Revisiting Distance Perception with Scaled Embodied Cues in Social Virtual Reality

Zubin Choudhary* University of Central Florida
Matthew Gottsacker† University of Central Florida
Kangsoo Kim‡ University of Central Florida
Ryan Schubert§ University of Central Florida
Jeanine Stefanucci¶ University of Utah
Gerd Bruder‖ University of Central Florida
Gregory F. Welch** University of Central Florida

Abstract

Previous research on distance estimation in virtual reality (VR) has well established that even for geometrically accurate virtual objects and environments users tend to systematically mis-estimate distances. This has implications for Social VR, where it introduces variables in personal space and proxemics behavior that change social behaviors compared to the real world. One yet unexplored factor is related to the trend that avatars' embodied cues in Social VR are often scaled, e.g., by making one's head bigger or one's voice louder, to make social cues more pronounced over longer distances.

In this paper we investigate how the perception of avatar distance is changed based on two means for scaling embodied social cues: *visual head scale* and *verbal volume scale*. We conducted a human-subject study employing a mixed factorial design with two Social VR avatar representations (full-body, head-only) as a between factor as well as three visual head scales and three verbal volume scales (up-scaled, accurate, down-scaled) as within factors. For three distances from social to far-public space, we found that visual head scale had a significant effect on distance judgments and should be tuned for Social VR, while conflicting verbal volume scales did not, indicating that voices can be scaled in Social VR without immediate repercussions on spatial estimates. We discuss the interactions between the factors and implications for Social VR.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

1 Introduction

Virtual reality (VR) systems can provide users with a sense of feeling present in a virtual environment similar to perceiving that environment in the real world [42]. Modern Social VR setups can further provide users with embodied representations through one's avatar that other users can see and interact with in a shared virtual space. Such social spaces are highly interesting for a wide range of application contexts from entertainment and social gaming to virtual meetings and distributed work environments [5, 9, 34, 43].

Social VR spaces further provide affordances to developers and users that do not exist in the real world. Probably the most prominent example is that one's virtual avatar body does not necessarily have to match one's physical body in terms of its pose, geometry, appearance, or behavior. While there is some evidence that visually faithful virtual reconstructions of one's physical body can foster more natural social behaviors [50], other strains of research found that sometimes it is advantageous to deviate from realism even if such faithful avatars are available. For instance, the Social VR avatars in Facebook's Project Horizon [1] are deliberately presented with slightly enlarged heads, which magnifies their screen space on low-resolution consumer head-mounted displays (HMDs), allowing users to see each other's facial expressions even over longer distances. Recently, Choudhary et al. [10] formally investigated the effects of such head scales in Social VR on the perception of facial expressions and Uncanny Valley effects [30], revealing that people are surprisingly tolerant towards scaled heads and that there are practical advantages of and uses for such approaches.

Similar to the scaling of visual aspects of users in Social VR, this approach can also be applied to non-visual aspects such as users' voices in virtual space. Leaving aside technical reasons, such as the need for adjustments due to the placement or quality of one's microphone or headphones, it is common practice to amplify audio volumes in Social VR until one can hear each other clearly—a feat that is not easily replicated in the real world, e.g., when talking to people with a low natural voice. In the real world, one would need to move closer to the origin of the sound waves to perceive them with a higher volume.

However, while there are good reasons for scaling visual and audio cues in Social VR, less is known about how these change the social behavior among users. A key advantage of Social VR over video conferencing or voice chats is that the interlocutors can feel present and interact with each other in a shared virtual space as if they were together in a similar space in the real world. Unfortunately, previous research on spatial perception in VR has provided strong evidence that distances in virtual environments are perceived differently compared to the real world [28, 37]. This means that not only are distances to objects and other users in Social VR systematically underestimated for longer and overestimated for shorter distances [8], but also that natural proxemics and avoidance behaviors are changed [3]. What effects scaled visual or audio cues in Social VR have on top of these yet unsolved hardware and graphics related perceptual effects is still largely underexplored and ill-understood.

In this paper, we present a human-subject study in which we formalized and evaluated the effects of two means for scaling embodied social cues: *visual head scale* and *verbal volume scale*. For different distances from social to far-public space, we compared two Social VR avatar representations (full body, head-only) as well as three visual head scales and three verbal volume scales (up-scaled, accurate, down-scaled). The results of our study provide insights for our main research questions:

- **RQ1**: Is interpersonal distance perception in Social VR affected by scaled embodied visual cues (*visual head scale*)?
- **RQ2**: Is interpersonal distance perception in Social VR affected by scaled embodied audio cues (*verbal volume scale*)?

*e-mail: zubinchoudhary@knights.ucf.edu
†e-mail: gottsacker@knights.ucf.edu
‡e-mail: kangsoo.kim@ucf.edu
§e-mail: rschuber@ist.ucf.edu
¶e-mail: jeanine.stefanucci@psych.utah.edu
‖e-mail: bruder@ucf.edu
**e-mail: welch@ucf.edu

- **RQ3**: Do the effects of scaled embodied cues on distance perception depend on users' Social VR avatar representation (from *full-body* to *head-only*)?

This paper is structured as follows. Section 2 discusses related work. Section 3 describes the experiment. Section 4 presents our results and Section 5 discusses our findings. Section 6 concludes the paper.

2 Related Work

In this section, we present related work on distance estimation in VR and methods for assessing distance perception in the scope of Social VR.

2.1 Distance Perception in Virtual Reality

VR display setups can provide users with a highly realistic impression of a computer-generated three-dimensional virtual environment. In particular in Social VR, a compelling immersive experience and an accurate perception of sizes, distances, and spatial interrelations are highly important to afford natural spatial interactions between users and their environment. For Social VR, distance perception defines users' proxemics behavior among users and their environment. In Hall's structure of proxemics, the space around a person can be divided into intimate (up to 0.46 m), personal (up to 1.22 m), social (up to 3.7 m), and public space ranges [21]. In both the real world and in Social VR, it is common that people feel uncomfortable when another person enters their intimate or personal space, and they choose to maintain socially and culturally defined distances from one another [23]. Distance perception also governs users' locomotion behavior. For example, Bailenson et al. found that users allowed more space between them and an avatar when approaching the avatar from the back compared to the front [4]. Similarly, Sanz et al. demonstrated that when given an obstacle avoidance task with virtual objects and humans, VR users respected the personal space of virtual humans and maintained a significantly longer distance from them than from inanimate objects [3].

Modern real-time rendering systems offer most of the spatial visual cues we can find in the real world, including perspective, interposition, lighting, and shadows [45]. However, distance and size perception were repeatedly found to be non-veridical in VR, meaning that users tend to overestimate or underestimate spatial relations—often by magnitudes of 50% or more [28, 46]. Many hardware and graphics related factors influencing a user's spatial perception in VR have been isolated over the last decades, but holistic theories and practical remedies remain elusive [12, 25, 37, 51].

In the real world, humans use a variety of visual and audio depth cues to perceive distances: pictorial cues such as occlusion, relative size, relative density, height in the visual field, aerial perspective, and non-pictorial cues such as motion parallax, convergence and accommodation, and binocular disparity [13], as well as audio volume and sound reverberations due to humanly discernible characteristics of the mechanics of sound wave propagation, in particular when people are talking in social settings. Several of these cues can be reliably transduced into immersive VR, but there are limitations for some of them due to limitations in the capabilities and fidelity of current-state VR setups. For example, researchers have found that field of view restrictions, added weight, rendering quality of pictorial depth cues, and the vergence-accommodation conflict affect distance perception with HMDs [37]. Moreover, VR hardware and software are not yet capable of rendering fully accurate audio sources and interactions [39], which may lead to distance misperception. In fact, Serafin et al. and Nielsen et al. have successfully used this misperception to redirect the movement and attention of participants [31, 40]

In this paper, we focus on modern peculiarities of Social VR environments and investigate how two methods to scale visual and audio signals impact distance perception. In particular, we look at visual head scaling, which denotes a computer graphics approach to usually up-scale the heads of virtual characters (sometimes called "big heads") to make them more visible on low-resolution displays or over longer distances [18]. Similarly, we also look at verbal volume scaling, which is commonly used in Social VR to amplify voices to be heard over a longer distance than they would usually be perceptible [24].

2.2 Assessing Distance Perception

A person's egocentric distance perception cannot be directly observed, so researchers have developed multiple methods for assessing perceived distances [37]. Straightforward *verbal estimates* are commonly used, but this method incorporates the participants' potentially biasing cognitive influences rather than relying solely on perceptual faculties [28] and results in high variance [35]. For example, if participants are asked to estimate the distance between their position and an avatar, they may use one of its body parts as a reference before making their estimate. However, if the body part changes size or is subjected to some other manipulation in VR, the reference becomes unreliable and their cognitive estimation process is affected. Loomis and Philbeck showed that participants were able to accurately verbally assess near distances, but they tend to underestimate farther ones [29] .

Another assessment methodology, *perceptual matching*, asks participants to consider a reference object when estimating the distance to a target object [41]. Depending on the implementation of the method, participants manipulate the reference object or use it as a purely visual aid to determine the distance to the target. *Perceptual bisection*, a variation of perceptual matching, involves adjusting the position of the reference object until it is halfway between the participant and the target object [6]. Loomis and Philbeck argue that the perceptual matching method reduces biases caused by human cognition [11, 29, 36].

The most common category of distance estimation methods leverages *visually directed actions*. In this methodology, the participant usually first observes their egocentric distance to a virtual object, after which they close their eyes or are blindfolded. They are subsequently tasked to perform a physical action to indicate the distance at which they perceive the object from them. The actions may include reaching, throwing, walking, or pointing. In *blind walking*, participants are tasked with walking straight toward the remembered location of the target object [28]. In *blind triangulated walking*, the participant turns an oblique angle to the target object and walks a few steps before turning back to the direction in which they remember the target object to be located. These actions yield a side and an angle of a triangle, which can be used to compute the perceived distance to the object. A variant of this method is *blind triangulated pointing*, where users usually take a few side-steps before pointing with their arm towards the remembered location of the target object [8]. Visually directed actions are useful because researchers can directly infer the participants' perception of distance from their proprioceptive and kinesthetic systems [27]. Previous results suggest that these approaches are highly accurate up to at least 10 meters [8, 27, 28].

Due to spatial constraints at our lab, and for these reasons, we decided to perform both verbal estimates and blind triangulated pointing for embodied distance responses, but no blind walking.

3 Experiment

In this section, we describe the experiment that we conducted to analyze the perception of Social VR avatar distances with scaled social cues by changing the visual head scale and verbal volume scale along with two avatar representations (full-body and head-only). In the different trials, participants saw a male avatar standing in front of them, while they could hear the avatar speaking with a lab member's normal speaking voice.

(a) Full body representation (b) Head only representation

Figure 1: Social VR avatars with different visual head scales used in the experiment: (a) full-body and (b) head-only avatar with, from left to right, up-scaled visual head (2×), accurate head scale (1×), and down-scaled visual head (0.5×)

3.1 Participants

After initial pilot tests, we estimated the effect size of the expected strong effects, and based on a power analysis, we decided on recruiting the number of participants from our university community. In the end, 38 participants took part in our experiment (28 male and 10 female; ages between 18 and 32, $M = 21.82$ $SD = 3.56$). All of the participants had normal or corrected-to-normal vision, while eight wore glasses and three wore contact lenses during the experiment. None of the participants reported any known visual or vestibular disorders that were not corrected, such as color or night blindness, dyschromatopsia, or a displacement of balance. The participants were either students or non-student members of our university, who responded to open calls for participation, and received monetary compensation for their participation.

3.2 Material

For our study, we used an Oculus Rift S HMD, which we connected to a graphics workstation (Intel Core i7-7820HK CPU @ 2.90 GHz, 32GB RAM, NVIDIA GTX 1070 graphics card, Windows 10 Pro). The workstation was used by the experimenter to run the rendering and logging software and monitor the participant's view and activities in the VR environment. The HMD provided a 110 degree field of view, and had a resolution of 1280×1440 pixels per eye at a refresh rate of 80 Hz. During the experiment, participants were standing in a marked location on the floor in our lab, wearing the HMD, while they were instructed to hold an Oculus Touch controller in their dominant hand. This controller was used to advance experiment trials and for the blind triangulated pointing task described below.

To investigate participants' distance perception with Social VR avatars, we prepared a virtual hallway space (see Figure 1) in the Unity game engine (version 2019) in which two avatars appeared at different distances from participants under different social scaling conditions. For the Social VR avatar in our study, we chose a rigged and animated life-size 3D male virtual human model (1.8 meters tall) from Mixamo (see Figure 1) with a neutral idle standing animation. The virtual hallway had the dimensions 5 m (width) × 3 m (height) × 15 m (length), which consisted of subtlety textured curtains and a floor below, rendered with the "Lit" shader in Unity. We decided on this limited-cue environment to provide a comparable stimulus as used in related studies [7, 8, 47]. For our verbal audio stimulus, we had a member of our lab record his voice in high fidelity while reading at his natural speaking volume from an article, which we looped.

3.3 Methods

3.3.1 Study Design

We used a 2×3×3×3 mixed factorial design. Our between-subject variable was *body type* and our within-subject variables were *visual head scale, verbal volume scale* and *distance*. The variables are described below.

- **Body Type (2 levels)**: We tested two different body representations for the avatar's body. We decided on the extremes among current Social VR setups.
  - **Full-Body**: A full-body avatar representation is shown to the participants. This representation is often used by Social VR setups with full-body tracking.
  - **Head-Only**: Only the avatar's head is shown to the participants. This representation is often seen in Social VR setups with head-only tracking.
- **Visual Head Scale (3 levels)**: We decided to test three different head scales, which were scaled to match three relative distances with respect to the avatar's current distance.
  - **Accurate**: This is the normal head scale of the avatar, calibrated to match human proportions.
  - **Up-scaled**: The head is up-scaled to a size that matches the visual angles subtended by a normal-scaled head that is 2 meters closer to the participant.
  - **Down-scaled**: The head is down-scaled to a size that matches the visual angles subtended by a normal-scaled head that is 2 meters farther away from the participant.

  We tested three target distances $d \in \{3, 5, 7\}$ in meters. For each distance, the agent's head was scaled with one of three ratios $r \in \mathbb{R}$, depending on the three experimental conditions (*accurate*, *up-scaled*, *down-scaled*).

$$r = \begin{cases} 1, & \text{if } accurate \\ \frac{d}{d-2}, & \text{if } up-scaled \\ \frac{d}{d+2}, & \text{if } down-scaled \end{cases}$$

  The agent's head scale was computed by multiplying the ratio r with the agent's default head scale. Effectively, this scaled the head size to that of an agent that was standing at that distance (accurate), 2 meters closer (up-scaled) or 2 meters farther away (down-scaled).
- **Verbal Volume Scale (3 levels)**: We decided on three different volumes for our verbal audio stimulus which matched three relative distances with respect to the avatar's distance. Therefore, we used Unity's default logarithmic volume roll-off algorithm without reverberation.
  - **Accurate**: The verbal volume is calibrated to match the distance of the avatar.

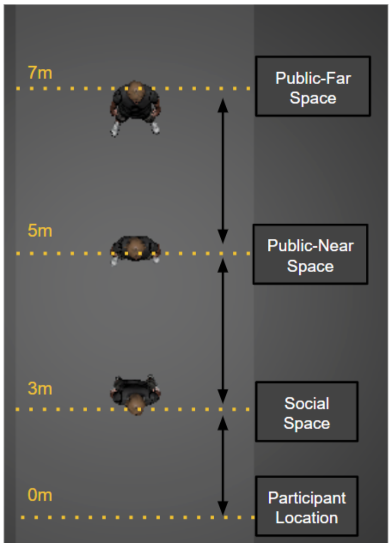


Figure 2: Three avatar distances (3, 5, and 7 meters) tested in the experiment according to Hall's Proxemics Theory [22].

- **Up-scaled**: The verbal volume is increased to match that of a person talking 2 meters closer to the participant.
- **Down-scaled**: The verbal volume is decreased to match that of a person talking 2 meters farther away from the participant.

- **Distance (3 levels)**: For the experiment, the avatar appeared at three different distances. We followed Hall's Proxemics Theory [22] with the choice of our avatar distances (see Figure 2).
  - **3 meters**: The avatar is located in *social space*.
  - **5 meters**: The avatar is located in *near-public space*.
  - **7 meters**: The avatar is located in *far-public space*.

Each participant only experienced one body type, which we decided on for improved external validity, considering that current Social VR setups tend to foster uniformity among the body type of avatar representations. A total of 20 participants experienced the head-only conditions and 18 participants experienced the full-body conditions.

3.3.2 Distance Measurement Protocols

We used two measurement protocols to assess participants' perception of the distance to the Social VR avatars in the different conditions of this experiment. The protocols were chosen to capture both verbal and proprioceptive responses, which are among the most common approaches (see Section 2) and give insights into cognitive and perception-action responses to the stimuli.

Blind Triangulated Pointing: Participants had to judge the distance to the seen avatar using the method of blind triangulated pointing, which we adapted to the configuration of our setup. Similar to previously introduced procedures [8, 16, 27], participants held a controller in their dominant hand as they observed the target. A colored red ray was protruding from the controller in the virtual environment. At the beginning of each trial, participants were instructed to hold out their arm and point the ray straight at the avatar that is shown in front of them. When participants felt ready to judge the distance to the target, they had to close their eyes, trigger the button of the controller to fade the rendered scene to black and removing the ray on the HMD, take two steps to the right, and point the controller at the target (see Figure 3). Participants were instructed to point at the target as accurately as possible while performing the side stepping at a reasonable speed to reduce effects of decreased precision caused by changes in the remembered position of the target over time [32]. Participants received no feedback about their pointing accuracy in order to minimize the effects of perception–action motor recalibration in the response method while assessing distance perception. Participants were trained on how to perform this task prior to the experiment trials.

Verbal Distance Estimation: Participants judged the distance to the avatar in the different conditions by looking at the target in front of them and speaking out loud how far away they believe it is from them. We did not impose time constraints on the participants but asked them to judge the distance once they felt ready. The verbal estimates were recorded by the experimenter in either feet or meters based on each participant's preference. For the analysis, we converted all responses to meters.

3.3.3 Procedure

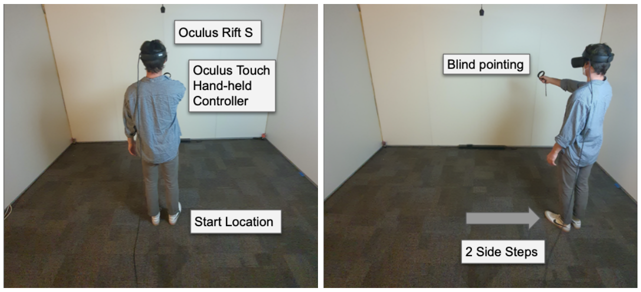
As participants arrived, the experimenter greeted and handed them an informed consent form. Once they agreed to participate in the study by giving their consent, the experimenter asked them to fill out the Simulator Sickness Questionnaire (SSQ) [26] to measure their pre-experiment sickness symptoms. The experimenter then introduced the study, where the participants will estimate the distance to a Social VR avatar standing at different distances, while hearing an audio loop of the avatar speaking. The experimenter also explained that they will experience some visual and audio changes while performing the distance estimation tasks throughout the experiment. However, the methods and conditions were not disclosed to the participants. Participants then donned the Oculus Rift S HMD, on which they could see the avatar standing in a virtual hallway facing towards them. Each participant either saw the full-body or head-only representation of the avatar (between-factor). During each trial, participants judged distances first via the verbal estimation method, followed by the blind triangulated pointing method as described above.

We did not restrict the participants' head movements. They were allowed to freely move their heads vertically and laterally which allowed them to scan the environment and look at the floor. This matches the unconstrained settings of Social VR spaces. In particular, the following depth cues were available to participants: pictorial cues such as, relative size, height in the visual field, and non-pictorial cues such as motion parallax, convergence and binocular disparity.

Before the experiment trials started, we included a practice session so that the participants had a chance to get acquainted with the two distance estimation methods. During the practice session, the avatar was standing at four meters in front of the participant without any visual/audio scaling. Once the participants were ready, they proceeded to the experimental trials. We varied the experiment conditions randomly according to the factors described in Section 3.3.1.

In total, each participant experienced 27 main trials in randomized order. However, to reduce any potential learning effects among the three distances (3, 5 and 7 meters), we included 18 additional trials between each three main trials, where we showed two randomly chosen intermediate distances (4–6 meters) that did not occur in the main trials. We excluded these from the analysis. After completing 30 trials, participants were given a 5–10 minute break.

Once they completed all trials, they were asked to complete the post-experiment SSQ questionnaire for simulator sickness, as well as the Slater-Usoh-Steed presence questionnaire [48] and the co-presence questionnaire by Garau and Slater [17]. The presence questionnaires were included to assess participants' overall sense of feeling present in the virtual environment as well as their feelings of co-presence with the avatars. Before ending the study, participants further completed a debriefing questionnaire, which asked them about their general preferences and reasoning behind their estimated distances, followed by a demographics questionnaire. The study



(a) (b)

Figure 3: Annotated photos showing a participant performing the blind triangulated pointing task in the laboratory setup wearing the Oculus Rift S HMD and pointing at the virtual target with the Oculus Touch controller.

ended with a monetary compensation. The study protocol was approved by our university's institutional review board.

4 Results

In this section, we present the results of the experiment. We highlight the significant results and go into detail on the findings for the two social cue scaling methods. All results were normally distributed according to QQ plots and a Shapiro–Wilk test at the 5% level. We analyzed the results with a mixed ANOVA and Tukey multiple comparisons at the 5% significance level with Bonferroni correction. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

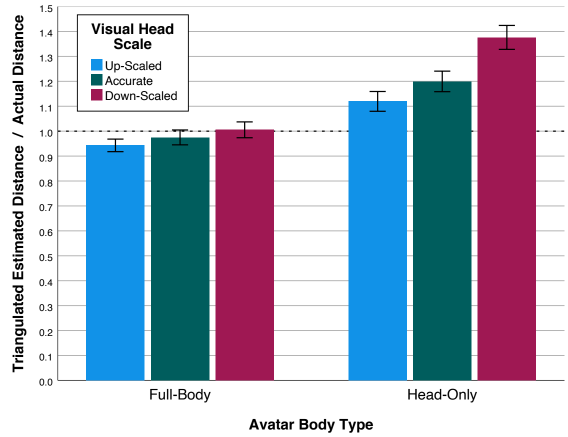
Avatar Body Type: Figure 4 shows the pooled results for the two avatar representations. We found a significant main effect of *body type* on distance judgments both for the triangulation distance measure, $F(1,36) = 6.22$, $\boldsymbol{p = 0.017}$, $\eta_p^2 = 0.147$, as well as for the verbal distance measure, $F(1,36) = 5.90$, $\boldsymbol{p = 0.02}$, $\eta_p^2 = 0.141$. This result is not unexpected, based on the previous literature on distance perception in VR [37], considering that head-only avatar representations provide fewer depth cues compared to full-body avatar representations. On average, participants *underestimated* distances for the *full-body* avatar representation by -5.5% ($SD = 1.7\%$) for triangulation responses and -30.1% ($SD = 1.5\%$) for verbal responses. In comparison, on average, participants *overestimated* distances for the *head-only* avatar representation by $+23.2\%$ ($SD = 2.5\%$) for triangulation responses and $+5.0\%$ ($SD = 3.2\%$) for verbal responses.

Avatar Distance: We found a significant main effect of the avatar's *distance* on distance judgments both for the triangulation distance measure, $F(1.52, 54.61) = 148.94$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.805$, as well as for the verbal distance measure, $F(1.03, 37.08) = 84.87$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.702$. This result was expected: participants estimated longer distances for avatars that were farther away, meaning that the system worked properly.

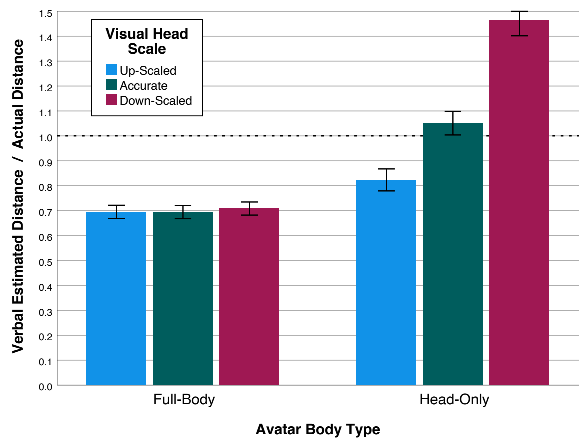
4.1 Head-Only Avatars

Figure 5 shows the pooled results for the head-only avatar representations. The x-axes show the three avatar distances, and the y-axes show the judged target distances. The gray diagonal line indicates expected results for veridical distance judgments. The colored lines show the distribution of judged distances in the different conditions.

We found a significant main effect of *visual head scale* on distance judgments for the triangulation distance measure, $F(2, 38) = 15.03$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.442$, as well as for the verbal distance measure, $F(1.08, 20.50) = 67.67$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.781$. Post-hoc tests showed significant differences between each two visual head scales



(a) Relative Triangulated Distance Judgments



(b) Relative Verbal Distance Judgments

Figure 4: Pooled responses for the two avatar body types in the experiment for the three visual head scales. The y-axes show the relative judged distances using the (a) triangulation method and (b) verbal method. The relative distances are computed by dividing the estimated distances by the actual distances. Hence, values close to 1 indicate accurate distance estimates, while values > 1 indicate distance overestimation and < 1 distance underestimation. The vertical error bars show the standard error of the mean.

(all $\boldsymbol{p < 0.002}$), except between *up-scaled* and *accurate* scales ($p = 0.76$) for triangulation responses, suggesting that "big heads" as used by Facebook's Project Horizon have less of an effect on distance estimation than down-scaled heads. This applies in particular to distances of five meters or more, which is where Choudhary et al. [10] found the biggest benefits of big heads for Social VR in terms of an improved perception of social cues like facial expressions.

We found no significant main effect of *verbal volume scale* on distance judgments for the triangulation distance measure, $F(2, 38) = 0.56$, $p = 0.58$, $\eta_p^2 = 0.029$, nor for the verbal distance measure, $F(2, 38) = 2.73$, $p = 0.08$, $\eta_p^2 = 0.126$. However, the trend in the results for verbal responses suggests that further experimentation with an increased participant sample may show a small effect of verbal volume scales on distance estimates. Overall, the results indicate that verbal volumes had no strong effects on distance estimation, suggesting that voices can be scaled in Social VR without having to worry about degraded distance estimation.

We found no interaction effects between any of the independent variables except between *verbal volume scale* and the avatar's *dis-*

tance for verbal responses, $F(4,76) = 3.09$, $\boldsymbol{p} = \mathbf{0.021}$, $\eta_p^2 = 0.140$, indicating that participants' tendency to verbally overestimate distances was aggravated the farther the avatar was away from them.

4.2 Full-Body Avatars

Figure 6 shows the pooled results for the full body avatar representations. The *x*-axes show the avatar distances, and the *y*-axes show the judged target distances. The colored lines show the distribution of judged distances in the different conditions.

We found no significant main effect of *visual head scale* on distance judgments for the triangulation distance measure, $F(2,34) = 3.16$, $p = 0.055$, $\eta_p^2 = 0.157$, nor for the verbal distance measure, $F(2,34) = 1.62$, $p = 0.21$, $\eta_p^2 = 0.087$. However, the trend in the results for triangulation responses suggests that an increased participant sample may show a small effect of visual head scales on distance estimates. Compared to the results for head-only avatar representations, the results for full-body avatars indicate that distance estimates were largely not or less affected by the scaling of the avatar's head. The results suggest that benefits related to the findings of Choudhary et al. [10] for scaled cues in Social VR can be gained for full-body avatars without negative effects on distance estimation.

We found no significant main effect of *verbal volume scale* on distance judgments for the triangulation distance measure, $F(2,34) = 0.23$, $p = 0.79$, $\eta_p^2 = 0.014$, nor for the verbal distance measure, $F(2,34) = 1.13$, $p = 0.34$, $\eta_p^2 = 0.062$. Similar to the results for head-only avatars, our results consistently indicate that verbal volumes had no discernible effect on distance estimation, suggesting that voices may be scaled in Social VR without degraded distance estimation.

We found no interaction effects between any of the independent variables except between *verbal volume scale* and the avatar's *distance* for triangulation responses, $F(4,68) = 2.57$, $\boldsymbol{p} = \mathbf{0.046}$, $\eta_p^2 = 0.131$.

4.3 Questionnaires

We measured a mean pre-SSQ sickness score of $M = 14.5$ ($SD = 27.7$) before the experiment and a mean post-SSQ score of $M = 53.5$ ($SD = 57.9$) after the experiment. On an absolute scale, these post-SSQ scores indicate a low amount of simulator sickness. The increase in simulator sickness symptoms was significant, $t(37) = 3.92, \boldsymbol{p} < \mathbf{0.001}$.

The mean Slater-Usoh-Steed questionnaire score for participants' sense of presence in the virtual environment was $M = 4.1$ ($SD = 0.76$), which suggests a reasonably high level of presence [48].

The mean score for the reported sense of feeling co-present in the virtual environment was $M = 3.7$ ($SD = 0.94$) for the full-body avatar and $M = 3.6$ ($SD = 0.87$) for the head-only avatar for the co-presence questionnaire by Garau and Slater [17]. These scores indicate a high sense of feeling co-present in the virtual environment.

5 Discussion

In this section, we refer back to our initial research questions (see Section 1) and summarize the main findings of our study. We discuss implications of the two body representations with scaled embodied cues on distance perception in Social VR, while also addressing limitations of our experiment.

5.1 Scaling of Visual Embodied Cues Changes Interpersonal Distance Estimation

In response to **RQ1**, we found that our tested visual head scales had a significant and strong effect on distance estimation for *head-only* avatar representations. The effects were straightforward in that participants judged a larger head to be closer and a smaller head to be farther away. While the magnitude of these effects differed between triangulated and verbal responses, together with the inherent biases of these response methods, both showed the same relative effects.

This finding is likely a result of a phenomenon called *size constancy* in perceptive psychology [33], indicating that humans perceive an object as having a fixed size, despite it being projected onto a larger or smaller area of our retina. In other words, an increased or decreased retinal size of a known object is usually not perceived as a change in its scale but rather as a change in its distance from the observer. While retinal size is not the only cue for distance perception, it is known to dominate other depth cues when they are in conflict, such as accommodation, convergence, or binocular disparity. This is in particular the case for stereoscopic display environments due to their inherent vergence-accommodation conflicts [8].

Our finding that visual head scales in Social VR can strongly affect the perceived distance to head-only avatars indicates their practical importance for such spaces. Our results show that designers and practitioners working with head-only avatars in Social VR need to carefully consider how they want to scale their avatar heads. Upscaled or down-scaled head sizes could in fact "correct" distance overestimation or underestimation effects that are inherent to the VR technology used (see Figure 4). On the other hand, practitioners should aim to avoid unintentional effects on distance perception by limiting the amount of head scaling.

In contrast to our results for head-only avatars, we found no observable effect of visual head scales on *full-body* representations. Due to the missing body parts for the head-only avatars, participants were not able to make relative size comparisons for the scaled head conditions—potentially leading to overestimation. The shift towards overestimation for our head-only avatars is in line with previous work on distance estimation with floating objects [38], but seems to imply a higher magnitude, which may be related to the known size cues of the human body parts in this experiment. While testing with a larger study population might reveal small/moderate effects related to such scaling, we did not find any strong effects as for head-only avatars. This is quite a positive finding for designers in Social VR as it implies that "big head" methods based on full-body avatars like those employed by Choudhary et al. [10] or Facebook's Project Horizon can be applied without having to worry about negative repercussions of such head scalings on distance perception.

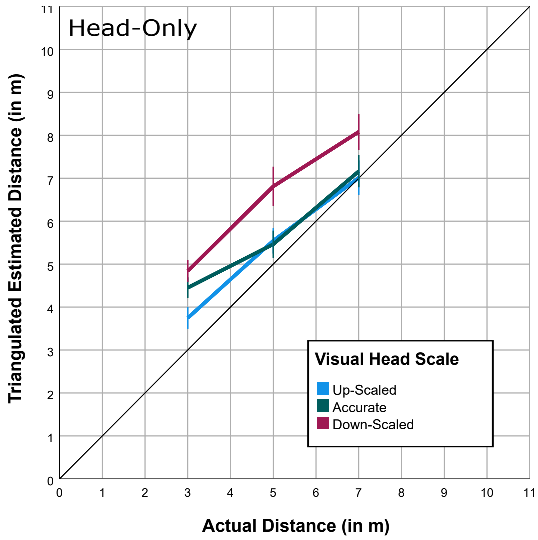
5.2 Scaling of an Avatar's Verbal Volume Levels in Social VR has no Discernible Effect on Interpersonal Distance Estimation

Regarding **RQ2** we can say that the tested audio volumes of the avatars' verbal cues had no noticeable effect on our results. We cannot rule out that audio volumes may affect distance estimation in other scenes or situations [2, 14, 15], such as in Social VR "in the dark," but our results are in line with the literature in that we found that visual distance cues tend to dominate audio cues when they are in conflict [19, 20, 44, 49], which was true at least for the verbal audio cues we tested in our Social VR setup.

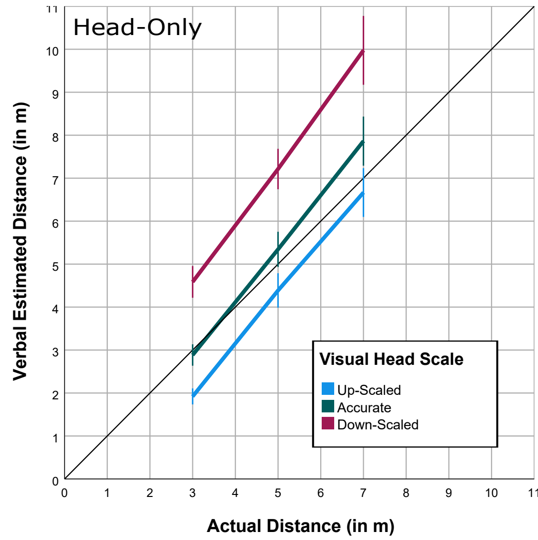
This is quite a positive result for practitioners in Social VR as it implies that audio volumes can be scaled without negative effects on distance estimation. However, we would like to acknowledge that we only tested audio cues from a single avatar in the virtual environment. For follow-up studies, it would be informative to look at spatial constellations of multiple avatars, each speaking at their own volume to fully understand if one may scale their volume levels independently or has to match relative volume differences in line with their constellations. Hearing two avatars speak at different volumes may provide important distance cues related to their interposition.

5.3 Effects of Interpersonal Distance Estimation Depend on Social VR Avatar Representations

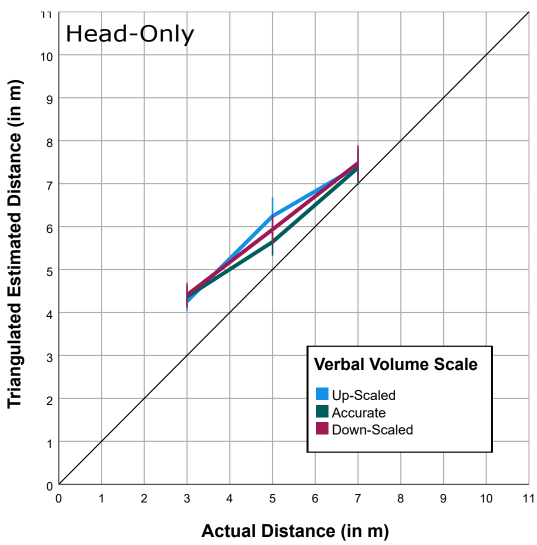
In response to **RQ3**, as already alluded to in the discussion above, we can say that our distance estimation results depended strongly on



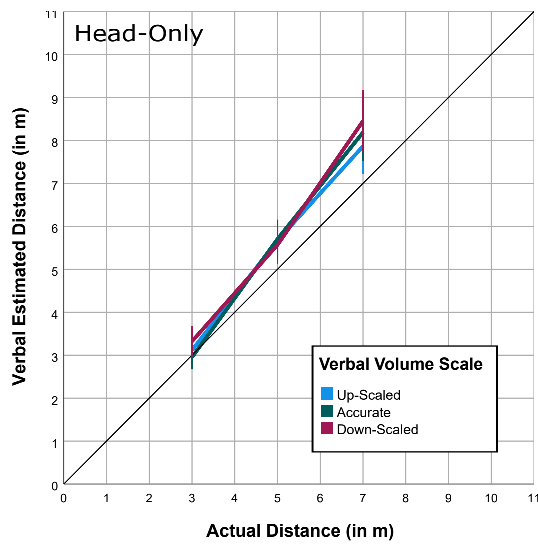
(a) Triangulation Measure — Visual Head Scaling



(b) Verbal Measure — Visual Head Scaling



(c) Triangulation Measure — Verbal Volume Scaling



(d) Verbal Measure — Verbal Volume Scaling

Figure 5: Pooled responses for the *head-only* avatar body type in the experiment. The *x*-axes show the three tested avatar distances. The *y*-axes show the judged target distances for (a+c) the triangulation distance measure and (b+d) the verbal distance measure, respectively. The colored lines show the distribution of judged distances for the different scaling methods for (a+b) visual head scales and (c+d) verbal volume scales, respectively. The vertical error bars show the standard error of the mean.

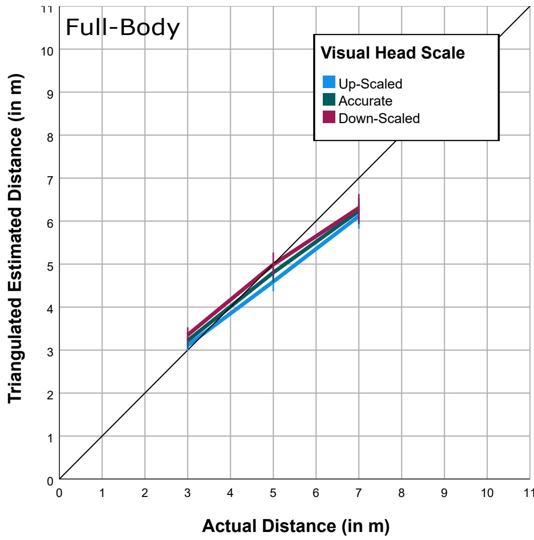
the avatar representation, in particular for visual head scales. The two tested extremes of avatar body representations (head-only and full-body) did indeed show different results.

A limitation of our study is that we did not sample and test the intermediate levels between these extremes. We tested head-only representations because these are common among low-cost HMD setups with head-only tracking, and we tested full-body representations because these are common among professional full-body-tracked HMD setups. On the range from head-only to full-body avatars, there are more versions like "head and hands," which are characterized by a tracked HMD and hand-held controllers, or "head and hands and torso," such as Facebook's Project Horizon, where the torso is not tracked but presented as a virtual reference to ground the movements of the head and hands around a centered torso object. Based on our results for the extremes, what we expect to see is a gradual interpolation of the distance estimates at the intermediate levels. It stands to reason that the more embodied cues are available, the more accurate distance estimation may be. However, these assumptions should be verified in future work.

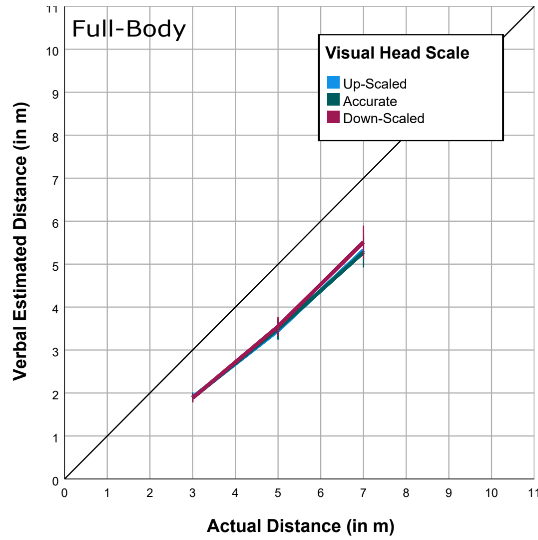
While not commonly the case in Social VR at the moment, we would also like to point out that it is possible to scale other body parts than one's head, such as one's hands, which may prove useful similar to up-scaled heads. Related cue conflicts should be carefully analyzed.

6 Conclusion

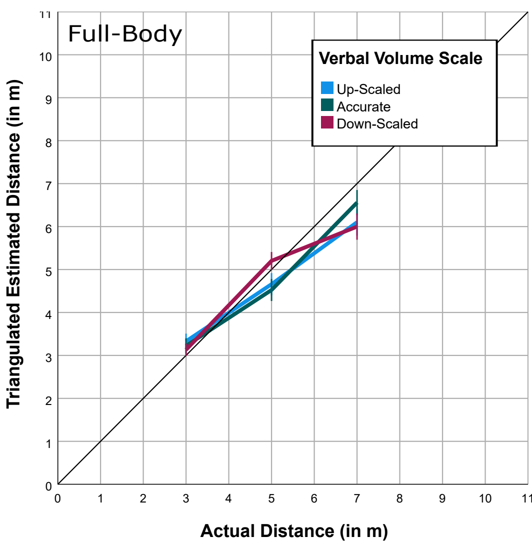
In this paper we analyzed human perception of avatar distance in Social VR with a controlled human-subject study. We compared two Social VR avatar representations, three visual head scales, and three verbal volume scales at different distances from social space



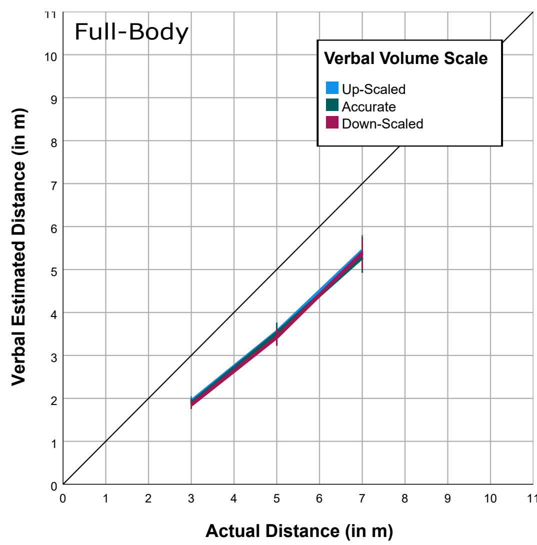
(a) Triangulation Measure — Visual Head Scaling



(b) Verbal Measure — Visual Head Scaling



(c) Triangulation Measure — Verbal Volume Scaling



(d) Verbal Measure — Verbal Volume Scaling

Figure 6: Pooled responses for the *full-body* avatar body type in the experiment. The x-axes show the three tested avatar distances. The y-axes show the judged target distances for (a+c) the triangulation distance measure and (b+d) the verbal distance measure, respectively. The colored lines show the distribution of judged distances for the different scaling methods for (a+b) visual head scales and (c+d) verbal volume scales, respectively. The vertical error bars show the standard error of the mean.

to far-public space. Our results show that visual head scale had a significant effect on distance judgments, which differed for the tested avatar representations, indicating that Social VR environments need to weight off their effects on users. We also found that verbal volume scales did not have a noticeable effect on distance judgments, which leads us to suggest that voices in Social VR can be scaled without immediate repercussions on spatial estimates.

In future work, we believe it would be informative to extend these results to less controlled interactive experiences involving social behavioral dynamics, such as groups of Social VR users mixing and mingling in a shared virtual space. We hypothesize that scaled visual cues, in particular scaled heads, will change how much personal space users maintain from one another, which would be informative for the design of virtual meeting spaces, their size and structure, as well as constraints when designing avatar representations for professional Social VR spaces.

Acknowledgments

This material includes work supported in part by the National Science Foundation under Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the Office of Naval Research under Award Numbers N00014-17-1-2927 and N00014-18-1-2964 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting institutions.

References

[1] Facebook Horizon | Oculus. https://www.oculus.com/facebook-horizon/, Nov 13, 2020.

[2] P. W. Anderson and P. Zahorik. Auditory/visual distance estimation: accuracy and variability. *Frontiers in psychology*, 5:1097, 2014.

[3] F. Argelaguet Sanz, A. H. Olivier, G. Bruder, J. Pettre, and A. Lecuyer. Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments. In *Proceedings of IEEE Virtual Reality (VR)*, pp. 75–80, 2015.

[4] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.

[5] M. W. Bell, S. Smith-Robbins, and G. Withnail. *This is Not a Game – Social Virtual Worlds, Fun, and Learning*, pp. 177–191. Springer London, London, 2010. doi: 10.1007/978-1-84996-047-2_10

[6] B. Bodenheimer, J. Meng, H. Wu, G. Narasimham, B. Rump, T. P. McNamara, T. H. Carr, and J. J. Rieser. Distance estimation in virtual and real environments using bisection. In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, APGV '07, p. 35–40. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1272582.1272589

[7] G. Bruder, F. Argelaguet, A.-H. Olivier, and A. Lécuyer. Distance Estimation in Large Immersive Projection Systems, Revisited. In *Proceeding of IEEE Virtual Reality (VR)*, pp. 27–32, 2015.

[8] G. Bruder, F. Argelaguet, A.-H. Olivier, and A. Lécuyer. Cave size matters: Effects of screen distance and parallax on distance estimation in large immersive display setups. *Presence: Teleoperators and Virtual Environments*, 25(1):1–16, 2016.

[9] A. D. Cheok, K. H. Goh, W. Liu, F. Farbiz, S. W. Fong, S. L. Teo, Y. Li, and X. Yang. Human pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing. *Personal and ubiquitous computing*, 8(2):71–81, 2004.

[10] Z. Choudhary, K. Kim, R. Schubert, G. Bruder, and G. F. Welch. Virtual big heads: Analysis of human perception and comfort of head scales in social virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 425–433. IEEE, 2020.

[11] S. H. Creem-Regehr and B. R. Kunz. Perception and action. *WIREs Cognitive Science*, 1(6):800–810, 2010. doi: 10.1002/wcs.82

[12] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception*, 34(2):191–204, 2005. PMID: 15832569. doi: 10.1068/p5144

[13] J. E. Cutting, P. M. Vishton, and P. A. Braren. How we avoid collisions with stationary and moving objects. *Psychological Review*, 102(4):627–651, 1995. doi: 10.1037/0033-295X.102.4.627

[14] D. J. Finnegan, E. O'Neill, and M. J. Proulx. Compensating for distance compression in audiovisual virtual environments using incongruence. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 200–212, 2016.

[15] D. J. Finnegan, E. O'Neill, and M. J. Proulx. An approach to reducing distance compression in audiovisual virtual environments. In *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–6. IEEE, 2017.

[16] S. S. Fukusima, J. M. Loomis, and J. A. Da Silva. Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1):86, 1997.

[17] M. Garau, M. Slater, D.-P. Pertaub, and S. Razzaque. The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators & Virtual Environments*, 14(1):104–116, 2005.

[18] Giant Bomb. Big Head Mode. https://www.giantbomb.com/big-head-mode/3015-403/, Nov 6, 2019.

[19] M. Gonzalez-Franco, A. Maselli, D. Florencio, N. Smolyanskiy, and Z. Zhang. Concurrent talking in immersive virtual reality: on the dominance of visual speech cues. *Scientific reports*, 7(1):1–11, 2017.

[20] M. Gorzel, D. Corrigan, G. Kearney, J. Squires, and F. Boland. Distance perception in virtual audio-visual environments. In *25th UK Conference of the Audio Engineering Society: Spatial Audio In Today's 3D World (2012)*, pp. 1–8, 2012.

[21] E. Hall. A system for the notation of proxemic behavior. *American Anthropologist*, 65(5):1003–1026, 1963.

[22] E. Hall. *The Hidden Dimension: man's use of space in public and in private*. Anchor Books, 1969.

[23] T. Iachini, Y. Coello, F. Frassinetti, G. Ruggiero, and F. Ric. Body Space in Social Interactions: A Comparison of Reaching and Comfort Distance in Immersive Virtual Reality. *PLoS ONE*, 9(11):e111511, nov 2014. doi: 10.1371/journal.pone.0111511

[24] L. B. Jaloza. Inside facebook reality labs research: The future of audio. https://tech.fb.com/inside-facebook-reality-labs-research-the-future-of-audio/, Sep 3, 2020.

[25] J. A. Jones, E. A. Suma, D. M. Krum, and M. Bolas. Comparability of narrow and wide field-of-view head-mounted displays for medium-field distance judgments. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '12, p. 119. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2338676.2338701

[26] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[27] E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt. Measurement protocols for medium-field distance perception in large-screen immersive displays. In *2009 IEEE Virtual Reality Conference*, pp. 107–113. IEEE, 2009.

[28] J. M. Loomis, J. M. Knapp, et al. Visual perception of egocentric distance in real and virtual environments. *Virtual and adaptive environments*, 11:21–46, 2003.

[29] J. M. Loomis and J. W. Philbeck. Measuring spatial perception with spatial updating and action. *Embodiment, Ego-space, and Action*, pp. 1–43, 2008.

[30] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, 2012.

[31] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. C. M. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafin. Missing the point: An exploration of how to guide users' attention during cinematic virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, p. 229–232. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2993369.2993405

[32] C. C. Pagano, R. P. Grutzmacher, and J. C. Jenkins. Comparing verbal and reaching responses to visually perceived egocentric distances. *Ecological Psychology*, 13(3):197–226, 2001.

[33] S. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, 1999.

[34] S. Parsons. Learning to work together: Designing a multi-user virtual reality game for social collaboration and perspective-taking for children with autism. *International Journal of Child-Computer Interaction*, 6:28 – 38, 2015. doi: 10.1016/j.ijcci.2015.12.002

[35] A. Peer and K. Ponto. Evaluating perceived distance measures in room-scale spaces using consumer-grade head mounted displays. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 83–86. IEEE, 2017.

[36] D. R. Proffitt. Distance perception. *Current Directions in Psychological Science*, 15(3):131–135, 2006.

[37] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments - A review. *ACM Computing Surveys*, 46(2):1–40, 2013. doi: 10.1145/2543581.2543590

[38] C. S. Rosales, G. Pointon, H. Adams, J. Stefanucci, S. Creem-Regehr, W. B. Thompson, and B. Bodenheimer. Distance judgments to on- and off-ground objects in augmented reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 237–243, 2019. doi: 10.1109/VR.2019.8798095

[39] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE Computer Graphics and Applications*, 38(2):31–43, 2018. doi: 10.1109/MCG.2018.193142628

[40] S. Serafin, N. C. Nilsson, E. Sikstrom, A. De Goetzen, and R. Nordahl. Estimation of detection thresholds for acoustic based redirected walking techniques. In *2013 IEEE Virtual Reality (VR)*, pp. 161–162, 2013. doi: 10.1109/VR.2013.6549412

[41] M. J. Sinai, W. K. Krebs, R. P. Darken, J. Rowland, and J. McCarley. Egocentric distance perception in a virutal environment using a perceptual matching task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 43, pp. 1256–1260. SAGE Publications Sage CA: Los Angeles, CA, 1999.

[42] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535), 2009.

[43] F. Steinicke, N. Lehmann-Willenbrock, and A. L. Meinecke. A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality. In *Symposium on Spatial User Interaction*, SUI '20. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3385959.3422699

[44] Y. Sugita and Y. Suzuki. Implicit estimation of sound-arrival time. *Nature*, 421(6926):911–911, 2003.

[45] W. Thompson, R. Fleming, S. Creem-Regehr, and J. K. Stefanucci. *Visual perception from a computer graphics perspective*. CRC press, 2011.

[46] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, 13(5):560–571, 2004. doi: 10.1162/1054746042545292

[47] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators & Virtual Environments*, 13(5):560–571, 2004.

[48] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503, 2000.

[49] J. Vroomen and B. De Gelder. Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. *The handbook of multisensory processes*, 3(4):1–23, 2004.

[50] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1643–1652, 2018. doi: 10.1109/TVCG.2018.2794629

[51] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Trans. Appl. Percept.*, 6(2), Mar. 2009. doi: 10.1145/1498700.1498702