Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality

Scott W. Greenwald(✉), Zhangyuan Wang, Markus Funk, and Pattie Maes

MIT Media Lab, 75 Amherst Street, Cambridge, MA, USA
scottgwald@media.mit.edu

Abstract. Room-scale virtual reality (VR) holds great potential as a medium for communication and collaboration in remote and same-time, same-place settings. Related work has established that movement realism can create a strong sense of social presence, even in the absence of photorealism. Here, we explore the noteworthy attributes of communicative interaction using embodied minimal avatars in room-scale VR in the same-time, same-place setting. Our system is the first in the research community to enable this kind of interaction, as far as we are aware. We carried out an experiment in which pairs of users performed two activities in contrasting variants: VR vs. face-to-face (F2F), and 2D vs. 3D. Objective and subjective measures were used to compare these, including motion analysis, electrodermal activity, questionnaires, retrospective think-aloud protocol, and interviews. On the whole, participants communicated effectively in VR to complete their tasks, and reported a strong sense of social presence. The system's high fidelity capture and display of movement seems to have been a key factor in supporting this. Our results confirm some expected shortcomings of VR compared to F2F, but also some non-obvious advantages. The limited anthropomorphic properties of the avatars presented some difficulties, but the impact of these varied widely between the activities. In the 2D vs. 3D comparison, the basic affordance of freehand drawing in 3D was new to most participants, resulting in novel observations and open questions. We also present methodological observations across all conditions concerning the measures that did and did not reveal differences between conditions, including unanticipated properties of the think-aloud protocol applied to VR.

Keywords: Room-scale virtual reality · Copresence · Non-verbal communication · Collaboration

Electronic supplementary material The online version of this chapter (doi:10.1007/978-3-319-60633-0_7) contains supplementary material, which is available to authorized users.

D. Beck et al. (Eds.): iLRN 2017, CCIS 725, pp. 75–90, 2017.
DOI: 10.1007/978-3-319-60633-0_7

1 Introduction

Embodied room-scale virtual reality endows users with a very different relationship to their own avatars and virtual environments than analogous non-immersive systems, which use input from keyboards, mice, gamepads, and joysticks. Users *embody* their avatars in a direct way – movements are one-to-one at physical scale, and they move and reach naturally in order to interact with objects. One dual implication of this fact is that when observing others' avatars in the virtual environment, they *look* human. That is, the same precise measurement of movement that is required to deliver the first-person VR experience allows these movements to be made visible to others with great fidelity as body movements. Consequently, when two people share a virtual space in this fashion, they each have a strong sense of being present with another human. Prior works have established the general principle that high *movement realism* achieves a strong sense of social presence, using comparatively low information-bandwidth.

Our system allows two users to interact in room-scale VR (i.e. six degree-of-freedom tracking of head and two handheld controllers) in the same-time, same-place setting, and is the first of its kind that we are aware of in the research community. The goals of this paper are to (1) establish the basic feasibility and utility of this kind of multi-user interaction, (2) pilot methodologies for studying behavior in this setting, (3) offer early results related to similarities, differences, advantages and disadvantages compared with face-to-face, (4) explore the use of freehand drawing in 3D for communicative interaction, and (5) propose future research directions. We made the choice to use minimal avatars to avoid complicating our results with effects related to the choice of body representation (Fig. 2).

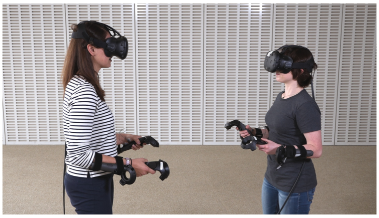


Fig. 1. Same-time, same-place interaction in room-scale VR

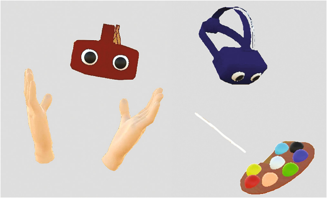


Fig. 2. Avatars for Charades and Pictionary

We designed a set of goal-oriented, communicative activities for pairs of participants to perform in an experimental setting. These were popular word-guessing games based on gesturing and freehand drawing that could be directly compared in face-to-face and VR settings. The different words that participants attempted to communicate represented a broad array of concepts and corresponding symbolic gestures. We view these as proxies for various communicative face-to-face activities. To explore the use of 3D drawing in communicative interaction, we had participants play an analogous game using freehand drawing in 3D instead of 2D. We evaluated the experiences using a combination of methods and metrics: the VR system itself provided data on movement; electrodermal activity was captured to measure engagement; users completed questionnaires measuring perceived mental load, presence, and other aspects of the experience; participants did think-aloud reflection while reviewing recordings; and semi-structured interviews were conducted with participants to gain further qualitative insights.

In the sections that follow, we first discuss related work, then we briefly describe the system that we built for collaboration in room-scale virtual reality in the same-time, same-place setting. Next we discuss the experiment we carried out, which required extensive modification and adaptation of the basic system, and present the corresponding results. Then we discuss the implications of the quantitative and qualitative results of the experiment. Finally, we conclude and highlight promising directions for future research.

2 Related Work

Two related bodies of research focus on (i) the psychological experience of interacting with human avatars or agents in immersive virtual environments, and (ii) methods and affordances for computer-mediated communication and collaboration. Studies of the psychological experience of interacting with *embodied agents* or *avatars* in immersive virtual environments have focused on *agency*, *presence*, *copresence* (or *social presence*), and *social influence* [2,3,6]. They employ self-reports, behavioral metrics, cognitive metrics, and qualitative methods to gain insight. Two factors shown to influence all of the above are *behavioral realism* and *photorealism* of the agent or avatar representations [1]. Importantly, a recent meta-analysis [6] showed that avatars have greater social influence than agents. That is to say, people react more strongly to other people than to non-human agents that purport to be people. In the present work, we are only concerned with the case of real-time interaction between people, so the upshot is that our use case resides at the end of the spectrum where social influence tends to be larger. A relevant study by Garau et al. [7] considers this case, also through the lens of behavioral and photo-realism. In their system, users' headsets and a single handheld controller are spatially tracked with six degrees of freedom. The authors define a metric for the *perceived quality of communication*, and test how this depends on *type of avatar* and *type of gaze*. The former refers to three different levels of realism, and the latter refers to two different methods for generating avatar eye gaze behavior. Results show a positive effect when gaze behavior mimics natural behavior. However, the said "natural behavior" is inferred from a model of speaker turn-taking, and not directly controlled by the user's real eye gaze. In contrast, our system does not use any indirect inference: it displays only the head orientation, and does not purport to represent eye movement. It also displays hand positions, supporting the use of unintentional and symbolic gestures.

The related work in the field of computer-mediated communication investigates the merits of different communication affordances from the perspective of collaboration. Isaacs and Tang [10] perform a systematic comparison of audio, video, and face-to-face as mediums for communication. They note increases in communication efficiency in video over audio-only communication due to the ability to indicate agreement using a nodding gesture, without interrupting the speaker. They note the great value of being able to point in the shared environment, as in face-to-face communication, but also highlight that video can be more efficient than face-to-face in cases where it removes distractions. Our system supports nodding to express agreement, and we also make observations about the removal of distractions in our somewhat different setup. In [12] from the same year, the authors focus on gaze and the representation of video avatars. They contend that the ability to judge which other participant is being gazed upon by each participant is important for group dynamics. Our system also allows each participant to see where other participants are looking through their head orientation, which we confirm to be an important feature. More recently, [11] uses see-through display augmented reality for remote collaboration. This work considers puzzle-solving as a collaborative task, and also underscores the importance of the affordance for pointing when collaborating in a shared space. Our system supports the ability to point in space, in a way that is directly analogous to the physical world except for the small physical disparity between the user's physical and virtual hands. The most similar prior work from the field of computer-mediated communication is *GreenSpace II* [4], a multi-user, six degree-of-freedom (or *6DoF*) system for architectural design review. Its two users would see stylized head and hand avatars (with one hand per user), and point in the shared space. Their physical movements were constrained to a small space – to make larger movements, they needed to use a 6DoF mouse. The paper demonstrates the feasibility of sharing an immersive virtual environment with spatially tracked head and hand avatars. A significant portion of the feedback provided in the qualitative evaluation focused on the limitations of the technology. The present work does confirm what is supposed there – namely that once the fidelity of the experience is improved (wider field of view, natural physical movement, better audio experience), the utility improves greatly, and the interaction feels natural.

3 System for Copresence in Room-Scale VR

We present a system to act as a foundation for exploring same-time, same-place collaboration in room-scale virtual reality. A later version of the system, described in Greenwald, et al. [8], is available for the community to use.[1] It allows each user to see head and hand avatars representing the other user, with their apparent virtual positions matching their respective physical positions, as illustrated in Fig. 1. The form of the head avatar corresponds closely to the

[1] CocoVerse, https://github.com/cocoverse.

physical headset. The hand avatars are customized according to the activity being performed.

The choice not to display a head or a body was made in order to be deliberately minimal – representing the hardware itself, so as to avoid making arbitrary choices that could significantly influence the experience. An entire field of related work (see e.g. [13]) concerns itself with how the representation of the body impacts the user's psychological experience, and we are just concerned with the baseline communication capabilities in the scope of this paper. Even so, we did opt for a few minor tweaks based on the results of preliminary testing. The headset is modified with the addition of simple, static eyes on the front, since users found that this dramatically increased the sense of social presence. In pilot testing, users had difficulty creating expressive hand gestures using a literal representation of a hand holding a controller. Instead, the default hand avatars are flat hands positioned vertically above the top of the controller, which proved to be more versatile.

Our system uses the HTC Vive, an off-the-shelf 6DoF VR system consisting of a headset, a pair of handheld controllers, and pair of tracking base stations. The Vive system requires one computer per headset, but several systems can share a set of base stations. Sharing is possible because the devices being tracked (headset and controllers) are receivers which observe optical signals from passive base stations. We calibrate a single coordinate system between the VR systems by sharing a set of configuration files between their host computers. Players' apparent virtual locations are made to match their physical locations, and the systems continually synchronize a virtual world representation over a local network. Our "naive" implementation sends updated headset and handheld controller positions from every user to every other user at 90 Hz, and has been tested with a maximum of five users in a single space. With that number of users two challenges arise: (i) with our "naive" implementation, network and graphics performance start to suffer, and (ii) physical cable management, with a cable running to each user's headset. Our environment was implemented in Unity, and we used a custom serialization protocol and TCP connection in the provided networking framework to synchronize the state of the environment between the host computers.

In order to be able to comprehensively study user interactions that take place in our system, we considered it an essential design requirement to be able to record and playback these interactions. Rather than screen recording, which is limited to one or two perspectives, we opted for recording of 3D paths of motion and orientation. This format supports visual inspection and quantitative analysis alike, allowing recordings to be viewed from any angle, and analyzed numerically. Viewing replays of VR interactions while actually in the VR space is a novel and insightful experience, and this topic will be discussed further in our experimental results.

4 Experiment Comparing Face-to-Face with VR

We sought reference activities to help us accomplish the stated goals of investigating advantages and disadvantages of VR vs face-to-face for communicative interaction, and exploring the use of freehand drawing in 3D in this setting. We identified the word-guessing games *Charades* and *Pictionary* that fit these constraints. These require the use of gestural communication that is both symbolic and expressive, and they are also composed of a sequence of short, goal-oriented subtasks exercising different means of non-verbal and gestural communication. Pictionary also has the property of being naturally extensible from its familiar 2D form into a 3D form – allowing for 2D and 3D interactions to be compared side-by-side as well, providing a baseline for investigating freehand drawing in 3D. In the Charades game, the focus of communication is on the body itself, while Pictionary makes use of a spatial medium to contain and convey drawings. This contrast should yield greater insight into the effectiveness of these two different communicative affordances, body movement and drawing, and allow us to conjecture what kinds of activities would be most amenable to this form of collaboration. It should also help identify the most limiting technological shortcomings, and hence provide recommendations about what improvements would be most worthy of effort. Overall, we see the communicative gestures and actions required by these two different word guessing games as a proxy for the many kinds of communication required for a variety of collaborative tasks. The tasks themselves are communicative, but only "collaborative" to a limited extent, since only one participant acts at a time. Isolating one-way communication in this fashion will act as a first step, paving the way for future research into more complex collaborative tasks using this configuration.

4.1 Method

To compare the effect of these independent variables (face-to-face vs. VR conditions, and the two game-based task settings), we conducted a user study. We designed our experiment following a repeated measures design with one independent variable: the word guessing game that is played (Charades or Pictionary) combined with whether the game was played in Virtual Reality (VR) or Face-to-Face (F2F). As dependent variables we measured the *Electrodermal Activity (EDA)* through sensors, *Task Load Index (TLX)*, *level of presence* as well as some other related aspects of the system usability through questionnaires. We counter-balanced the order of the conditions according to the Balanced Latin Square.

The two primary hypotheses related to the contrast between our independent variables were that (1) face-to-face and VR would be similarly effective, despite the ostensible differences in the richness of the communication channels, and (2) the games would reveal different quantitative and qualitative attributes of non-verbal communication across conditions, given their different uses of body movement vs. drawing.

4.2 Apparatus and Tasks

Room setup. Figure 3 depicts the physical space layout used for the experiment. Players act or draw in the *activity space*. Facilitators operate and control the game session from the *control space*. The *real-world whiteboard* is used for drawing during the F2F Pictionary game. Game play information such as the current word, timer, game mode etc. is shown on the *real-world display* during F2F conditions. The *camera* footage of the F2F games provides video for think-aloud review sessions.

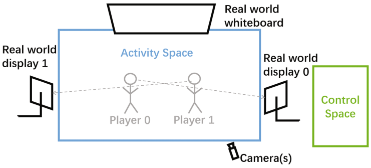


Fig. 3. Physical room layout for face-to-face and VR games.

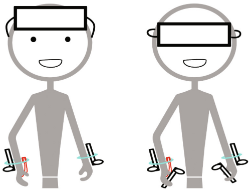


Fig. 4. Positioning of headset, controllers, and sensors during F2F and VR activities.

Positioning of devices on body. Figure 4 shows the positioning of the GSR sensors, VR controllers and headset on the body during F2F and VR activities, mounted with elastic velcro bands. The positional tracking devices worn during the activities collected movement data that could be directly compared between F2F and VR conditions. The GSR sensor was mounted to participants' dominant hand, with gel electrodes placed on the lower palm.

Quantitative data acquisition. Electrodermal activity data was collected using a *Shimmer GSR* sensor with *iMotions* software. After smoothing and detrending, *Coefficient of Variation (CV)* was calculated as a metric of arousal, as in [5].

Movement data was collected from the position of the headset and two arm-mounted controllers. The HTC Vive system provides positional data at a rate of 90 Hz. The sensors occasionally become momentarily occluded, causing tracking to be lost. We computed the average distance traveled per tracked frame (cm/frame) for each session and player.

Word selection for guessing games. The guessing words used during the study were selected from lists of varying difficulty provided by a game website.[2] For each game, we informally piloted candidate words, and observed the type of body gestures used while playing (fingers, hands, full-body, etc.), as well as the use of 3D space where applicable. Based on the results, we selected a final set

[2] The Game Gal, https://www.thegamegal.com/.

of words of varying difficulty that would sample a variety of gesture types and highlight different uses of 3D space.

Questionnaire design. We used the NASA Task Load Index (TLX) questionnaire and a custom set of questions. The TLX questions were presented using a slider with options from 0 to 100 in increments of 5, with the slider initially positioned at 50. Informed by our pilot tests, additional questions were presented to inquire about specific aspects of game play, the differences between F2F and VR, and the usability of the user interface.

4.3 Procedure

Subjects arrived in pairs, and experimental sessions began with a general introduction, before putting on VR devices and sensors. Pairs played through all five conditions (Charades and Pictionary 2D, each in F2F and VR, plus Pictionary 3D in VR) in the order dictated by their experimental group, with questionnaires administered as appropriate after each condition. At the start of each condition, participants were first given an opportunity to briefly familiarize themselves with the devices and physical or virtual space, and a simple warm-up task was provided. During game play, for each word the "acting" player was given 45 seconds to silently convey a word to the "guessing" player, with roles alternating as directed by the system. The facilitator determined when the word had been guessed correctly, and operated a control interface on one host computer to advance to the next word. After playing both F2F and VR variants of a game, participants would perform a retrospective think-aloud protocol and interview together. They reviewed the video and immersive VR playback (or just immersive VR playback, in the case of Pictionary 3D) in succession, in the order that they were played.

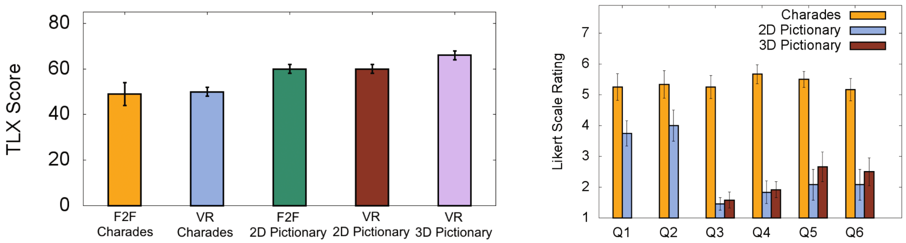
4.4 Participants

We invited 6 pairs of participants (4 female and 8 male) to take part in the study, with ages ranging from 19 to 50 (M = 31.0 years, SD = 10.62 y). The study took approximately 2.5 h, of which roughly 30 min were spent playing the games, 30 min reviewing recordings, 30 min filling out questionnaires, 30 min interviewing, and the remaining time used for breaks and setup. Participants were compensated with a $25 gift card.

4.5 Quantitative Results

Here we present the data that was collected during the user study. To analyze the NASA-Task Load Index (TLX), we used a one-way repeated measures ANOVA. For the questionnaire we applied a non-parametric Friedman test. Bonferroni correction was used for all post-hoc tests.

NASA-TLX. When comparing the TLX between the five conditions, the F2F Charades led to the least perceived cognitive load (M = 49.33, SD = 18.6), followed by the VR Charades ($M = 50.17$, $SD = 10.26$), the 2D VR Pictionary ($M = 60.34$, $SD = 9.55$), the 2D F2F Pictionary ($M = 60.58$, $SD = 9.26$), and the 3D Pictionary ($M = 66.17$, $SD = 9.60$). Mauchly's test of sphericity indicated that we can assume a sphericity of the data ($p > 0.05$). The one-way repeated measures ANOVA revealed a significant difference between the conditions, $F(1, 4) = 6.589$, $p < .001$. As a post-hoc test, pairwise comparisons revealed a significant difference between the VR Charades condition and the 3D VR Pictionary condition ($p < 0.05$). The effect size shows a large effect ($\eta^2 = .375$). Figure 5a shows the results graphically.



(a) (b)

Fig. 5. (a) The NASA-Task Load Index results of the user study for all conditions and (b) The quantitative results of the Likert scale questionnaire for the different games. Questions Q1-Q6 are explained in the text. All error bars depict the Standard Error.

When analyzing the Likert questions of the questionnaire, we used a non-parametric Friedman test. All Likert items were 7-point Likert items meaning: 1 = strongly disagree and 7 = strongly agree. For Q3-Q6 we used Wilcoxon signed-rank post-hoc tests with an applied Bonferroni correction for all conditions resulting in a significance level of $p < 0.017$. All results of the questionnaire are depicted in Fig. 5b.

Q1: *"Overall the experience playing the game in VR was different than playing F2F."* Participants found that the overall experience playing the Charades game in VR was more different from playing it F2F ($M = 5.25$, $SD = 1.49$) than in the 2D Pictionary game ($M = 3.75$, $SD = 1.42$). The Friedman test revealed a significant difference between the two games, $\chi^2(1) = 6.0$, $p = 0.014$.

Q2: *"Playing the game in VR was harder than playing F2F."* Further, the participants rated playing the Charades game to be harder in VR compared to F2F ($M = 5.33$, $SD = 1.56$), compared to the 2D Pictionary game ($M = 4.00$, $SD = 1.76$). The Friedman test did not reveal a significant difference between the two games ($p > 0.05$).

Q3: *"The absence of a body avatar was a problem in VR."* Considering the absence of a body avatar, the participants the participants rated the Charades game the most problematic ($M = 5.27$, $SD = 1.35$), followed by the 3D VR

Pictionary (M = 1.55, SD = .93), and the 2D VR Pictionary (M = 1.45, SD = .69). The Friedman test revealed a significant difference between the games, $\chi^2(2) = 18.0$, $p < 0.001$. The post-hoc tests showed a significant difference between 2D Pictionary and Charades ($Z = -2.825$, $p = 0.005$) and 3D Pictionary and Charades ($Z = -3.072$, $p = 0.002$).

Q4: *"The absence of facial gesture representations was a problem in VR."* When analyzing if the absence of facial gesture representations were a problem in VR, the participants rated the Charades game as the most problematic for that aspect ($M = 5.67$, $SD = 1.07$), followed by the 3D VR Pictionary ($M = 1.92$, $SD = .9$), and the 2D VR Pictionary ($M = 1.83$, $SD = 1.27$). The Friedman test revealed a significant difference between the games, $\chi^2(2) = 20.14$, $p < 0.001$. The post-hoc tests showed a significant difference between 2D Pictionary and Charades ($Z = -3.075$, $p = 0.002$) and 3D Pictionary and Charades ($Z = -3.089$, $p = 0.002$).

Q5: *"The absence of hand gesture representations was a problem in VR."* Considering if the absence of hand gesture representations is problematic in the VR games, the participants rated the Charades game as the most problematic (M = 5.5, SD = .905), followed by the 3D VR Pictionary (M = 2.67, SD = 1.67), and the 2D VR Pictionary ($M = 2.08$, $SD = 1.73$). The Friedman test revealed a significant difference between the games, $\chi^2(2) = 17.077$, $p < 0.001$. The post-hoc tests showed a significant difference between 2D Pictionary and Charades ($Z = -2.842$, $p = 0.004$) and 3D Pictionary and Charades ($Z = -2.952$, $p = 0.003$).

Q6: *"The absence of finger gesture representations was a problem in VR."* Finally, when analyzing whether the absence of finger gesture representation was problematic for playing the VR game, the participants rated the Charades game as the most problematic ($M = 5.17$, $SD = 1.267$), followed by the 3D VR Pictionary (M = 2.50, SD = 1.567), and the 2D VR Pictionary (M = 2.08, SD = 1.73). The Friedman test revealed a significant difference between the games, $\chi^2(2) = 11.73$, $p = 0.003$. The post-hoc tests showed a significant difference between 2D Pictionary and Charades ($Z = -2.739$, $p = 0.006$) and 3D Pictionary and Charades ($Z = -2.823$, $p = 0.005$).

Considering the players' analysis of their experience in both games we were asking additional questions comparing their VR and F2F experience.

Q7: *"Reviewing videos/the VR recordings helped me remember my experience during the games."* When analyzing where the participants found it better to review their experience, the participants found the VR recording of the games better ($M = 6.00$, $SD = 1.27$) than the video recording ($M = 5.58$, $SD = .51$). A non-parametric Friedman test could not find a significant difference between the video recording and the VR recording.

Q8: *"Reviewing videos in VR/ on video helped me gain new insights into my interactions".* Considering gaining new insights on the participants interactions during the game, the participants rated the VR recording to provide more insights (M = 6.00, SD = 1.20) compared to the traditional video recording

($M = 4.91$, $SD = 1.37$). A Friedman test revealed a significant difference between the two recording systems, $\chi^2(1) = 4.500$, $p = 0.034$.

Electrodermal Activity. Considering the analysis of the EDA using the CV, the results revealed that the 3D VR Pictionary led to the most EDA activity ($M = .24$, $SD = .18$), followed by the 2D VR Pictionary ($M = .20$, $SD = .20$), the F2F Charades ($M = .15$, $SD = .11$), the VR Charades ($M = .14$, $SD = .07$), and the 2D F2F Pictionary ($M = .11$, $SD = .04$). Mauchly's test of sphericity indicated that we cannot assume a sphericity of the data ($p < 0.001$). Therefore, we apply a Greenhouse-Geisser correction to adjust the degrees of freedom. Unfortunately, a one-way repeated measures ANOVA could not reveal a significant difference between the conditions ($p > .05$).

Head Movement. When analyzing the head movements the participants made during the different conditions, the 3D Pictionary ($M = .117$, $SD = .038$), Charades F2F ($M = .115$, $SD = .045$), and the Charades VR ($M = .111$, $SD = .041$) lead to similarly frequent head movements, followed by the F2F Pictionary 2D ($M = .105$, $SD = .045$). The Pictionary 2D in VR led to the least head movements ($M = .078$, $SD = .018$). A one-way repeated measures ANOVA revealed a significant difference between the conditions, $F(4, 32) = 2.670$, $p = .049$. However, a post-hoc did not reveal a significant difference.

Left Hand Movement. For the movements of the participants' left hands, we found that the F2F Charades led to the most hand movement ($M = .27$, $SD = .15$), followed by the 2D F2F Pictionary ($M = .21$, $SD = .10$), the VR Charades ($M = .19$, $SD = .05$), the 3D Pictionary ($M = .13$, $SD = .04$), and the 2D VR Pictionary ($M = .10$, $SD = .03$). Mauchly's test of sphericity indicated that we cannot assume a sphericity of the data ($p < 0.001$). Therefore, we apply a Greenhouse-Geisser correction to adjust the degrees of freedom. A one-way repeated measures ANOVA showed a significant difference between the conditions, $F(1.477, 14.771) = 8.775$, $p = .005$. The post hoc tests showed a significant difference between 2D VR Pictionary and all other conditions. Further there was a significant difference between VR Charades and 3D Pictionary (all $p < .05$).

Right Hand Movement. We found that the 3D Pictionary led to the most right hand movement ($M = .26$, $SD = .12$), followed by F2F Charades ($M = .24$, $SD = .11$), the 2D VR Pictionary ($M = .23$, $SD = .18$), the 2D F2F Pictionary ($M = .23$, $SD = .11$), and the VR Charades ($M = .21$, $SD = .06$). A one-way repeated measures ANOVA could not reveal a significant difference between the conditions ($p > .05$).

4.6 Qualitative Results

The questionnaire questions reported above captured many of the most salient trends we discovered during our prior informal pilots. The qualitative results presented in this subsection are focused on ideas that are either more complex and nuanced, or first became apparent in the main study. In this section we report factual aspects of this feedback, and save a discussion of its significance and relationship to our quantitative results for the *Discussion* section that follows.

One idea that was important but also very subtle to interpret was the degree of expressivity participants perceived in the gestures of others. This subject was always brought up in the interview at the end of the entire session. All participants agreed that, as expected, the smoothness and precision of the representation of movement in the space led to a high degree of expressivity and sense of being able to perceive some aspects of emotion or other non-verbal reactions. It was difficult for participants to describe this explicitly, because in the same-time, same-place setting, it seemed very natural that the other person's emotions could be interpreted through movement, and therefore not noteworthy on its own. For this reason it was primarily during the process of viewing VR recordings that participants were able to consider in isolation what kind of information avatar movements contained. Several participants found their own movements and those of their partners to be distinctive and recognizable. Other participants disagreed, and felt that they would not be able to distinguish a playback of their own avatar actions from actions of unknown others. This on-the-fence status was well summarized by one participant's comment that there were "glimpses of humanity" that would appear sporadically throughout the process of viewing. Another participant reported "they're very emotive" and "you can definitely tell it's you."

Recounting briefly some comments about the general relationship between the face-to-face and VR experiences, participants mentioned most frequently that VR Charades was challenging because of the lack of face and body avatars. After initial reports that the VR 2D Pictionary experience was qualitatively highly similar to its face-to-face counterpart, the facilitators questioned participants for more detail. Because participants rarely look to each others' faces for feedback during gameplay, the entire focus was really on the board, and they found the experience of drawing on the physical whiteboard versus the virtual whiteboard nearly identical. They cited several advantages for VR over face-to-face: the virtual board erases automatically between words, switching colors was faster using the VR color palette than physically switching markers, and in VR the body does not occlude the drawing surface, so it was never an issue that the actor's body was blocking the view. One corollary that came out in interviews was that VR offered the advantage of removing some aspects of face-to-face interaction that are distracting, awkward, or unpleasant. Attention to gender, ethnicity, body image, and certain visual social cues are impeded through the invisibility of the physical body.

Next, we review comments participants made about the process of reviewing video versus VR recordings. Several participants reported reviewing video to be unpleasant, mentioning they felt "silly" watching themselves play. In contrast, they described the experience of watching replays in VR as insightful and fun. In 3D Pictionary specifically, many participants reported that viewing the replay from a different perspective allowed them to see how their drawings were not as decipherable from their partners' perspective as from their own.

One last area of participant feedback that we'll highlight in this section is the description of 3D versus 2D drawing. Nearly all participants described drawing in 3D as challenging, but some enjoyed the challenge while others found it frustrating. There was broad agreement that drawing in 3D was typically slower, but there were cases where it offered advantages. The biggest challenge was becoming accustomed to considering multiple viewing perspectives. There was a weak consensus that drawing on a virtual 2D plane would be a winning strategy if emphasis was placed on finishing quickly. In contrast, participants in our experiment participants were given time limits, but were not otherwise incentivized to finish quickly. This observation is highly coupled to the specific task of Pictionary play, and may have been accentuated by the fact that the word list was designed for 2D Pictionary (Fig. 6).

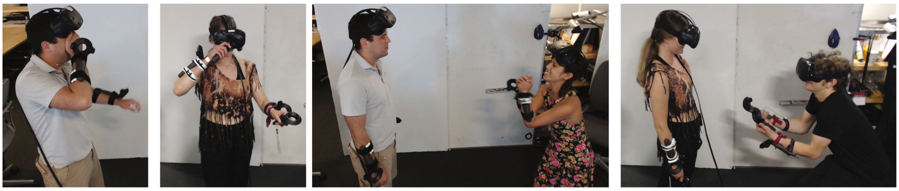


Fig. 6. Expressive poses in F2F/VR acting out "blind" (left) and "beg" (right)

5 Discussion

The previous section presents a disparate set of results from our five data sources. In this section we highlight some salient relationships between these results.

We begin by observing that participants felt strongly that (1) the communication medium was not sufficient for Charades, while feeling that (2) the medium was entirely sufficient for Pictionary in 2D and 3D, as evidenced by the questionnaire responses. In the former, the absence of facial gestures, finer hand gestures, finger movements, and a body for non-verbal communication were considered highly problematic, while in Pictionary they were considered irrelevant. Further underscoring this was the response to Q1. At the Likert scale value of 3.75 participants were very close to "neutral" on the question. We interpret this as a strong statement about two aspects of the interaction: (1) the adequacy of the hand-held controllers at approximating the face-to-face experience of drawing on a whiteboard, and (2) the expressiveness of the avatars. We know that when the focus of the interaction is on the body itself, as in Charades, the simple avatars were inadequate. Despite participants' reports to this effect, even the most difficult words we tested were guessed correctly by a subset of groups – meaning that the communicative affordances were nonetheless powerful enough to admit creative workarounds. Furthermore, the qualitative feedback indicated that the avatars were perceived as quite expressive and emotive. Reconciling these statements, we propose the following guideline, pertaining to systems equivalent to ours: a collaborative task that is communicative, but with a central focus that is not on the face or body itself, when facilitated by well-adapted task-specific interface affordances, will yield an overall experience comparable to face-to-face. Stated more broadly, minimal avatars provide a powerful and versatile baseline set of communication affordances. Roughly speaking, the two games we tested define a spectrum between the worst and best-adapted activities for our simple head and hand avatars. We conclude that, when designing system for a certain form of collaboration in VR, one should ask whether it is more Charades-like or more Pictionary-like in order to decide whether the additional effort of embodying a more sophisticated avatar is justified.

Next, comparing movement, TLX, and EDA data for 2D Pictionary reveals an interesting correlation. In particular, it was a high-EDA activity, and a somewhat high perceived cognitive load (TLX) activity, while being the lowest-movement activity overall. This indicates a mode of mental engagement corresponding to decreased physical movement. If there were any coupling between physical movement and EDA, it would work against this result, hence it is interesting to highlight.

Now we review true advantages of VR over face-to-face that were shown in our results, beginning with those relating to efficiency of task performance. First, the virtual whiteboard did not need to be manually erased, and therefore decreased the time and energy required to perform an equivalent task in VR vs. F2F. Next, the transparency of the body in VR minimized occlusion of the virtual whiteboard – the drawing player could stand right in front of the board without preventing the guessing player from seeing the drawing. Next, a psychological benefit was reported in participants' observation that masking the physical body can be beneficial to focus and decrease social anxiety in collaborative interactions. All of these can be viewed as advantages of "programming" the virtual visual environment, by instantly changing its properties in ways that require time and effort, or aren't possible at all, in the physical world. Indeed, they "satisfy needs of communication," physically and psychologically, in a way that is not possible face-to-face, and hence go *beyond being there* [9].

Now we turn briefly to the methodological implications of this experiment. Although our EDA data did not uncover significant differences between our activities, it was close enough that we would conjecture that further refinement of the method to reveal significant differences would be possible – for instance subdividing overall games into smaller components, or applying peak detection algorithms. Next, discussing movement data, the only significant result was that the left (palette) hand stays very still during 2D Pictionary. While this is not exciting on its own, the prospect of doing more sophisticated analysis of body movement with absolute positional data rather than (or in addition to) accelerometry is very exciting. This is firm evidence that activity analysis and recognition can be applied to the positional data collected by the Lighthouse system, and certainly any other system with similar or greater precision that comes along. Finally, our significant result about the difference between video and VR review of games is worthy of note. Participants found VR review equally good (i.e. not significantly different) for *recall* of the experiment, but significantly better at providing new insights. Not only does this provide a basis for researchers to obtain highly nuanced qualitative feedback from participants, it also suggests that review of VR activities could be used in the context of learning or training – leveraging the reflective power of scrutinizing ones' own performance in a way that is demonstrably better than video.

6 Conclusion

Same-time, same-place interaction in virtual reality has been shown without any doubt as a practical medium for communication and collaboration, which carries with it a sense of social presence that is adequate for a variety of non-verbal methods of communication mediated by hand gestures, head gestures, and overall spatial movement. If facial gestures, torso, or leg movements are particularly relevant to the communicative task, the minimal system we built would need to be extended to support these in some fashion before being applied for the use case. It was shown that drawing in 3D is challenging but highly promising due to the new space for expression that it opens up. It was observed that interacting in VR has the advantage of masking aspects of physical appearance and the body that can be distracting during collaborative interaction. Reviewing interaction in VR allowed participants to gain new insight into how their own communicative processes did and didn't work, and this could be useful as a tool for reflection or coaching. We see all three of these as fruitful directions for future research in collocated and remote computer-mediated communication using room-scale virtual reality.

Acknowledgements. We would like to thank Hisham Bedri and Ronen Zilberman for their contributions to the project, and Chris Schmandt for his useful feedback.

References

1. Bailenson, J.N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., Blascovich, J.: The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. Presence Teleoperators Virtual Environ. **14**(4), 379–393 (2005). http://www.mitpressjournals.org/doi/10.1162/105474605774785235
2. Bailenson, J.N., Yee, N.: Digital chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments. Psychol. Sci. **16**(10), 814–819 (2005). http://dx.doi.org/10.1111/j.1467-9280.2005.01619.x
3. Blascovich, J.: Social influence within immersive virtual environments. In: Schroeder, R. (ed.) The Social Life of Avatars. Computer Supported Cooperative Work, pp. 127–145. Springer, London (2002). doi:10.1007/978-1-4471-0277-9_8

4. Davidson, J.N., Campbell, D.A.: Collaborative design in virtual space - greenspace ii: a shared environment for architectural design review. In: McIntosh, P., Ozel, F. (eds.) Design Computation: Collaboration, Reasoning, Pedagogy: Acadia Conference Proceedings, pp. 165–179. ACADIA, Tucson, October 1996. http://papers.cumincad.org/cgi-bin/works/Show?c7d4
5. Doberenz, S., Roth, W.T., Wollburg, E., Maslowski, N.I., Kim, S.: Methodological considerations in ambulatory skin conductance monitoring. Int. J. Psychophysiol. **80**(2), 87–95 (2011)
6. Fox, J., Ahn, S.J.G., Janssen, J.H., Yeykelis, L., Segovia, K.Y., Bailenson, J.N.: Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. Hum. Comput. Interact. **30**(5), 401–432 (2015). http://www.tandfonline.com/doi/full/10.1080/07370024.2014.921494
7. Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., Sasse, M.A.: The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In: Proceedings of the Conference on Human Factors in Computing Systems - CHI 2003, p. 529. ACM Press, New York, April 2003. http://portal.acm.org/citation.cfm?doid=642611.642703
8. Greenwald, S.W., Corning, W. Maes, P.: Multi-user framework for collaboration and co-creation in virtual reality. In: Proceedings of the 12th International Conference on Computer Supported Collaborative Learning (2017). http://hdl.handle.net/1721.1/108440
9. Hollan, J., Stornetta, S.: Beyond being there. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1992, pp. 119–125. ACM, New York (1992). http://doi.acm.org/10.1145/142750.142769
10. Isaacs, E.A., Tang, J.C.: What video can and cannot do for collaboration: a case study. Multimedia Syst. **2**(2), 63–73 (1994). http://link.springer.com/10.1007/BF01274181
11. Kim, S., Lee, G., Sakata, N., Billinghurst, M.: Improving co-presence with augmented visual communication cues for sharing experience through video conference. In: 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 83–92. IEEE, September 2014. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6948412
12. Okada, K.I., Maeda, F., Ichikawaa, Y., Matsushita, Y.: Multiparty videoconferencing at virtual social distance. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work - CSCW 1994, pp. 385–393. ACM Press, New York, October 1994. http://portal.acm.org/citation.cfm?doid=192844.193054
13. Slater, M., Sanchez-Vives, M.V.: Transcending the self in immersive virtual reality. Computer **47**(7), 24–30 (2014)