This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3067783, IEEE Transactions on Visualization and Computer Graphics

The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments

Sahar Aseeri and Victoria Interrante, *Senior Member, IEEE*

Fig. 1: The three types of avatar representations compared in our real time social virtual environment (left to right): No_Avatar, Scanned_Avatar, and Real_Avatar.

Abstract—Current avatar representations used in immersive VR applications lack features that may be important for supporting natural behaviors and effective communication among individuals. This study investigates the impact of the visual and nonverbal cues afforded by three different types of avatar representations in the context of several cooperative tasks. The avatar types we compared are No_Avatar (HMD and controllers only), Scanned_Avatar (wearing an HMD), and Real_Avatar (video-see-through). The subjective and objective measures we used to assess the quality of interpersonal communication include surveys of social presence, interpersonal trust, communication satisfaction, and attention to behavioral cues, plus two behavioral measures: duration of mutual gaze and number of unique words spoken. We found that participants reported higher levels of trustworthiness in the Real_Avatar condition compared to the Scanned_Avatar and No_Avatar conditions. They also reported a greater level of attentional focus on facial expressions compared to the No_Avatar condition and spent more extended time, for some tasks, attempting to engage in mutual gaze behavior compared to the Scanned_Avatar and No_Avatar conditions. In both the Real_Avatar and Scanned_Avatar conditions, participants reported higher levels of co-presence compared with the No_Avatar condition. In the Scanned_Avatar condition, compared with the Real_Avatar and No_Avatar conditions, participants reported higher levels of attention to body posture. Overall, our exit survey revealed that a majority of participants (66.67%) reported a preference for the Real_Avatar, compared with 25.00% for the Scanned_Avatar and 8.33% for the No_Avatar. These findings provide novel insight into how a user's experience in a social VR scenario is affected by the type of avatar representation provided.

Index Terms—interpersonal communication, trust, communication satisfaction, social presence, behavioral cues, virtual environment.

1 Introduction

As virtual reality (VR) technology assumes an increasingly prevalent role in applications requiring interpersonal interactions among co-located users, additional work is needed to provide avatar representations that optimally support interpersonal communication experiences.

The research we present in this paper is driven by the expressed needs of our architecture colleagues for a system that supports collaborative discussion between multiple stakeholders in the context of an immersive design review. Each person wears a head-mounted display (HMD) and can freely walk within a large, open tracked area in order to be able to evaluate a proposed plan from a first-person perspective, and all users are immersed together to facilitate the discussion about specific aspects of the design. Similar requirements are common to design review scenarios in many other fields, as well as to other collaborative immersive applications involving nuanced communication among professional partners.

• *Sahar Aseeri and Victoria Interrante are with University of Minnesota. E-mail: {aseer002, interran}@umn.edu*

Manuscript received 09 Sept. 2020; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

The minimum requirements identified by our colleagues as essential to support their needs were (1) the ability to see where each person is standing, in the co-located real and virtual environments, to avoid people being afraid of bumping into each other while walking around; (2) the ability to see where each person is directing their attention to facilitate the social process of engaging others in conversation; and (3) the ability to support deictic gestures (e.g. pointing) and basic communicative head gestures, such as nodding in agreement. For those purposes, a minimalist avatar representation consisting of a rendering of the headset and controllers could potentially be sufficient, in conjunction with some sort of simple augmentation to facilitate identifying each unique individual. However, we felt that it could be possible to better support some more nuanced aspects of deliberative interpersonal communication and engagement through the use of a more sophisticated avatar representation. In particular, we suspected that the use of video-see-through technology might afford increased communication effectiveness—even under conditions in which the eyes and most of the upper face would be obscured by the HMD—by providing access to subtle aspects of facial or body expressions that current real-time avatar animation systems cannot yet robustly support.

To better understand the relative merits of the various best-available embodiment options, we ran an experiment that compared multiple qualitative and quantitative measures of interpersonal communication, including social presence, interpersonal trust, communication satisfac-

tion, attention to behavioral cues, mutual gaze, and unique word counts, between the following three conditions: Real_Avatar, Scanned_Avatar, and No_Avatar (see Figure 1). To maximize the power of our experiment, we decided to use a within-subjects design, in which each participant was embodied in each of the three avatar conditions and conversed with a confederate represented similarly. To minimize the potential of behavioral anchoring and/or carry-over effects, we structured the engagement with each avatar in the context of a different communication-centered task, with the pairing of avatar to task randomized between participants.

Our experiment found significantly better outcomes on multiple measures with the use of the 3D video-based avatar, compared to the scanned avatar model and the animated HMD and controllers only. Specifically, we found that in the Real_Avatar condition, participants spent more time attempting to engage in mutual gaze behaviors, showed a greater level of attentional focus on facial expressions, and reported higher levels of interpersonal trust. Additionally, a significant majority of our participants reported an overall preference for the Real_Avatar. Our results confirm prior work showing benefits in the use of full-body vs. minimal avatars, and extend prior work by showing that the additional expressivity afforded by the video representation remains relevant and impactful even when the entire upper half of the conversational partner's face is obscured by the HMD.

In section 2 of this paper we provide an overview of related work in virtual social communication, followed by an overview then detailed description of our experiment in sections 3 and 4. Section 5 presents the results of our experiment, along with some brief discussion for each result. General discussion is given in section 6, followed by conclusions and future work in section 7.

2 Related Work

Many prior studies have investigated how different forms of avatar representation influence interpersonal communication in VR.

In the realm of desktop applications, Bente et al. [2] compared ratings of social presence, interpersonal trust, and perceived communication quality among pairs of remote participants collaborating on a management decision task using text, audio, audio+video, or representation as a low-fidelity or high-fidelity avatar seen on a computer monitor. They found significantly lower ratings in the text-only mode, but no significant differences between any of the non-text conditions. Similarly, Lockwood and Massey [16] compared participants' impressions of the communicative ability, benevolence, and trustworthiness of a project leader presenting advice via audio, video, or as an informally or formally dressed avatar, and found that ratings on all measures were significantly lower in the audio-only condition, but not significantly different between any of the other conditions. However, Riegelsberger et al. [19] studied the found a significantly greater tendency for participants to seek advice from advisors represented via audio or video than via a photo+text or a 2D animated avatar, and the negative impact of the avatar representation was particularly strong when the advisor was an expert versus non-expert.

In a 3D teleconferencing system, Jo et al. [14, 15] found that participants were less likely to seek advice from, and slower to trust, a remote assistant represented as a computer-generated (CG) character avatar, than as an articulated scanned avatar when both assistants were concurrently available. Trustworthiness ratings of the CG avatar were particularly low when it appeared in the context of a stereo video-see-through view of a real-world background versus a CG background. However, in a remote interview task with a known remote conversant, they found that participants who interacted with the CG avatar reported higher levels of co-presence than those who interacted with the scanned avatar, despite its higher ratings of visual realism and naturalness; lowest levels of co-presence were reported in the 2D video condition.

Altogether, these results suggest a complex relationship between visual realism, trust, and co-presence in avatar-mediated communication via non-immersive technologies. Credibility seems to suffer when the avatar representation lacks realism, while co-presence suffers when a more realistic-looking avatar model lacks realistic facial animation.

In an augmented reality context, Yoon et al. [28] found no significant main effect of character style (realistic vs. cartoon) on aggregate ratings of social presence, despite most users expressing an overall preference for seeing their partner represented as a realistic whole body avatar. They found that co-presence rankings were consistently lower when avatars were incompletely represented as hands only, however.

Additionally, multiple authors have compared various measures of interpersonal engagement using full or partial avatars in immersive VR with real world interaction as a control condition. Greenwald et al. [11] compared participants' experiences during gameplay in the real world to gameplay in VR using a minimal avatar represented as an anthropomorphized HMD and two disconnected, rigid-body 3D hand models. They found that the absence of body, hand, and facial gestures in the VR condition was problematic in some contexts, but not when participants' attention was more focused on the activity than on their partner. Several of their participants reported that not being able to see their partner had some advantages, e.g. in reducing their social anxiety. However, Pan and Steed [18] found that participants who interacted using fully embodied avatars exhibited higher levels of demonstrative trust in their partner than participants who interacted using a controllers-only embodiment, and levels of self-reported trust that were not significantly different than participants in a face-to-face condition. Their fully-embodied participants also completed cooperative tasks as fast as the participants interacting face-to-face, and faster than participants in the controllers-only condition. Likewise, Heidicker et al. [13] found that participants reported significantly higher ratings of presence when they solved a collaborative task while embodied in a full motion-tracked avatar than with motion-tracked head and hands only; on other measures, including co-presence and social presence, no significant differences were found however.

Similarly, Smith and Neff [25] observed qualitatively similar communication patterns between face-to-face and full body avatar conditions across two different negotiation and consensus-building tasks, but found a significant drop-off in several measures when participants were not embodied, primarily due to the inability of nonverbal communication. However, they found significantly lower ratings of social awareness in the full-body avatar condition than in face-to-face interaction, due to the avatars' lack of eye movement and facial expression cues. Overall, these papers [13, 18, 25] show that some aspects of interpersonal engagement can be supported as well using full body avatars in VR as in real life, although lack of eye movement and facial expressions can have some negative effects. There is less evidence to support the effectiveness of minimal embodiments, though they may be sufficient in some cases.

In the realm of architectural design, Abbas et al. [1] compared the strengths and weaknesses of using immersive VR (IVR) versus in-person meetings for stakeholder discussions about construction projects. Groups of participants conferred about optimal design options, first using traditional face-to-face discussion with the building information modeling (BIM) information displayed on a computer monitor, and then while immersed as non-articulated 3D avatars in a 3D virtual model of the designed spaces with the BIM information superimposed. They reported that the face-to-face discussion mode was rated significantly higher than the IVR mode in communication accuracy (less likely to be misunderstood by others) and communication appropriateness, primarily due to the inability in VR to know if others were paying attention to what one was saying. Ratings were equivalent between the VR and real world conditions, however, for: discussion quality, communication richness, and communication openness. They also found no significant differences in any of their measures of the quality of the discussions. The authors conclude that IVR-based design reviews have enormous potential, but that improvements are needed in supporting non-verbal communication cues within virtual environments.

Reproducing accurate, realistic facial expressions on avatars representing users who are wearing HMDs remains a challenging problem. In recent years, Roth et al. [20], Wei et al. [27], and others, have developed systems that use inward-facing sensors to capture information about a user's gaze direction and related facial muscle poses underneath and around the HMD, and use that information to animate corresponding facial expressions on the user's avatar. Although such technologies

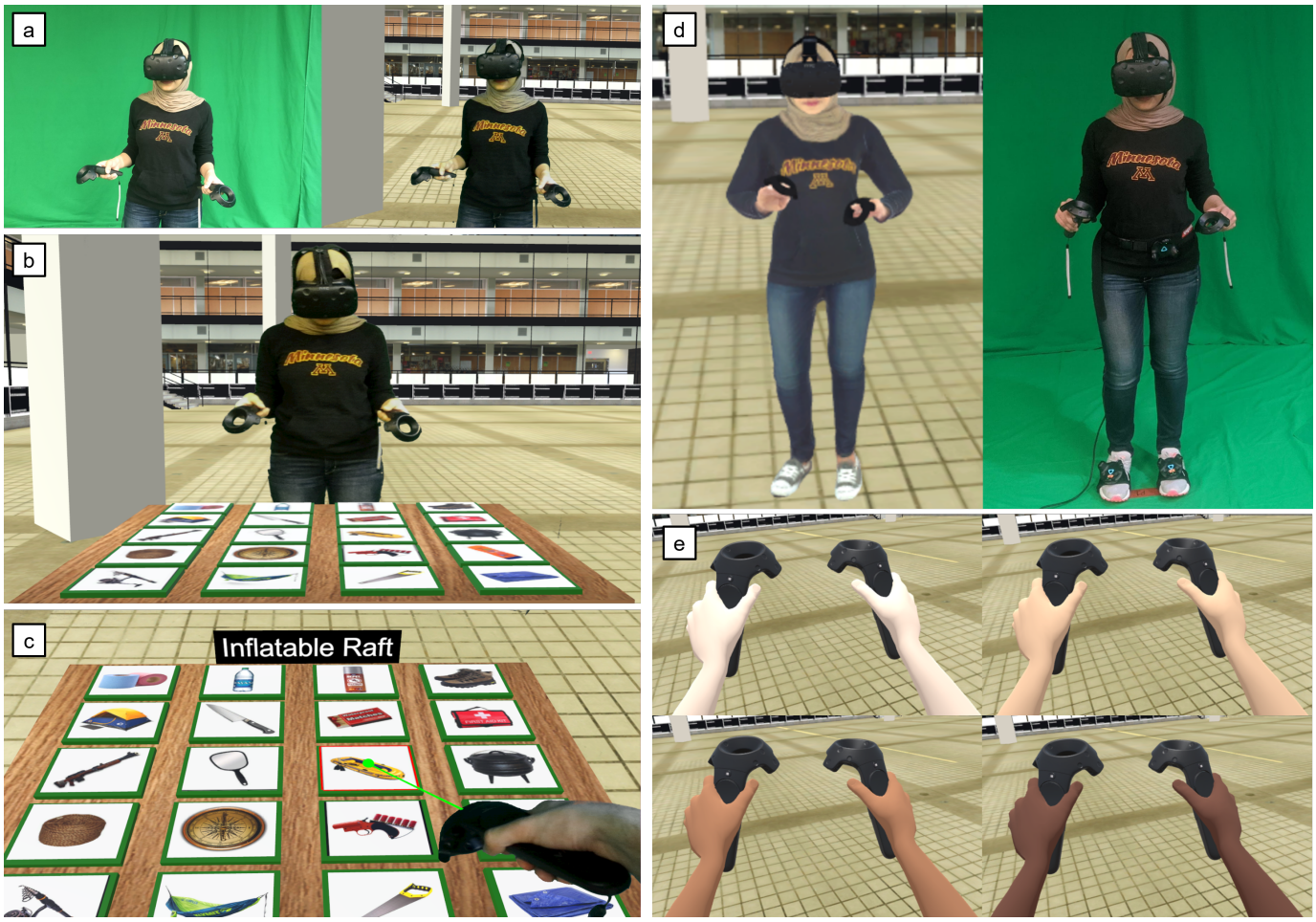

Fig. 2: The avatar representations for both participants (the IP and PE): the Real_Avatar representation is on the left side of the image and the Scanned_Avatar is on the right side. (a) The captured image of the PE with the green screen background subtraction. (b) The IP's view of the PE during the Survival Item task. (c) IP's view of themselves during the Survival Item task. (d) The PE's scanned virtual model (left) with HMD and 5 Vive trackers for full-body real-time synchronization. (e) The IP's virtual body from a first-person perspective with different skin color options.

are not yet widely available, they may soon offer promising possibilities for improved communication outcomes. Garau et al. [10] have shown that the quality of an avatar's eye gaze behavior can significantly affect perceived communication quality during dyadic social interaction in a shared, immersive virtual environment.

Similarly to our work, Cho et al. [8] compared participants' ratings of social presence in a shared virtual environment when their co-located confederate was depicted using real-time texture-mapped RGBD volumetric capture verus as an animated pre-scanned avatar whose mouth motion was generated by interpolation between three pre-scanned visemes. However, their confederate was not wearing an HMD. They found that social presence was significantly higher with the RGBD representation, and many of their participants complained about the artificial appearance of the pre-scanned alternative. Unlike Cho et al. [8], our target application requires all users to wear headsets. Therefore, we follow the example of Salzmann and Froehlich [23] and have all of our avatars similarly depicted wearing HMDs.

Considering everything that is known from prior related work [2, 8, 13–16, 18, 19, 25, 28], we felt fairly confident in seeing clear advantages from using a full-body avatar, over a minimal avatar representation. Yet our core question remained: Would we see an advantage in measures of social presence, trust, and communication effectiveness from using a video-see-through approach over scanned avatars to represent users to each other in a shared virtual environment context when all the users are wearing headsets?

3 Experiment Setup

We designed our experiment to evaluate interpersonal communications between two physically co-located users in a shared virtual environment. The experiment took place within an approximately 2-3m x 4-5m curved section of our 30'x29' lab space, within which the floor and the 10' high walls were completely covered by large sheets of green-screen fabric. We used two laptops to immerse the two users (the participant and the experimenter/confederate) in the same virtual scene—a photorealistic replica of our architecture building's indoor courtyard. The virtual environment was modeled in Sketchup and textured using photographs of the real place. It was rendered using the Unity game engine (Unity 2018.1.5) on MSI GT72 Dominator Pro-445 gaming laptops, each equipped with an 8 GB NVIDIA Geforce GTX 980M graphics card and a 2.80 GHz Intel Core i7 4980 HQ processor, with 32 GB of memory and a 512 GB SSD. Each user wore an HTC Vive headset and held a Vive controller in each hand. All objects were tracked in a common world coordinate system established by a single pair of Lighthouse emitters.

To guarantee consistency, we set up the tracking space on one laptop and copied the resulting config, vrchap and json files to the other. We used the Photon Unity Networking plugin to make all of the tracking data available to both systems. The participant's headset had a ZED mini stereo camera attached to the front, which was turned off when not in use. The ZED features dual 2K image sensors that capture video over a 110°field of view (FOV) at 60Hz using a fixed camera separation of 63mm and a USB-C port for streaming this output. Each user also wore

a Sennheiser XSW 1-ME2 wireless lavalier microphone that recorded their audio to a PreSonus AudioBox for subsequent analysis.

Our experiment compared multiple measures of communication effectiveness (described in more detail in section 4.2) between the invited participant (IP) and the primary experimenter (PE) across three different conditions of body representation. In each case, the participant saw themselves and the experimenter represented using the same type of representation, while the experimenter always saw the participant represented as a floating HMD with two controllers across all of the experiment conditions. Below we describe what participant saw in the No_Avatar, Scanned_Avatar, and Real_Avatar conditions.

No_Avatar condition: The IP saw themselves and the PE each represented as a floating HMD with two controllers, conveying the tracked positions and orientations of each user's head and hands in real time.

Scanned_Avatar condition: The IP saw themselves and the PE each represented as a full body virtual avatar. The PE's body model was obtained using a high-quality 3D body scan from Me3D2, a commercial 3D scanning company with a 360°photo booth in our city (see Figure 2 (d)). We rigged this model using Maya 2018 and scaled it to match the PE's width and height to ensure we had the correct size. The PE wore two Vive trackers on their feet and a third around their torso, and we used the FinalIK Unity plugin to apply an Inverse Kinematic (IK) solver to ensure that the pose of the scanned model accurately mimicked the PE's movements in real time. We did not attempt to animate the face of the scanned avatar. For each IP, we used MakeHuman to create a gender-matched avatar, re-sized based on the IP height, with different options of skin color (see Figure 2 (e)).

Real_Avatar condition: The IP saw stereo video images of everything located in front of the green screen fabric, including their own body and the body of the PE, integrated into the context of the virtual environment (see Figure 2 (b, c)). We used the ZED SDK Unity Plugin to adjust the chroma key to remove the green screen from the captured image (see Figure 2 (a)). Throughout the experiment, the IP was situated in the lab space such that the area covered by the green screen material filled their entire FOV across all of the natural rotational positions of their head. As the ZED mini camera has a smaller FOV (90°H x 60°V) than the HTC Vive (∼110°H x ∼110°V), less of the periphery of the virtual environment was visible in this condition.

4 Experiment Overview

4.1 Participants

We recruited 36 participants (16 F, 20 M), ranging from 18 to 27 years of age (M = 20.94, SD = 2.67). Participants were a mix of graduate and undergraduate students, recruited via posted flyers, and all had corrected-to-normal vision. Each participant was given a $30 gift card for participating.

4.2 Metrics

We collected multiple subjective measures via surveys of social presence, interpersonal trust, interpersonal communication satisfaction, and attention to behavioral cues. All survey responses were collected on a 7-point Likert scale with "Strongly Disagree" at option 1 and "Strongly Agree" at option 7. Our objective measures included the percentage of time the participant spent attempting to engage in mutual gaze with the experimenter and the count of unique words they spoke while conversing. We also conducted an exit survey to learn more about the user's experience of each avatar type, including which representation they preferred overall and why.

4.2.1 Social Presence

Social presence is a measure used to evaluate the level of "being there" with another person and having the ability to understand the partner's thoughts and emotions [3, 17]. To measure social presence we used 30 questions from the Networked Minds Social Presence Inventory (2002). These questions spanned five sub-scales: co-presence, attentional engagement, emotional contagion, comprehension, and behavioral interdependence, with each sub-scale containing 6 prompts [4, 5]. The six questions we omitted from the original survey pertained to communication between remotely located partners, and to the evoked emotions of sadness and nervousness, which were not applicable in our communication scenarios. A full list of the used and unused questions is provided in Appendix A (supplemental materials).

4.2.2 Interpersonal Trust

Our interpersonal trust survey was developed in–house, and included eight prompts: "I often believed what my partner was telling me", "I was often comfortable dealing with my partner", "I was often relying on my partner's advice", "I was often sharing personal information with my partner", "Overall, I would trust my partner", "I was often convinced by my partner to change my thoughts", "I often felt confidence in performing tasks with my partner", and "I was often feeling secure with my partner."

4.2.3 Interpersonal Communication Satisfaction

We used eleven questions from the 19-item Interpersonal Communication Satisfaction Scale by Hecht (1978) [12]. The questions we chose focused on communication effectiveness, conversational flow, satisfaction/enjoyment, and rapport. The questions we omitted were either redundant to the questions we had already selected, or pertained to topics that we felt were outside the scope of our scenarios. Some examples of omitted questions are: "Nothing was accomplished" and "I had something else to do." A full list of the selected and omitted items is provided in Appendix B (supplemental materials).

4.2.4 Communication Behavioral Cues

We also used the attention to behavioral cues questionnaire constructed by Roth et al. [21]. This survey consists of two parts. The first part asked the users to indicate how much attention they paid to different behavioral cues, including gesture, body posture, facial expression, speech, head gaze movement, body proxemics, and others. The second part asked users how much they had missed the same behavioral cues, referring to the absence or lack of quality of these cues [21, 22].

4.2.5 Mutual Gaze

Mutual gaze is an objective measure that can be used to assess an implied social presence between individuals [7]. Because our headset did not have built-in eye tracking, we adapted the technique described by Clay et al. (2019) [9] to use the forward-facing direction of the participant's headset. To account for the fact that people do not always orient their eyes in the exact same direction as their head, we did some pilot experiments ahead of time to determine the appropriate size for a sphere collider that would typically contain the head of one's conversational partner when one was looking towards them. At every frame (60x/sec) we kept track of whether the approximated gaze ray from the participant's headset intersected the sphere collider around the experimenter's headset, and from this data we calculated the total percentage of time the participant attempted to engage in mutual gaze.

4.2.6 Unique Words

People use utterance for communication, and multiple features can be extracted to give insight into the quality of the communication and related social processes [26]. On the advice of Professor Ben Munson, chair of the Department of Speech-Language-Hearing Sciences at the University of Minnesota, we decided to use the *unique word count* metric—the number of distinct words spoken over a specified period of time-as our text analysis measure to quantify conversational activity and fluency. Researchers have previously noted an association between a higher number of unique words spoken and more fluent interpersonal communication [24]. We used a commercial service to transcribe the recorded audio to text for each participant over two time-limited interaction tasks. We then separated the participants' utterances from those of the PE, and wrote a short program to compute the unique word count.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3067783, IEEE Transactions on Visualization and Computer Graphics

4.3 Hypotheses

The following hypotheses were formulated based on our main research question, which was: to what extent and under what conditions do the most promising different methods of representing co-located users to each other affect the quality of their interpersonal communications in a shared virtual environment?

H1 Experiencing interpersonal communications with a Real_Avatar will enable users to feel a stronger sense of social presence and mutual engagement with their partners in a shared virtual environment.

H2 Experiencing interpersonal communication with a Real_Avatar will allow users to feel more interpersonal trust with their partners in a shared virtual environment.

H3 Experiencing interpersonal communication with a Real_Avatar will result in users feeling a higher level of satisfaction in their communications with their partners in a shared virtual environment.

4.4 Study Design

Our experiment involved three avatar types and three tasks. We paired a different task with each avatar exposure to avoid carry-over effects. Thus, each participant experienced each avatar and each task exactly once, with the assignment of avatar to task counterbalanced between participants using a 3x3 Latin Square. This generated three blocks of twelve participants each. The six possible presentation orders of each avatar/task combination were then counterbalanced within each block. The experimenter was the participant's partner for each task.

We chose three tasks that had been used in previous research on communication effectiveness: Conversation Cards [14, 15], Survival Items [5, 8], and Charades [11]. Each task had a prescribed duration of seven minutes, and was carried out with the participant and experimenter standing and facing each other at a comfortable conversational distance of about 1m.

The **Conversation Cards task** is an activity in which IP is encouraged to share personal information and stories by answering a series of general life questions. The task's goal is to evoke a deep conversation between the IP and PE through sharing information the IP knows fluently. The IP and PE took turns asking questions. The **Survival Items task** is an activity in which the PE presents the IP with a story and the IP must choose five items to use to survive the described scenario. The items were presented as cards laid out on a virtual table, which the participant selected using their controller (See Figure 2 (b, c)). After the IP made their choice, the PE tried to convince them to change their decision by presenting new facts. Each IP experienced two different stories with the same 20 items. The **Charades task** involved guessing different verbs that were acted out, such as reading, writing, and cooking. The IP and PE took turns acting and guessing words.

4.5 Procedure

Participants came individually to our lab, where they underwent screening tests for visual acuity and stereo vision, which all passed. They then signed a consent form to participate in the experiment and filled out a demographic survey. Next, they completed Eysenck's Personality Inventory (EPI) [6] to quantify their extroversion/introversion traits. After that, they read the experiment instructions, and prepared 5–7 questions for the conversations cards task, and 8–10 words to act out for charades. The experiment duration was 1.5 hours, including 20 minutes for each of the three conditions, with 10-minute breaks in between. The participant took off the headset after each exposure condition and completed questionnaires evaluating their experience of the interpersonal communication in the virtual social environment. At the end of the experiment, the participant filled out an exit survey including open-ended questions to elicit feedback about their experience and to ask which was their preferred avatar representation.

5 Results and Discussion

We used the IBM SPSS tool to perform the data analysis and the R programming language to plot the results. For readability, we have grouped the results and discussion for each metric in the sections below. All the results are summarized in table 1.

5.1 Social Presence

5.1.1 Results

Because we had removed some of the questions, we conducted a reliability test on the Networked Minds Social Presence Inventory by calculating Cronbach's alpha, which yielded an excellent internal consistency of $\alpha = 0.91$. Aggregate social presence was computed for each participant as the median of their Likert responses over all of the questions. A Shapiro-Wilk test of this data showed a significant departure from normality, $W(108) = 0.86$, $p < .01$, so a non-parametric within-subjects Friedman test was conducted which found no significant differences in aggregate social presence between the three avatar conditions. We also computed the per-participant median scores for each of the Networked Minds Social Presence Inventory sub-scales: co-presence, attentional engagement, emotional contagion, comprehension, and behavioral interdependence. A significant difference was found for co-presence $\chi^2(2) = 17.20$, $p < .01$, $W = 0.24$. Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction were used to compare all pairs of groups. Median (IQR) levels for the No_Avatar, Scanned_Avatar, and Real_Avatar conditions were 6 (5 to 6.75), 7 (6 to 7), and 6.5 (5.62 to 7), respectively. The test indicated that there was a difference between Real_Avatar and No_Avatar (Z = -2.77, p = .02, r = -0.33) and the mean rank of Real_Avatar was higher than No_Avatar. Also, there was a difference between Scanned_Avatar and No_Avatar (Z = -2.70, p = .02, r = -0.32) and the mean rank of Scanned_Avatar was higher than No_Avatar. However, there was no significant difference between the Real_Avatar and Scanned_Avatar.

The results for the sub-scales attentional engagement, emotional contagion, comprehension, and behavioral interdependence indicated non-significant differences between avatar types. See table 1, which includes a summary of all the p-values that were computed for the social presence measurements. Also, Figure 3 presents the individual scores and aggregated means and medians for all the social presence sub-scales.

5.1.2 Discussion

Based on the overall social presence results, we have to reject our hypothesis **H1**, which is that experiencing interpersonal communication with a Real_Avatar will enable users to feel a stronger sense of social presence and mutual engagement with their partners in a shared virtual environment. However, when we look at the results of the social presence sub-scale co-presence, which is defined as two or more people feeling as if they are together in the same place and time, we can see

Fig. 3: The means and medians of the social presence sub-scales metrics. Co-presence shows a significant difference between the conditions.

Metric	Factor		p-value	Post-hoc
Social Presence	Overall social presence		$\chi^2(2) = 4.16$, $p = .13$	*ns*[*]
	Co-presence		$\chi^2(2) = 17.20$, $p < .01$	Real_Avatar, Scanned_Avatar > No_Avatar
	Attentional engagement		$\chi^2(2) = 4.83$, $p = .09$	*ns*
	Emotional contagion		$\chi^2(2) = 3.53$, $p = .17$	*ns*
	Comprehension		$\chi^2(2) = 2.49$, $p = .29$	*ns*
	Behavioral interdependence		$\chi^2(2) = 0.02$, $p = .99$	*ns*
Interpersonal Trust			$\chi^2(2) = 11.81$, $p < .01$	Real_Avatar > Scanned_Avatar, No_Avatar
Interpersonal Communication Satisfaction			$\chi^2(2) = 2.00$, $p = .37$	*ns*
Behavioral Cues	How much did you attend to <cue>?	Gesture	$\chi^2(2) = 1.12$, $p = .57$	*ns*
		Body posture	$\chi^2(2) = 6.96$, $p = .03$	Scanned_Avatar > Real_Avatar, No_Avatar
		Facial expression	$\chi^2(2) = 13.25$, $p < .01$	Real_Avatar > No_Avatar
		Speech	$\chi^2(2) = 0.54$, $p = .76$	*ns*
		Head-Gaze	$\chi^2(2) = 1.33$, $p = .52$	*ns*
		Body proxemics	$\chi^2(2) = 1.01$, $p = .60$	*ns*
	How much did you miss <cue>?	Gesture	$\chi^2(2) = 2.88$, $p = .24$	*ns*
		Body posture	$\chi^2(2) = 14.21$, $p < .01$	No_Avatar > Real_Avatar, Scanned_Avatar
		Facial expression	$\chi^2(2) = 12.33$, $p < .01$	No_Avatar, Scanned_Avatar > Real_Avatar
		Speech	$\chi^2(2) = 2.88$, $p = .24$	*ns*
		Head-Gaze	$\chi^2(2) = 1.12$, $p = .57$	*ns*
		Body proxemics	$\chi^2(2) = 0.72$, $p = .70$	*ns*
Mutual Gaze	Conversation Cards		$H(2) = 6.62$, $p = .04$	Real_Avatar > Scanned_Avatar
	Survival Items		$F(2, 33) = 0.79$, $p = .46$	*ns*
	Charades		$F(2, 33) = 4.68$, $p = .02$	Real_Avatar > No_Avatar
Unique Words	Conversation Cards		$F(2, 33) = 1.19$, $p = .32$	*ns*
	Survival Items		$F(2, 33) = 3.25$, $p = .05$	No_Avatar > Real_Avatar

[*] Not Significant.

Table 1: A summary of all the p-value results and the post-hoc tests that have been conducted in the experiment for all the study metrics

significant differences between the avatar types. The lower score on the co-presence subscale for the No_Avatar condition is consistent with the findings from previous related work. However, we were surprised to not see a difference between Real_Avatar and Scanned_Avatar conditions. We had expected that the Real_Avatar, with all of its visual details and behavioral cues, would evoke a higher sense of co-presence than the Scanned_Avatar. However, it may be that the existence of the avatar body in both conditions was sufficient, regardless of the other cues.

Based on participants' comments in the exit survey, the natural body movement and facial expressions of the Real_Avatar seemed to play a main role in promoting social presence and engagement. However, some participants reported that the Scanned_Avatar seemed more compatible with the 3D virtual world, which made the whole social experience more immersive. One user mentioned that the Scanned_Avatar was accurate enough to be suited for different types of games. Four participants favored No_Avatar for social presence and engagement. One of them who experienced No_Avatar with the Survival Items task, and another who experienced No_Avatar with the Charades task, said that they focused their attention on the movements and social cues of their partner more in this case than with the other two avatars, which made them feel more engaged in the social environment.

5.2 Interpersonal Trust

5.2.1 Results

We conducted a reliability test on the interpersonal trust metric to check the internal consistency of the questions, which yielded an acceptable internal consistency of $\alpha = 0.76$. A median score was computed for each participant across the eight questions. A Shapiro-Wilk test of these data showed a significant departure from normality, $W(108) = 0.96$, $p < .01$. A non-parametric within-subjects Friedman test showed a significant difference in interpersonal trust between avatar types $\chi^2(2) = 11.81$, $p < .01$, $W = 0.16$. Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction were used to compare all pairs of groups. Medians (IQR) for the No_Avatar, Scanned_Avatar, and Real_Avatar were 6 (4.6 to 6), 5.25 (4.1 to 6) and 6 (5 to 6), respectively. The test indicated a difference between Real_Avatar and Scanned_Avatar (Z = -3.54, $p < .01$, r = -0.42) and the mean rank of Real_Avatar was higher than Scanned_Avatar. There was also a difference between Real_Avatar and No_Avatar (Z = -2.57, p = .03, r = -0.3) and the mean rank of Real_Avatar was higher than No_Avatar. However, the differences between the Scanned_Avatar and No_Avatar were not significant.

5.2.2 Discussion

Based on these results we can accept our hypothesis **H2**, which is that experiencing interpersonal communication with a Real_Avatar will allow users to feel more interpersonal trust with their partners in a shared virtual environment. Our expectation was that the Real_Avatar would evoke a higher sense of interpersonal trust because the live camera view gives an authentic sense of the partner's affective state.

The interpersonal trust question in the exit survey gives several reasons the majority of the participants trusted the Real_Avatar more than the other two avatars. First, because it is a real image of a person that has partial facial expressions and natural body movements. Most participants may be more inclined to trust a person they can see than a hidden person represented by a CG avatar. The second reason given is that it is easier to establish social communication with the Real_Avatar. One user said, "The most trustful avatar is the Real_Avatar, because it was easier to communicate with and therefore easier to maintain a conversation and build trust with." However, two participants felt that the Real_Avatar and Scanned_Avatar were the same in terms of trustworthiness, and two other participants felt that the Scanned_Avatar was the most trustworthy because they preferred to talk to an unreal person. Three participants favored No_Avatar; one said about the No_Avatar in the Conversation Cards task, "I felt like the No_Avatar was most comfortable because that was the one used in the most relaxing and comfortable game, the conversation." The other two users preferred to speak with No_Avatar because it did not have a face, which one of them said led them to think that the avatar would not remember anything they said.

5.3 Interpersonal Communication Satisfaction

5.3.1 Results

Because we had removed some questions, we conducted a reliability analysis on the Interpersonal Communication Satisfaction Inventory and found good internal consistency $\alpha = 0.87$. A Shapiro-Wilk test of that data showed a significant departure from normality, $W(108) = 0.85$, $p < .01$. A non-parametric within-subjects Friedman test was conducted and the result was not significant; see table 1 for more details.

5.3.2 Discussion

Because the statistical analysis shows that there are no differences between the avatar types for the Interpersonal Communication Satisfaction metric, we must reject hypothesis **H3**, which is that experiencing interpersonal communication with a Real_Avatar will result in users feeling a higher level of satisfaction in their communications with their partners in a shared virtual environment. However, when we asked participants in the exit survey about which avatar was the easiest and most satisfying to communicate with in the social environment, 25 of them were the most satisfied with the Real_Avatar compared to the other avatar types. One participant mentioned that Real_Avatar was the easiest avatar to communicate with because he felt like he could better understand the avatar and how it reacts in certain situations. Another said the presence of facial expressions and non-verbal cues help to have an effective conversation and understand what the other person is saying. Also, another pointed out that Real_Avatar affords more opportunities to see the full body representation, gestures, head gaze, and body posture, which was most human-like and genuine.

Five participants preferred communicating with Scanned_Avatar, two who experienced it with the Conversation Cards task saying that it allows them to talk indirectly with the partner. One participant said, "I felt it was easiest to communicate with the Scanned_Avatar because I was able to talk to people indirectly. It made speaking candidly easier because it did not feel as real and as if there would be as many consequences for being honest." One positive aspect of Scanned_Avatar is that it felt natural to be in a virtual space with a virtual likeness of a character. The CG avatar may lower users' anxiety when they communicate with a stranger in a social environment. Also, when communicating with the Scanned_Avatar, users can still see the basic body language of the avatar's responses to what they are saying.

Six participants indicated that the No_Avatar was easiest for communication with in the social environment. In some cases, that may be because they almost felt like they were talking to an automated machine that would not judge them, helping them to feel free and secure to speak and act with their partner. One user said, "The No_Avatar made it easiest for me to communicate because it felt like there was less pressure of having someone standing in front of me."

5.4 Attention to Behavioral Cues

5.4.1 Results

We conducted separate analyses on each part of the questionnaire about attention to behavioral cues. A Shapiro-Wilk test showed a significant departure from normality on each item.

Attentional Focus A non-parametric within-subjects Friedman test found significant differences between avatar types in the attention to body posture $\chi^2(2) = 6.96$, $p = .03$, $W = 0.10$ and attention to facial expressions $\chi^2(2) = 13.25$, $p < .01$, $W = 0.18$, while the rest of the attentional allocations were not significantly different between avatar types, see table 1.

We used a post-hoc Wilcoxon signed-rank test with Bonferroni correction to compare attention to body posture and attention to facial expressions between all pairs of avatar conditions. Medians (IQR) for attention to body posture in the No_Avatar, Scanned_Avatar, and Real_Avatar conditions were 3 (1.25 to 6), 6 (5 to 6) and 5 (2.25 to 6), respectively. The test indicated that there was a significant difference between Scanned_Avatar and No_Avatar (Z = -3.31, $p < .01$, r = -0.39) and between Scanned_Avatar and Real_Avatar (Z = -2.63, p = .03, r = -0.31). The mean rank of Scanned_Avatar was higher than Real_Avatar and No_Avatar. However, there was no significant difference between Real_Avatar and No_Avatar. Medians (IQR) for attention to facial expression in the No_Avatar, Scanned_Avatar, and Real_Avatar conditions were 1 (1 to 1), 2 (1.25 to 4.75) and 3.50 (2 to 6), respectively. The test indicated that there was a significant difference between Real_Avatar and No_Avatar (Z = -2.92, $p < .01$, r = -0.34) and the mean rank of Real_Avatar was higher than No_Avatar. However, there were no differences between Real_Avatar and Scanned_Avatar or between Scanned_Avatar and No_Avatar. Figure 4 (a) shows the box-plot for all three types of avatars and each behavioral cue.

Missing Behavioral Cues A non-parametric within-subjects Friedman test found significant differences between avatar types in participants' sense of missing cues to body posture $\chi^2(2) = 14.21$, $p < .01$, $W = 0.20$ and facial expressions $\chi^2(2) = 12.33$, $p < .01$, $W = 0.17$; no significant differences were found for the other cues, see table 1.

Medians (IQR) for missing cues to body posture in the No_Avatar, Scanned_Avatar, and Real_Avatar conditions were 5 (2.25 to 6), 2 (2 to 3) and 2 (2 to 3.75), respectively. We used a post-hoc Wilcoxon signed-rank test with Bonferroni correction to do paired comparisons and found a significant difference between No_Avatar and Scanned_Avatar (Z = -3.46, $p < .01$, r = -0.41) and between No_Avatar and Real_Avatar (Z = -3.58, $p < .01$, r = -0.42). The mean rank of No_Avatar was higher than Scanned_Avatar and Real_Avatar. However, there was no significant difference between the Scanned_Avatar and Real_Avatar.

Medians (IQR) for missing facial expressions in the No_Avatar, Scanned_Avatar, and Real_Avatar conditions were 6 (5 to 7), 6 (3.50 to 7) and 4.50 (2 to 6), respectively. A post-hoc Wilcoxon signed-rank test with Bonferroni correction indicated a significant difference between No_Avatar and Real_Avatar (Z = -3.03, $p < .01$, r = -0.36) and between Scanned_Avatar and Real_Avatar (Z = -2.49, p = .04, r = -0.29). The mean rank of Real_Avatar was lower than No_Avatar and Scanned_Avatar. However, there was no significant difference between the Scanned_Avatar and No_Avatar. Figure 4 (b) shows the corresponding box-plot.

5.4.2 Discussion

Our finding of differences in the allocation of attention to body posture and facial expressions between the three types of avatars suggests that differences in cues to body posture and facial expressions between avatars may have an effect on the quality of communication. Therefore, including an optimal representation of body posture and facial expressions seems important. The lack of a difference in the amount of attention allocated to gesture could be because of the similarity of the gestural capabilities of the three avatar representations.

The greater attentional focus on the body posture of the Scanned_Avatar may be because the absence of other cues such as facial expressions forced participants to rely more heavily on body posture for non-verbal communication. Roth et al. [21] found similar results when they compared attentional allocation during interactions using embodied wooden mannequins versus in real life. We find it a bit surprising that the results show no significant difference in attentional allocation to body posture between the Real_Avatar and No_Avatar conditions. This may be because the Real_Avatar has many accessible behavioral cues that split the attentional focus, or because the representation is so natural that sustained attentional focus is not necessary. In the No_Avatar condition it makes sense that body posture is not attended to because there is nothing to see. The difference in attentional focus on facial cues between the Real_Avatar and No_Avatar is expected. Although the HMD covers the avatar's face, we still can see the lip movement in the Real_Avatar representation. We also anticipated a difference between Real_Avatar and Scanned_Avatar on this measure, but similar amounts of attentional focus may arise for different reasons. Artifacts, for instance, could capture attentional focus without being desirable.

For the Missing Behavioral Cues, the high rank of No_Avatar can be explained by the absence of a body or a face. The lack of a difference between the Scanned_Avatar and Real_Avatar conditions in how much people missed behavioral cues from the body posture makes sense due to their structural similarity. Likewise, more strongly missing cues from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3067783, IEEE Transactions on Visualization and Computer Graphics

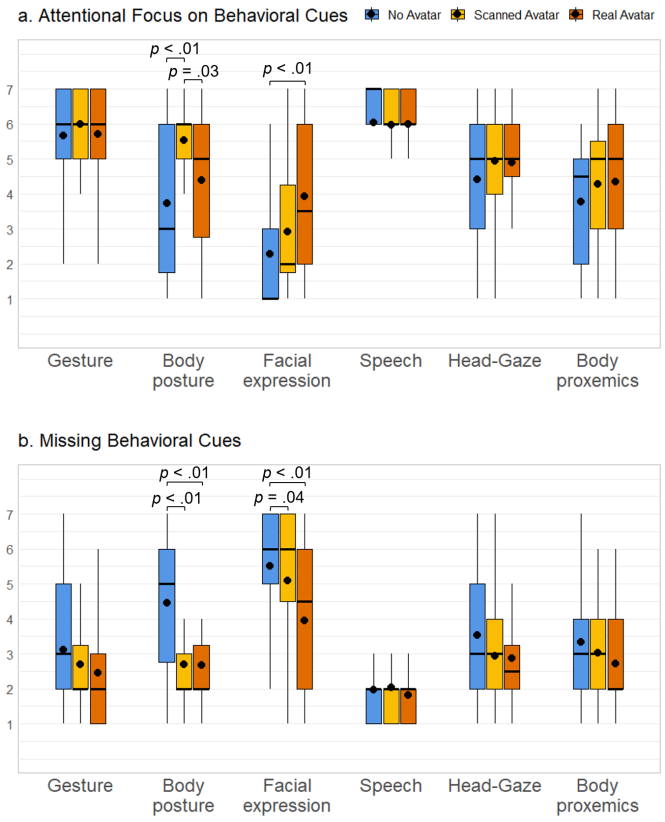

Fig. 4: Box-plots for both parts of the behavioral cues metric. (a) shows mean, median, and the significant p-values for each sub behavioral cue under the Attentional Focus. (b) shows mean, median, and the significant p-values for each sub behavioral cue under the Missing Behavioral Cues.

facial expressions in the No_Avatar and Scanned_Avatar compared to the Real_Avatar conditions makes sense in that the Real_Avatar was the only representation where facial expressions could be seen. We view this as a sanity check more than anything. Facial expressions are important to communication between people and increase the understanding of individuals' expressiveness.

We did not anticipate any differences in attention to speech, head-gaze, and body proxemics because all of these attributes were similarly present across our avatar conditions. A chart of the attentional allocation to each of the behavioral cues is provided in figure 4.

5.5 Mutual Gaze

5.5.1 Results

Anticipating that the rate of mutual gaze could be significantly affected by the different task requirements, we decided to use between-subject analyses to compare the effect of avatar within each task separately.

In this case, the sample size for each avatar type is 12 instead of 36 participants. For the Conversation Cards task, a Shapiro-Wilk test showed a significant departure from normality, $W(36) = 0.74$, $p < .01$. Therefore, a non-parametric between-subjects Kruskal-Wallis test was done which showed a significant difference between the avatar types in mutual gaze $H(2) = 6.62$, $p = .04$, $\eta_H^2 = 0.08$. Post-hoc Mann-Whitney tests using a Bonferroni correction were used to compare all pairs of groups. Medians (IQR) for the No_Avatar, Scanned_Avatar, and Real_Avatar were 97.46 (93.8 to 98.8), 93.6 (92.1 to 99.12) and 98.9 (97.6 to 99.9), respectively. The test indicated a significant difference ($U = 31.5$, $Z = -2.34$, $p = .05$, $r = -0.28$) between Real_Avatar and Scanned_Avatar and the mean ranks showed Real_Avatar was higher than Scanned_Avatar. However, there were no significant differences between No_Avatar compared with Real_Avatar and the No_Avatar compared with Scanned_Avatar.

A Shapiro-Wilk test showed a non-significant departure from normality in measures of mutual gaze during the Survival Items and Charades tasks. Therefore, we conducted a parametric between-subjects one-way ANOVA test for comparison in both tasks separately. For the Survival Items task, the test showed no significant difference in mutual gaze between the avatar types $F(2, 33) = 0.79$, $p = .46$. For the Charades task, the test did show a significant difference $F(2, 33) = 4.68$, $p = .02$, $\eta_p^2 = 0.22$. Post-hoc comparisons using Bonferroni correction indicated that there was a significant difference between Real_Avatar and No_Avatar ($p = .03$). The mean differences showed that time spent in mutual gaze in the Real_Avatar condition ($M = 80.68$, $SD = 10.32$) was higher than with the No_Avatar ($M = 65.65$, $SD = 13.15$). However, there was no difference between the Real_Avatar and Scanned_Avatar as well as the No_Avatar and Scanned_Avatar conditions. See figure 5 for more details.

5.5.2 Discussion

Our results show that the avatar type has an impact on mutual gaze in the Conversation Cards and Charades tasks but not in the Survival Items task. This makes sense in that participants were mainly looking at the cards during the Survival Items task, rather than at each other. The fact that mutual gaze was attempted for longer durations in the Real_Avatar condition even though the eyes were hidden is interesting, as it suggests that participants may have inferred a stronger potential affordance for mutual gaze in that condition than in the others. The majority of the participants mentioned in the exit survey that they preferred to converse with a real person to watch the partner's behavioral feedback whilst they shared personal information. Other participants said sometimes they avoided looking at the Scanned_Avatar because it looked weird and that interrupted them from thinking. In the Charades task, the longer periods of attempted mutual gaze may reflect participants' sense that facial expressions could potentially provide information helpful to guessing the action.

5.6 Unique Words

5.6.1 Results

As the results of the Eysenck Personality Inventory showed equivalent numbers of extroverts in each avatar/task condition, we did not control for extroversion as a potential co-factor. Because the Charades task did not involve speaking, we only conducted the audio analysis for the Conversation Cards and Survival Items tasks.

For each of those tasks, a Shapiro-Wilk test showed a non-significant departure from normality in the number of unique words spoken. Therefore, a parametric between-subjects one-way ANOVA was conducted to compare the effect of the avatar type on unique words counts for both tasks. For the Conversation Cards task, there was no significant effect of avatar type $F(2, 33) = 1.19$, $p = .32$. For the Survival Items task, there was a significant effect $F(2, 33) = 3.253$, $p = .05$, $\eta_p^2 = 0.17$. Post-hoc comparisons using Bonferroni correction indicated a significant difference between No_Avatar and Real_Avatar ($p = .05$). The mean differences showed higher unique words count with No_Avatar ($M = 148$, $SD = 38.598$) than Real_Avatar ($M = 112.50$, $SD = 24.303$). However, there are no significant differences between the other conditions. See figure 5 for more details.

5.6.2 Discussion

The fact that we did not find any difference in the number of unique words spoken between avatar conditions in the Conversation Cards task may be a reflection of the required nature of the conversation during that task. The result that relatively *more* words, rather than fewer, were spoken in the Survival Items task in the No_Avatar condition was a surprise to us. We can think of several possible explanations for these results. First, some of the participants pointed out in the exit survey that they talked and expressed themselves more in the No_Avatar condition than the other two conditions. This implies that the absence of the avatar could lead these participants to elaborate with more words. The other possibility is related to the main goal of the Survival Items task,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3067783, IEEE Transactions on Visualization and Computer Graphics

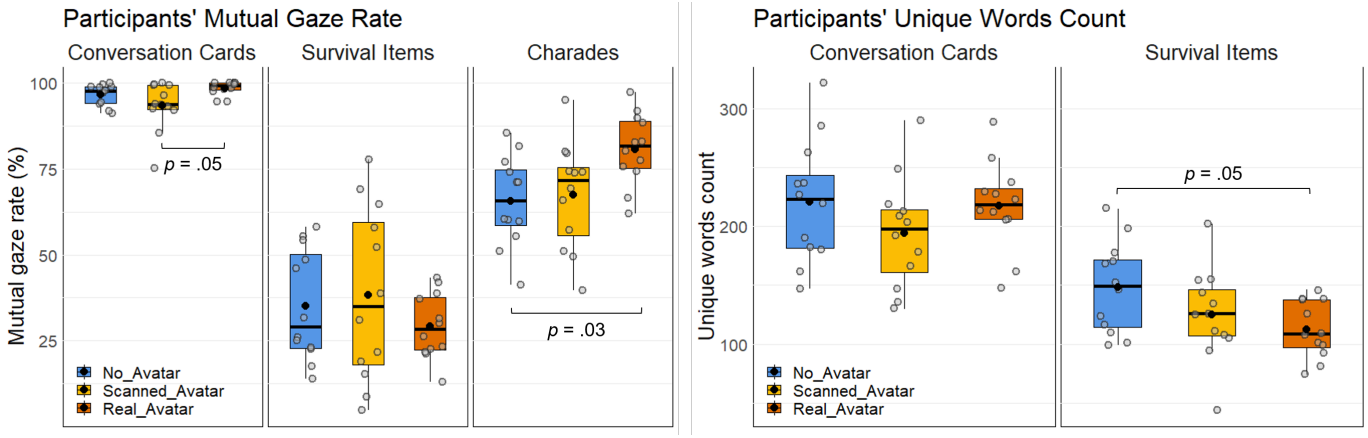

Fig. 5: In the left side we can see the means and medians of mutual gaze rates for Conversation Cards, Survival Items, and Charades. In the right side we can see the means and medians of unique words count for Conversation Cards and Survival Items.

which is convincing individuals to change one or two of the selected items. In this case, the presence of the Real_Avatar or Scanned_Avatar may help to more easily convince the participants to change their items, which makes them talk less and end the task with less unique words.

6 General Discussion

In the exit questionnaire, the majority of participants (66.67%) stated an overall preference for interacting in the Real_Avatar mode, while 25.00% preferred the Scanned_Avatar and 8.33% preferred No_Avatar. For those who preferred the Real_Avatar, their comments were that Real_Avatar was easy to communicate with, the same as face-to-face communication. Even though the HMD covers part of the face in the Real_Avatar condition, it was still more enjoyable and interactive for them compared to the other avatar types. This could be due to the greater availability of facial expressions and other subtle behavioral cues that could deliver more emotions and comprehensive thoughts. Further, the comments suggested that interacting with Real_Avatar created a deeper connection between the two users, the participant and the experimenter, and made them more socially confident to share information, especially in the Conversation Cards task. One participant who did not like the Real_Avatar said that it detracted from their immersion in the virtual environment because its style did not match that of the rest of the virtual world.

For participants who preferred the Scanned_Avatar, they liked it due to how advanced it was, and that it gave a good middle ground for immersiveness in the virtual environment compared to the other two avatars. Also some said that it felt the most natural, realistic, and easy to have an actual communication flow, due to the ability to see the avatar's body language and hand movements. At the same time, others who disliked the Scanned_Avatar felt it did not seem real enough, and they remarked that it was glitchy, fake, choppy, and uncomfortable to communicate with, because of the lack of facial expression and that the fingers could not move. Among the participants who preferred No_Avatar, some said it felt more natural compared to other avatar types, maybe because the other avatars tried to show reality in something not real. Others said it almost made them feel like they were talking to a computer that would not judge them on some points. Participants who disliked the No_Avatar said they felt almost like they were alone in comparison to the other avatars, and the conversation felt less real. Some participants said it reminded them of talking to a robot or a phone call. The majority of the participants said that No_Avatar was hard to communicate with in some tasks. Overall, we find both advantages and disadvantages in using each avatar type.

Going back to our main research question, our results extend prior work by showing that the Real_Avatar, with its fully detailed and unmediated facial and body expression features afforded by the video-see-through representation, better supports several important aspects of interpersonal communication than the Scanned_Avatar and No_Avatar representations, in contexts relevant to immersive architectural design reviews where all users are wearing HMDs. Although the No_Avatar representation is less encumbering than the Scanned_Avatar approach, and simpler to achieve in terms of hardware and software requirements than either the Real_Avatar or Scanned_Avatar representations, our results show that it is associated with less satisfactory outcomes on multiple measures of interpersonal communication effectiveness that could be important to the success of collaborative negotiations in the context of shared immersive virtual environments.

7 Limitations and Future Work

Our experiment has several limitations that need to be explicitly acknowledged. First, the avatar appearance is not the only thing that differs between our three different avatar conditions. Additional differences include the smaller FOV in the Real_Avatar condition, and potential differences in the end-to-end system latency between all three conditions, which we did not quantify. Potential effects of these factors on the overall participant satisfaction with each avatar type cannot be ruled out. Second, the ZED mini provides pass-through images from the points of view of cameras whose exit pupils are not precisely aligned with the exit pupils of the eyes; in particular, there is a non-trivial offset in the forward direction which causes some perceptual distortion. Using a true RGBD capture system could enable a more accurate re-rendering of the real world imagery from the point of view of the participant's own eyes. Finally, as our participants were not architects, and our tasks did not involve design review discussions, extrapolation of our results to a professional immersive architectural design review context need to be done with caution.

Nevertheless, our findings provide some take-away points for future enhancement in avatar representation and in the design of experiments for improving users' experience in virtual social environments. In the Scanned_Avatar case, participants remarked that adding lip synchronization and finger movement, and enabling smoother real-time body movement would be desirable. Supporting eye contact is also important. Ultimately, it would be ideal to be able to capture and accurately portray a full range of subtle facial expressions on the Scanned_Avatar model, including mouth movements that are accurate enough to support lip reading. One user said, "If you can get the mouths to move even somewhat realistically on the Scanned_Avatar, I think it will present the best combination of immersion and human representation." The Real_Avatar condition could be improved by the use of a camera system that has a wider FOV and provides a more geometrically correct view. An important practical necessity is to achieve robust real time background subtraction using depth data instead of relying on a green screen. Ultimately, it would be desirable to be able to portray the occluded parts of each user's face on their avatar, instead of showing the HMD.

Future experimental work could aim to further elucidate the different task-dependent requirements for the avatar representation type. It

seems clear that the highest fidelity representations are not necessarily required or even optimal for all situations. For instance, one participant said, "Seeing a CG avatar decreased the pressure of answering anything right or wrong. We think using that for an interview preparation program would be really helpful." Another said, "With the No_Avatar, having no face probably helped me in freely expressing myself without reservations. This could probably be used for therapy, I guess." Future experiments involving designers as participants could additionally help to better inform design review applications specifically.

8 Conclusions

There is a growing need for virtual reality technologies that provide better support for interactive communication between co-immersed users. Current systems still have multiple limitations in terms of their ability to accurately support subtle social cues that can be important in a professional environment. In this paper, we compared multiple measures of communication effectiveness across three different avatar representation approaches. We found that Real_Avatar (video-see-through) evoked more interpersonal trust than Scanned_Avatar (wearing an HMD) or No_Avatar (HMD and controllers only), more sustained attempts at mutual gaze and less attention to body posture than Scanned_Avatar, and higher co-presence and a greater attentional focus on facial expressions than No_Avatar. Co-presence was also higher with Scanned_Avatar than No_Avatar. While most of our participants expressed an overall preference for the Real_Avatar, this sentiment was not universal and multiple advantages were also cited for the other two avatar representations. Thus, individual differences as well as differences in task requirements add complexity to the determination of an optimal avatar type to support interpersonal communication in VR.

Acknowledgments

This work was supported in part by the National Science Foundation (CHS-1526693). The authors wish to thank Professor Emeritus Lee B. Anderson, from the Department of Architecture at the University of Minnesota, for his guidance in the conceptual stages of this research, and Professor Benjamin Munson, chair of the Department of Speech-Language-Hearing Sciences at the University of Minnesota, for his valuable guidance on audio data analysis. We also thank Olivia Yang for data-input and audio-transcription work, and Ville Cantory for assisting us with the photos of our experiment and setup. The first author was supported by the Saudi Arabian Cultural Mission Scholarship (SACM).

References

[1] A. Abbas, M. Choi, J. Seo, S. H. Cha, and H. Li. Effectiveness of immersive virtual reality-based communication for construction projects. *KSCE Journal of Civil Engineering*, 23:4972–4983, 2019.

[2] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg. Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Human communication research*, 34(2):287–318, 2008.

[3] F. Biocca. The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of computer-mediated communication*, 3(2):JCMC324, 1997.

[4] F. Biocca and C. Harms. Networked minds social presence inventory: |(scales only, version 1.2) measures of co-presence, social presence, subjective symmetry, and intersubjective symmetry. 2003.

[5] F. Biocca, C. Harms, and J. Gregg. 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*, pp. 1–9, 2001.

[6] A. M. Bodling and T. Martin. *Eysenck Personality Inventory*, pp. 1007–1008. Springer New York, New York, NY, 2011. doi: 10.1007/978-0-387-79948-3_2025

[7] R. Cañigueral and A. F. d. C. Hamilton. The role of eye gaze during natural social interactions in typical and autistic people. *Frontiers in Psychology*, 10:560, 2019.

[8] S. Cho, S.-w. Kim, J. Lee, J. Ahn, and J. Han. Effects of volumetric capture avatars on social presence in immersive virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 26–34. IEEE, 2020.

[9] V. Clay, P. König, and S. König. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12(1), Apr. 2019. doi: 10.16910/jemr.12.1.3

[10] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 529–536, 2003.

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

[12] M. L. Hecht. The conceptualization and measurement of interpersonal communication satisfaction. *Human Communication Research*, 4(3):253–264, 1978.

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

[14] D. Jo, K.-H. Kim, and G. J. Kim. Effects of avatar and background representation forms to co-presence in mixed reality (mr) tele-conference systems. In *SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, SA '16. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2992138.2992146

[15] D. Jo, K.-H. Kim, and G. J. Kim. Effects of avatar and background types on users' co-presence and trust for mixed reality-based teleconference systems. In *Proceedings the 30th Conference on Computer Animation and Social Agents*, pp. 27–36, 2017.

[16] N. S. Lockwood and A. P. Massey. Communicator trust across media: A comparison of audio conferencing, video conferencing, and a 3d virtual environment. In *2012 45th Hawaii International Conference on System Sciences*, pp. 839–848. IEEE, 2012.

[17] C. S. Oh, J. N. Bailenson, and G. F. Welch. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:114, 2018.

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

[19] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy. Rich media, poor judgement? a study of media effects on users' trust in expertise. In *People and Computers XIX—The Bigger Picture*, pp. 267–284. Springer, 2006.

[20] D. Roth, G. Bente, P. Kullmann, D. Mal, C. F. Purps, K. Vogeley, and M. E. Latoschik. Technologies for social augmentations in user-embodied virtual reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–12, 2019.

[21] D. Roth, M. E. Latoschik, C. Bloch, and G. Bente. When some things are missing: The quality of interpersonal communication in social virtual reality (presentation). In *Presentation on the 68th Annual Conference of the International Communication Association (ICA), May 24-28 2018, Prague, Czech Republic*, 2018.

[22] D. Roth, J.-L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar realism and social interaction quality in virtual reality. In *Proceedings of the 23rd IEEE Virtual Reality (IEEE VR) conference*, pp. 277–278, 2016.

[23] H. Salzmann and B. Froehlich. The two-user seating buck: Enabling face-to-face discussions of novel car interface concepts. In *2008 IEEE Virtual Reality Conference*, pp. 75–82. IEEE, 2008.

[24] T. Sen, M. R. Ali, M. E. Hoque, R. Epstein, and P. Duberstein. Modeling doctor-patient communication with affective text analysis. In *Seventh international conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 170–177. IEEE, 2017.

[25] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[26] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.

[27] S.-E. Wei, J. Saragih, T. Simon, A. W. Harley, S. Lombardi, M. Perdoch, A. Hypes, D. Wang, H. Badino, and Y. Sheikh. Vr facial animation via multiview image translation. *ACM Transactions on Graphics (TOG)*, 38(4):1–16, 2019.

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