayer Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16).* 4337–4350. https://doi.org/10.1145/2858036.2858132

[58] Kwan Min Lee. 2004. Presence, Explicated. *Communication theory* 14, 1 (2004), 27–50.

[59] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. 2006. Are Physically Embodied Social Agents Better than Disembodied Social Agents?: The Effects of Physical Embodiment, Tactile Interaction, and People's Loneliness in Human–Robot Interaction. *International Journal of Human-Computer Studies* 64, 10 (Oct. 2006), 962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002

[60] Meng-Yun Liao, Ching-Ying Sung, Hao-Chuan Wang, Wen-Chieh Lin, and Fu-Yin Cherng. 2019. Embodying Historical Learners' Messages as Learning Companions in a VR Classroom. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, 1–6. https://doi.org/10.1145/3290607.3312861

[61] Lorraine Lin and Sophie Jörg. 2016. Need a Hand? How Appearance Affects the Virtual Hand Illusion. In *Proceedings of the ACM Symposium on Applied Perception (SAP '16)*. 69–76. https://doi.org/10.1145/2931002.2931006

[62] Lorraine Lin, Aline Normovle, Alexandra Adkins, Yu Sun, Andrew Robb, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2019. The Effect of Hand Size and Interaction Modality on the Virtual Hand Illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 510–518.

[63] Stefan Liszio, Katharina Emmerich, and Maic Masuch. 2017. The Influence of Social Entities in Virtual Reality Games on Player Experience and Immersion. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*. ACM, Hyannis Massachusetts, 1–10. https://doi.org/10.1145/3102071.3102086

[64] T. Luong, Y. Cheng, M. Mobus, A. Fender, and C. Holz. 2023. Controllers or Bare Hands? A Controlled Evaluation of Input Techniques on Interaction Performance and Exertion in Virtual Reality. *IEEE Transactions on Visualization & Computer Graphics* 29, 11 (nov 2023), 4633–4643. https://doi.org/10.1109/TVCG.2023.3320211

[65] Divine Maloney, Guo Freeman, and Donghee Yvette Wohn. 2020. "Talking without a Voice": Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–25. https://doi.org/10.1145/3415246

[66] Joseph Edward McGrath. 1984. *Groups: Interaction and Performance*. Vol. 14. Prentice-Hall Englewood Cliffs, NJ.

[67] Ryan P McMahan, Alexander Joel D Alon, Shaimaa Lazem, Robert J Beaton, David Machaj, Michael Schaefer, Mara G Silva, Anamary Leal, Robert Hagan, and Doug A Bowman. 2010. Evaluating Natural Interaction Techniques in Video Games. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 11–14.

[68] David McNeill. 2008. *Gesture and Thought*. University of Chicago Press.

[69] Albert Mehrabian. 2017. *Nonverbal Communication*. Routledge.

[70] Nicoleta Meslec and Petru Lucian Curşeu. 2013. Too Close or Too Far Hurts: Cognitive Distance and Group Cognitive Synergy. *Small Group Research* 44, 5 (Oct. 2013), 471–497. https://doi.org/10.1177/1046496413491988

[71] Meta Quest. 2022. VrApi Input API | Oculus Developers. https://developer.oculus.com/documentation/native/android/mobile-vrapi-input-api/. Accessed on 2022-07-12.

[72] Meta Quest Blog. 2019. Introducing Hand Tracking on Oculus Quest—Bringing Your Real Hands into VR. https://www.oculus.com/blog/introducing-hand-tracking-on-oculus-quest-bringing-your-real-hands-into-vr/. Accessed on 2021-06-14.

[73] Michael D. Michalisin, Steven J. Karau, and Charnchai Tangpong. 2004. The Effects of Performance and Team Cohesion on Attribution: A Longitudinal Simulation. *Journal of Business Research* 57, 10 (Oct. 2004), 1108–1115. https://doi.org/10.1016/S0148-2963(03)00042-0

[74] M. Moehring and B. Froehlich. 2011. Effective Manipulation of Virtual Objects within Arm's Reach. In *2011 IEEE Virtual Reality Conference*. 131–138. https://doi.org/10.1109/VR.2011.5759451

[75] Fares Moustafa and Anthony Steed. 2018. A Longitudinal Study of Small Group Interaction in Social Virtual Reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM, Tokyo Japan, 1–10. https://doi.org/10.1145/3281505.3281527

[76] Michael Neff, Yingying Wang, Rob Abbott, and Marilyn Walker. 2010. Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents. In *Intelligent Virtual Agents (Lecture Notes in Computer Science)*, Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova (Eds.). Springer, Berlin, Heidelberg, 222–235. https://doi.org/10.1007/978-3-642-15892-6_24

[77] Kristine L. Nowak and Frank Biocca. 2003. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 12, 5 (Oct. 2003), 481–494. https://doi.org/10.1162/105474603322761289

[78] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (2018). https://doi.org/10.3389/frobt.2018.00114

[79] John Porter III, Matthew Boyer, and Andrew Robb. 2018. Guidelines on Successfully Porting Non-Immersive Games to Virtual Reality: A Case Study in Minecraft. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. 405–415. https://doi.org/10.1145/3242671.3242677

[80] Richard E. Potter and Pierre A. Balthazard. 2002. Understanding Human Interactions and Performance in the Virtual Team. 4, 1 (2002), 1–23. https://www.proquest.com/docview/200016013/abstract/A2B7B1C2090C43AAPQ/1

[81] Mores Prachyabrued and Christoph W. Borst. 2014. Visual Feedback for Virtual Grasping. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 19–26. https://doi.org/10.1109/3DUI.2014.6798835

[82] Daniel Richert, Ammar Halabi, Anna Eaglin, Matthew Edwards, and Shaowen Bardzell. 2011. Arrange-A-Space: Tabletop interfaces and gender collaboration. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. 1495–1500. https://doi.org/10.1145/1979742.1979797

[83] Nick Salamon, Jonathan M. Grimm, John M. Horack, and Elizabeth K. Newton. 2018. Application of Virtual Reality for Crew Mental Health in Extended-Duration Space Missions. *Acta Astronautica* 146 (May 2018), 117–122. https://doi.org/10.1016/j.actaastro.2018.02.034

[84] Eva-Lotta Sallnäs. 2005. Effects of Communication Mode on Social Presence, Virtual Presence, and Performance in Collaborative Virtual Environments. *Presence: Teleoperators and Virtual Environments* 14, 4 (2005), 434–449. https://doi.org/10.1162/105474605774785253

[85] Martijn J. Schuemie, Peter van der Straaten, Merel Krijn, and Charles A.P.G. van der Mast. 2001. Research on Presence in Virtual Reality: A Survey. *CyberPsychology & Behavior* 4, 2 (April 2001), 183–201. https://doi.org/10.1089/109493101300117884

[86] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. 1577–1582. https://doi.org/10.1145/3025453.3025602

[87] Claude Elwood Shannon. 1948. A Mathematical Theory of Communication. *The Bell system technical journal* 27, 3 (1948), 379–423.

[88] Thomas B. Sheridan. 1992. Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environments* 1, 1 (Jan. 1992), 120–126. https://doi.org/10.1162/pres.1992.1.1.120

[89] John Short, Ederyn Williams, and Bruce Christie. 1976. *The Social Psychology of Telecommunications*. Toronto; London; New York: Wiley.

[90] Paul Skalski, Ron Tamborini, Ashleigh Shelton, Michael Buncher, and Pete Lindmark. 2011. Mapping the Road to Fun: Natural Video Game Controllers, Presence, and Game Enjoyment. *New Media & Society* 13, 2 (2011), 224–242. https://doi.org/10.1177/1461444810370949

[91] Harrison Jesse Smith and Michael Neff. 2018. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3173863

[92] Susan G Straus and Joseph E McGrath. 1994. Does the Medium Matter? The Interaction of Task Type and Technology on Group Performance and Member Reactions. *Journal of applied psychology* 79, 1 (1994), 87.

[93] Theresa Jean Tanenbaum, Nazely Hartoonian, and Jeffrey Bryan. 2020. "How Do I Make This Thing Smile?": An Inventory of Expressive Nonverbal Communication in Commercial Social Virtual Reality Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376606

[94] Osku Torro, Jani Holopainen, Henri Jalo, Henri Pirkkalainen, and Antti Lähtevänoja. 2022. How to Get Things Done in Social Virtual Reality – A Study of Team Cohesion in Social Virtual Reality–Enabled Teams. In *Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2022.057

[95] Unity Asset Store. 2022. Oculus Integration. https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022. Accessed on 2022-06-29.

[96] Roshan Venkatakrishnan, Rohith Venkatakrishnan, Balagopal Raveendranath, Christopher C. Pagano, Andrew C. Robb, Wen-Chieh Lin, and Sabarish V. Babu. 2023. How Virtual Hand Representations Affect the Perceptions of Dynamic Affordances in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2258–2268. https://doi.org/10.1109/TVCG.2023.3247041

[97] Jan-Niklas Voigt-Antons, Tanja Kojic, Danish Ali, and Sebastian Möller. 2020. Influence of hand tracking as a way of interaction in virtual reality on user experience. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–4.

[98] Spyros Vosinakis and Panayiotis Koutsabasis. 2018. Evaluation of Visual Feedback Techniques for Virtual Grasping with Bare Hands Using Leap Motion and Oculus Rift. *Virtual Reality* 22, 1 (March 2018), 47–62. https://doi.org/10.1007/s10055-017-0313-4

[99] Tzu-Yang Wang, Yuji Sato, Mai Otsuki, Hideaki Kuzuoka, and Yusuke Suzuki. 2019. Effect of Full Body Avatar in Augmented Reality Remote Collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1221–1222. https://doi.org/10.1109/VR.2019.8798044

[100] Yingying Wang, Jean E. Fox Tree, Marilyn Walker, and Michael Neff. 2016. Assessing the Impact of Hand Motion on Virtual Character Personality. *ACM Transactions on Applied Perception* 13, 2 (March 2016), 9:1–9:23. https://doi.org/10.1145/2874357

[101] Peter H. Waxer. 1977. Nonverbal Cues for Anxiety: An Examination of Emotional Leakage. *Journal of Abnormal Psychology* 86, 3 (1977), 306–314. https://doi.org/10.1037/0021-843X.86.3.306

[102] Barbara A. Wech, Kevin W. Mossholder, Robert P. Steel, and Nathan Bennett. 1998. Does Work Group Cohesiveness Affect Individuals' Performance and Organizational Commitment?: A Cross-Level Examination. *Small Group Research* 29, 4 (Aug. 1998), 472–494. https://doi.org/10.1177/1046496498294004

[103] Carolyn Y. Wei. 2006. Not Crazy, Just Talking On The Phone: Gestures And Mobile Phone Conversations. In *2006 IEEE International Professional Communication Conference*. 299–307. https://doi.org/10.1109/IPCC.2006.320363

[104] Nkenge Wheatland, Yingying Wang, Huaguang Song, Michael Neff, Victor Zordan, and Sophie Jörg. 2015. State of the Art in Hand and Finger Modeling and Animation. *Computer Graphics Forum* 34, 2 (2015), 735–760. https://doi.org/10.1111/cgf.12595

[105] Rand R Wilcox. 2012. *Introduction to robust estimation and hypothesis testing*. Academic press.

[106] Yuanjie Wu, Yu Wang, Sungchul Jung, Simon Hoermann, and Robert Lindeman. 2021. Using a Fully Expressive Avatar to Collaborate in Virtual Reality: Evaluation of Task Performance, Presence, and Attraction. *Frontiers in Virtual Reality* 2 (April 2021). https://doi.org/10.3389/frvir.2021.641296

[107] Amal Yassien, ElHassan B. Makled, Passant Elagroudy, Nouran Sadek, and Slim Abdennadher. 2021. Give-Me-A-Hand: The Effect of Partner's Gender on Collaboration Quality in Virtual Reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Article 392, 6 pages. https://doi.org/10.1145/3411763.3451601

[108] Yiran Zhang, Huyen Nguyen, Nicolas Ladevèze, Cédric Fleury, and Patrick Bourdot. 2022. Virtual Workspace Positioning Techniques during Teleportation for Co-located Collaboration in Virtual Reality using HMDs. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 674–682. https://doi.org/10.1109/VR51125.2022.00088