Virtual Big Heads: Analysis of Human Perception and Comfort of Head Scales in Social Virtual Reality

Zubin Choudhary* Kangsoo Kim† Ryan Schubert‡ Gerd Bruder§ Gregory F. Welch¶

University of Central Florida

Figure 1: Examples of virtual avatars with different head scales: (a) normal head scale (1x), (b) big head (2x) with the *fixed neck height* scaling method, in which the head protrudes upwards from the neck, (c) small head (0.5x) with the *fixed neck height* scaling method, (d) big head (2x) with the *fixed eye height* scaling method, in which the body is scaled down to maintain the same eye height, and (e) small head (0.5x) with the *fixed eye height* scaling method, in which the body is scaled up to maintain the eye height.

Abstract

Virtual reality (VR) technologies provide a shared platform for collaboration among users in a spatial context. To enhance the quality of social signals during interaction between users, researchers and practitioners started augmenting users' interpersonal space with different types of virtual embodied social cues. A prominent example is commonly referred to as the *"Big Head"* technique, in which the head scales of virtual interlocutors are slightly increased to leverage more of the display's visual space to convey facial social cues. While beneficial in improving interpersonal social communication, the benefits and thresholds of human perception of facial cues and comfort in such Big Head environments are not well understood, limiting their usefulness and subjective experience.

In this paper, we present a human-subject study that we conducted to understand the impact of an increased or decreased head scale in social VR on participants' ability to perceive facial expressions as well as their sense of comfort and feeling of "uncanniness." We explored two head scaling methods and compared them with respect to perceptual thresholds and user preferences. We further show that the distance to interlocutors has an important effect on the results. We discuss implications and guidelines for practical applications that aim to leverage VR-enhanced social cues.

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

*e-mail: zubinchoudhary@knights.ucf.edu
†e-mail: kangsoo.kim@ucf.edu
‡e-mail: rschuber@ist.ucf.edu
§e-mail: bruder@ucf.edu
¶e-mail: welch@ucf.edu

1 Introduction

Over the last decade, large strides have been made to improve augmented reality (AR) and virtual reality (VR) technologies, and new opportunities have arisen for applications in fields such as simulation, training, therapy, and healthcare. For many of these, collaboration between multiple users is an important aspect of the experience. In our daily (real) life, we leverage a combination of both verbal and nonverbal cues, including body posture and facial expressions, to communicate with other individuals. Collaborative environments try to facilitate effective communication of embodied social cues, and researchers have introduced different methods to accentuate such cues in collaborative experiences with VR/AR technologies [21, 25, 27–30].

A prime example of such accentuation is the so-called *Big Head* technique, aka *Donkey Kong* mode, due to the game character's proportions [11], where heads of interlocutors are disproportionately up-scaled relative to the rest of the body. An underlying assumption of this technique is that humans have a certain tolerance for seeing body parts at different sizes, and up-scaling certain parts can improve their respective effectiveness in conveying social cues. Aspects of this technique originated among game developers, who were looking for a solution to the problem of limited screen space and low pixel resolutions when trying to present game characters to players. Presenting heads at their regular scale with respect to the rest of the simulated game environment generally made it hard to make out differences in characters' heads and facial expressions. Up-scaling only the heads of the game characters enabled game developers to maintain a low scale for the environment while allowing players to distinguish characters and their respective visual game states. While, traditionally, the low resolution in desktop or game console display environments was the governing factor, which has subsequently been largely overcome, we are seeing similar challenges with respect to VR/AR displays today. Not surprisingly, some collaborative environments, such as Facebook's Social VR [34], have already started to leverage the approach, although the reasoning behind their design choices in terms of head scales and body proportions remains obscure.

In this paper, we present a human-subject study aimed at understanding the importance and influence of an increased or decreased head scale in social VR environments on a user's perception and sense of comfort. We identified preferences and thresholds for head scales over a range of distances up to ten meters in two task contexts. We present and evaluate two alternative head scaling methods, and discuss their respective benefits and drawbacks.

With this work, we aim to contribute to the research community by providing answers to the research questions below:

- **RQ1**: What are the thresholds for the amounts of head scale that can be introduced without limiting a user's ability to perceive facial expressions?
- **RQ2**: What are the thresholds that govern a user's sense of comfort or "uncanniness"?
- **RQ3**: How do the thresholds change with respect to distance from intimate space to public space?

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

2 Related Work

In this section, we present related work on social VR in general, sharing and perceiving social signals through embodied virtual avatars, and users' sense of comfort when viewing or spatially interacting with another avatar.

2.1 Social Virtual Reality

Social VR platforms, such as Second Life[1], VRChat[2], or Facebook Horizon[3], enable users to meet and share experiences together in an immersive virtual environment as if the users were together in a shared real place. As social VR increases in popularity, public interest in VR, and its consumer market, also continues to grow.

One core aspect of social VR is the idea that users can create or customize virtual characters to represent themselves within the space of the virtual world. These virtual self-representations, which are frequently humanoid, are referred to as *avatars* and can have an effect on other users' perception of some aspects of the real user behind the avatar, such as his or her personality or lifestyle [3].

Social VR platforms facilitate different types of social behaviors, such as communication (e.g., teleconferencing) or collaboration in real or virtual tasks (e.g., collaborative content creation). Most social VR settings use a single environment scale, while others allow users to change it [22, 36] (see movie "Ant Man," 2015). Although both vocal and textual communication are generally integral to social interactions within social VR contexts, visual aspects of communication, such as appearance, gestures, or expressions, also play an important role, by themselves and in addition to speech or text. Users implicitly recognize the importance of the visual characteristics of their avatars, particularly those aspects most noticeable or relevant to other users, as evident in the effort put into avatar customization [9].

2.2 Social Signals in Virtual Avatars

Embodied virtual self-representations in the form of a user's virtual avatar can reproduce some of the user's physical movements (e.g., if tracked in real time or controlled indirectly) to provide a higher sense of social presence in an immersive virtual environment by providing a mechanism for introducing often nuanced but important non-verbal aspects of communication [32].

Similar to gestures and movement, facial expressiveness and the resulting ability to convey emotion is another powerful source of non-verbal communication and influence in both real and virtual interpersonal interactions. Cafaro et al. found that non-verbal social signals such as smiling, gaze direction, and the proximity behavior of a virtual human affected a user's perception of that virtual human's personality and interpersonal attitude during a first encounter situation, even after a very short time of interaction [5]. Furthermore, emotion display through the facial expressions of a virtual human has been shown to have an effect on a user's behavior during social interaction, such as the likelihood to concede during a negotiation task [7]. Likewise, accurately simulating realistic eye gaze behavior for a virtual avatar can increase the level of participation in a conversation with the avatar [6]. Hager and Ekman showed that the accuracy of facial expression recognition when presented with a fixed-size image of a human face declined as the distance of the image from a viewer increased, decreasing the effective retinal size of the face to the viewer [12].

Given the importance of non-verbal social signals, such as facial expressions, in users' perceptions and behavior in social VR contexts, it is not surprising that some existing social VR platforms have explored mechanisms for greatly emphasizing or increasing the visibility of such signals, such as increasing the head size for users' avatars [11, 34], in situations where limited resolution or varied interpersonal distance may diminish the effectiveness of such cues. However, in addition to increasing the visibility of facial expressions, the up-scaling of avatar heads may have other impacts on users' perception and social interaction. Because of this, one aspect of our study also explored how comfortable users would be when interacting with an avatar with manipulated proportions.

2.3 Avatar Realism and the Sense of Comfort

There has been a strong interest among psychologists and practitioners in social VR and related fields in understanding peoples' sense of comfort (or perceived "uncanniness" or "creepiness") when presented with artificial representations of humans, including their perceived "humanness," and associated effects. In particular, the *Uncanny Valley* proposes that increasing a human representation's visual realism may not necessarily result in an increased sense of comfort when interacting with that entity [26]. Along these lines, with respect to head proportions for a humanoid robot, DiSalve et al. found that people were more comfortable with a head that was *less* realistic, with a wider head and wider eye spacing than a realistic human head, supporting the idea that less realistic proportions may in fact not detract from the effectiveness of a humanoid character [8].

Our sense of comfort with other real humans or their representations in a social situation is also dependent on *distance*: Hall presented a system to standardize how humans perceive and structure the space in our immediate surroundings from *intimate* space (up to 0.46 m), *personal* space (up to 1.22 m), *social* space (up to 3.7 m), and *public* space [13]. Hall further stated that humans tend to maintain some "buffer space" around themselves and each other. This has been shown to apply to both real and virtual entities [1, 24]. People in the real world but also in an immersive virtual environment generally feel uncomfortable if another person, in particular a stranger, invades their personal space or even their intimate space [15]. In such cases, humans generally experience an activation of the amygdala in response to this violation of social norms [16]. Bailenson et al. investigated the interpersonal distance maintained between participants and virtual avatars and found that participants gave more personal space when facing the back of a avatar than the front, when an avatar engaged in mutual gaze with them, and when the avatar invaded their personal space [2]. A breach of such social norms can reduce a person's sense of social presence—in particular comfort—with the entity.

[1] https://secondlife.com
[2] https://www.vrchat.com
[3] https://www.oculus.com/facebookhorizon

3 Experiment

In this section we present our user study, in which we evaluate the effects of a virtual avatar's head scale on participants' perception of the avatar in a social VR context with an immersive VR head-mounted display (HMD).

3.1 Participants

After initial pilot tests, we estimated the effect size of the expected strong effects, and based on a power analysis, we made the decision to recruit 23 participants from our university community (16 male and 7 female; ages between 18 and 31, $M = 23.96$ $SD = 4.11$). 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 known visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance. 17 participants had used a VR HMD before, and eight of them had experience with social VR. 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

To investigate the participants' perception of different avatar head scales in a shared VR environment, we prepared a virtual space containing an avatar in which participants could view as well as adjust the scale of the avatar's head. In this section, we describe the details of the avatar and the study settings for the experiment.

A life-size 3D female virtual human model was created and rigged in Autodesk Maya and Blender as the virtual avatar for our study (see Figure 1). The model was imported into the Unity game engine (version 2019.2.9.f1) and placed in a virtual hallway environment. For the virtual environment, we created a simulated hallway with the dimensions 5 m (width) × 3 m (height) × 15 m (length) as shown in Figure 2, so that participants could see the avatar at different distances in the hallway. The character had an neutral idle standing animation in the hallway, repeating slight body movements, and was able to perform various facial expressions using the blendshapes on her face, e.g., lip and eye brow movements, etc. Among different possible facial expressions, we decided to use four specific expressions: happy, sad, angry, and skeptical (see Figure 3), and the avatar performed these expressions continuously in a loop throughout the experiment. We confirmed that the visibility of the expressions varied with the prepared facial expressions, e.g., happy is the most noticeable, and skeptical is the most difficult to identify, through an internal pilot test. At the start of the experiment, the avatar's body height was adjusted to match that of the participant. The scaling methods were applied relative to the body height.

The avatar and the surrounding virtual environment were displayed through an immersive VR HMD, HTC Vive Pro, which was connected to a host PC (Intel Core i7-7820HK CPU @ 2.90 GHz, 32Gb Ram, NVIDIA GTX 1070 graphics card, Windows 10 Pro) for the experimenter to run the program and monitor the participant's view and activities in the VR environment. During the experiment, participants were standing on a marked location on the floor in our lab environment, wearing the HMD, and could adjust the scale of the avatar head using the Vive Pro Controller (see Figure 4). They could scale up or down the head by pressing the up or down button on the trackpad, and when they wanted to make a decision on the head scale, they could press the trigger button at the bottom of the controller, which saved their chosen head scale for the condition. Two Lighthouse units (HTC SteamVR Base Station 2.0) were set up in the experimental space to track the HMD's pose (position and orientation), and the tracked pose was used to render the virtual environment according to the participant's viewpoint, so that the participants could feel as if they were present in the environment facing the virtual avatar. After an initial familiarization period, participants were asked to remain standing at a fixed position in the lab. The HMD provided a 110 degree vertical and 100 degree horizontal field of view, and had a resolution of 1440×1600 pixels per eye at a refresh rate of 90 Hz. The corresponding angular resolution was 15 px/degree.

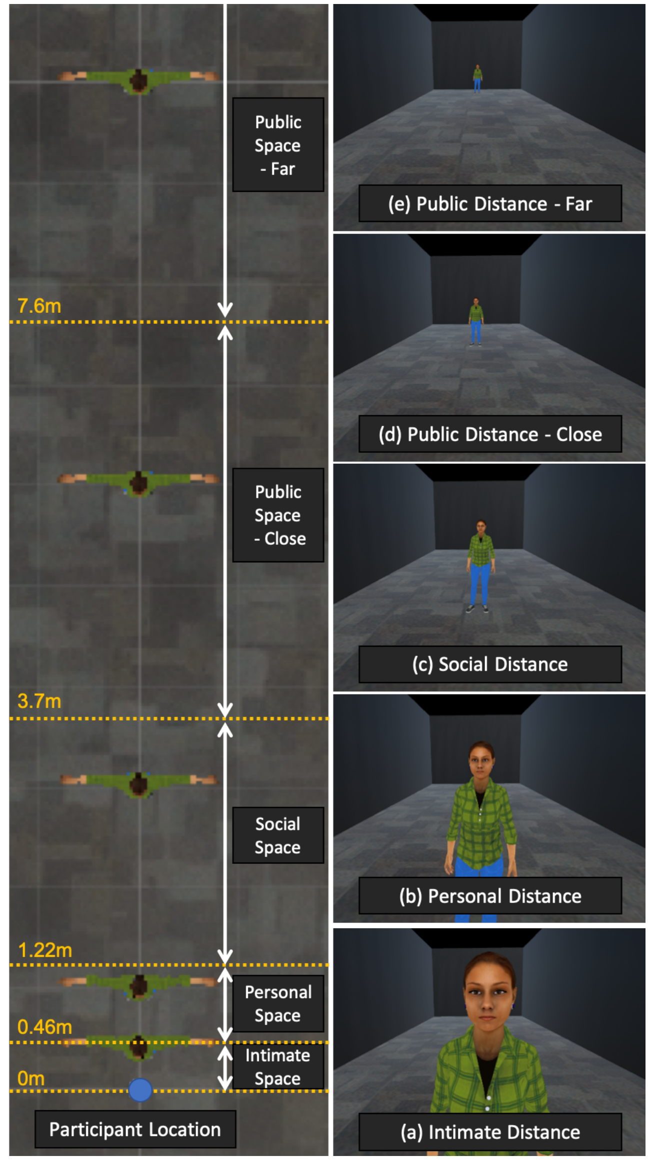


Figure 2: Illustration of the tested distances and proxemics in the experiment. Screenshots on the right show the avatar at distances of (a) 0.4 m (intimate space), (b) 1 m (personal space), (c) 3 m (social space), (d) 6 m (close public space), and (e) 10 m (far public space).

3.3 Methods

3.3.1 Study Design

We used a 2×2×5 full-factorial within-subjects design with three factors: *scaling method*, *task* and *distance*; each tested on three *target scales* as described below.

- **Scaling Method (2 levels)**: We developed two mechanisms to adjust the avatar's head scale relative to the rest of the avatar's body (see Figure 1).

Figure 3: The avatar's four facial expressions used for the experiment: (a) happy, (b) sad, (c) angry, and (d) skeptical.

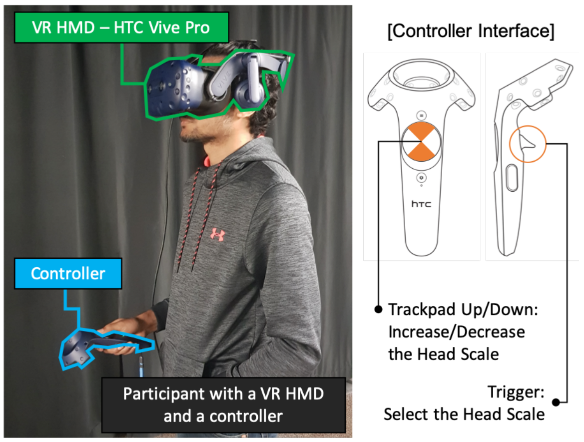


Figure 4: Annotated photo showing a participant in the experiment wearing the HMD and holding the controller during the task. The controller interface to adjust the scale of the avatar's head is shown on the right.

  - **Fixed Neck Height**: In our first scaling method, the avatar's head scale was changed upwards from the avatar's neck height, so that participants had to look up to the avatar's head when the head scale was increased. As a result, the eye height of the avatar changed depending on the head scale, meaning that with an increased head scale the avatar would look down at the participant, while a reduced head scale meant that the avatar would look up at the participant.
  - **Fixed Eye Height**: Considering the potential influence of a mismatch in the eye height of the avatar and the participant, we designed a second scaling method, in which up-scaling the head meant that the rest of the body was down-scaled, while down-scaling the head meant that the rest of the body was up-scaled. The result is that the head scale is adjusted as in the first scaling method, but the body scale is further adjusted so that the avatar always maintains the same eye height at all times.

- **Task Context (2 levels)**: We decided to test two separate task contexts in the avatar head scaling tasks.
  - **Facial Expression**: In this task context, participants were asked to adjust the scale of the avatar's head while focusing on their performance in recognizing the avatar's facial expressions. If the head scale was too small or too large, they would not be able to recognize the facial expressions. We prepared this task context considering the importance of understanding the social signals, particularly through facial expressions, from other users in social VR. We played four facial expressions in a loop, which are described in more detail in Section 3.2 (see Figure 3).
  - **Comfort Level**: In this task context, participants were asked to adjust the scale of the avatar's head while focusing on their perception of the avatar and their feeling of comfort when interacting with the avatar in social VR. If the head scale is too small or too large, a feeling of "uncanniness" (or "creepiness") tends to arise, which indicates that participants would not feel comfortable interacting with such an avatar in social VR.

- **Distance (5 levels)**: For the experiment, the avatar appeared at five different distances, considering the proxemics in social interactions based on Hall's Proxemics theory [14] (see Figure 2).
  - **Intimate**: Intimate space is considered for touching or whispering within a distance of 0.46 m, so we located the avatar at 0.4 m away from the participants.
  - **Personal**: Personal space is for interactions with close friends or family members within a distance of 1.22 m, so we located the avatar at 1 m away from the participants.
  - **Social**: Social space is for interactions with acquaintances within a distance of 3.7 m, so we located the avatar at 3 m away from the participants.
  - **Public-Close**: Public-Close space is for public speaking within a distance of 7.6 m, so we located the avatar at 6 m away from the participants.
  - **Public-Far**: Public-Far space is also for public speaking but targeting people in farther distances beyond 7.6 m, so we located the avatar at 10 m away from the participants.

  We asked the participants to scale the head size based on the following target scales.

- **Target Scales**: During each trial, participants were asked to either identify the *minimum*, *maximum*, or *ideal* head scale in the task contexts mentioned above.
  - **Minimum**: Participants were asked to identify and select the minimum head scale in the particular task context. For the facial expression task context, this meant that they would reduce the head scale until they could just barely recognize the facial expressions. For the comfort level task context, they reduced the head scale as far as they would still maintain a sense of visual comfort seeing the avatar in front of them, i.e., without feeling uncanny.
  - **Maximum**: Similar to the minimum target scale, participants were asked to identify and select the maximum head scale in the particular task context.
  - **Ideal**: Participants were asked to identify and select their preferred ideal head scale in the particular task context. For the facial expression task context, this meant the ideal head scale for recognizing the facial expressions. For the comfort level task context, this meant finding the ideal head scale without feeling uncanny.

In total, each participant experienced 60 trials in combinations of our factors while they were performing the tasks adjusting and selecting the avatar's head scale. A 2×2 Latin square was used to balance the order of the two scaling methods and two tasks for each participant. So as not to confuse participants, the task context for each task was fixed in the order of minimum, maximum, and optimum scale. The five distances were presented in random order to participants.

3.3.2 Procedure

Once participants arrived, a form of informed consent was provided, and they were asked to give their verbal consent to participate in the experiment. Afterwards, the experimenter verbally explained the study details with the task instructions, and made sure that the participants understood the tasks, in which they should adjust the scale of a virtual avatar's head for two different task contexts in an immersive virtual environment. The experimenter explained all the conditions that the participants would experience with respect to the factors, which we described in Section 3.3.1. Participants then donned an HTC Vive Pro HMD on which they could see the virtual avatar standing in a virtual hallway facing toward them. The virtual avatar was introduced to participants as a typical representation of a real human interlocutor in a social VR environment, such as those known from Facebook's Social VR. Before the actual experiment, there was a practice session so that the participants could have an idea about the avatar's facial expressions that they should look out for and how to change the avatar's head scale using a Vive controller for the experiment as described in Section 3.2. They could increase or decrease the head scale by pressing the up or down button on the trackpad of the controller, and make a decision on the head scale by pressing the trigger button, which would also advance the experiment to the next trial condition—different distance, target scale, task context, or scaling method.

Once participants were familiar with the interface and tasks, we started the experimental trials. For recognising the facial expressions, the virtual avatar exhibited four expressions in an infinite loop—happy, sad, angry, and skeptical (see Figure 3). The participants were asked to identify and select the head scales at the five different distances per target scale (minimum, maximum, and ideal scales) that corresponded to the current task context. After completing the three target scales they moved on to the next task context and scaling method based on the Latin square design. When the experiment was halfway completed, a two-minute break was provided to participants to minimize the effects of the HMD on participants' eyestrain or potential simulator sickness. After completing all conditions, they proceeded to complete a post-questionnaire, assessing their demographics and prior VR experience, and we asked their general perception and preference of the virtual avatar conditions as well as the reasoning behind their answers. Finally, the experiment ended with a monetary compensation.

3.4 Measures and Hypotheses

In this section, we describe the measures and hypotheses that we used for the experiment to understand the range of the avatar's head scales in the different conditions.

3.4.1 Avatar Head Scale

As described in Section 3.3, participants experienced the avatar head scaling tasks with 60 different conditions in terms of the four factors: scaling method, task context, target scale, and distance. We collected the participants' decisions on the avatar's head scale during the tasks to investigate how the four factors influence the participants' selection of the head scale. The general hypotheses we established were:

- **H1** For the facial expression task context, participants will increase the avatar's head scale as the distance between the avatar and themselves increases.
- **H2** For the comfort level task context, participants will maintain the avatar's head scale at the original scale even if the distance between the avatar and themselves increases.

For both task contexts and scaling methods, we expected to identify minimum and maximum thresholds showing a range of head scales over which avatar head scales can be varied in social VR.

With respect to Hypothesis **H1**, we further deliberated that, if the low resolution of the VR HMD denotes a cut-off for the recognition of facial expressions, we expected to see a linear relationship between the distance to the virtual avatar and the head scale chosen by our participants. In other words, if the facial expressions can be well recognized at a distance of one meter with a certain amount of screen space and pixels, participants would up-scale the head to twice its size if the avatar is presented at a distance of two meters, three-times its size at four meters, etc.

3.4.2 Preference of Scaling Method

We used a customized questionnaire to collect the participants' subjective preferences of the scaling methods with respect to the task contexts. On a 7-point scale from 1 (Fixed Neck Height) to 7 (Fixed Eye Height), we asked them which method the participants preferred in the facial expression task context and the comfort level task context. We also collected qualitative comments about the reasoning behind their preferences and their choices of head scales with respect to the distances. Regarding the preference of the scaling method, we had the following hypothesis considering the benefit of the fixed eye height method that renders the avatar's face at the same eye level as the participant's eyes, which makes it easier to see the avatar's facial expressions and potentially more comfortable due to more social eye-contact. The fixed eye height method also avoided any social issues caused by the virtual avatar looking down at the participants.

- **H3** For both task contexts, participants will prefer the scaling method with the fixed eye height compared to the method with the fixed neck height.

4 Results

The results are shown in Figure 5: (a) and (b) show the results when tasked with identifying a comfortable range of head scales, while (c) and (d) show the results when tasked with identifying head scales so that participants can still perceive the virtual avatar's facial expressions. The left column shows results for the *fixed neck height* scaling method, in which only the head size is scaled upwards/downwards, while the right column shows the results for the *fixed eye height* scaling method, which maintains eye level with the participants. The *x*-axes show the distances to the virtual avatar. The *y*-axes show the head scales defined by the participants; the results are presented on a logarithmic (base 2) scale on that axis. The colored red lines indicate the *maximum* and *minimum* head scales that the participants chose. The colored green area in between the thresholds indicates the range between those two extrema. The green lines indicate the *ideal* head scales for the distances, which the participants chose. The error bars indicate the standard error.

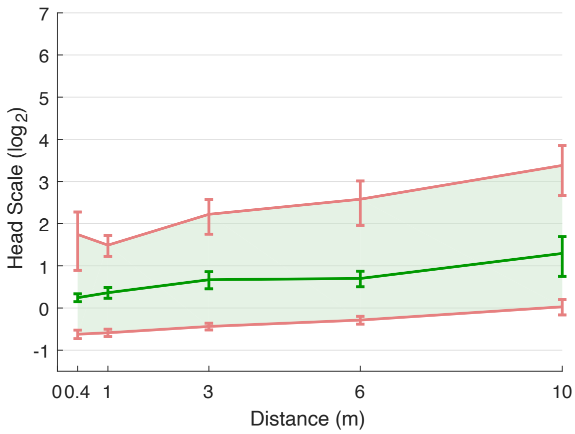
4.1 Avatar Head Scale

We analyzed the responses with repeated-measures ANOVAs and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. We confirmed the normality with Shapiro-Wilk tests at the 5% level and QQ plots. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated. We only report the significant effects.

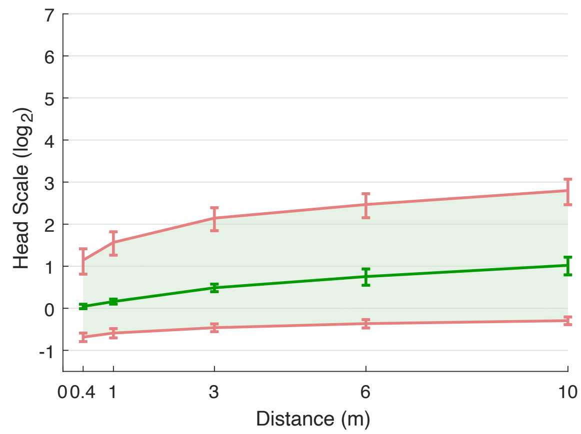
We found no interaction effect with a 2×2 ANOVA. We found a significant effect of **scaling method** on head scales, $F(1,21) = 7.93$, $\boldsymbol{p = 0.01}$, $\eta_p^2 = 0.27$, indicating that the *fixed eye height* scaling method resulted in larger head scales than the *fixed neck height* scaling method. We also found a significant effect of **task context** on head scales, $F(1,21) = 22.44$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.52$, indicating that the task to identify *facial expressions* resulted in larger head scales than the task to identify *comfort levels*.

4.1.1 Comfort Level

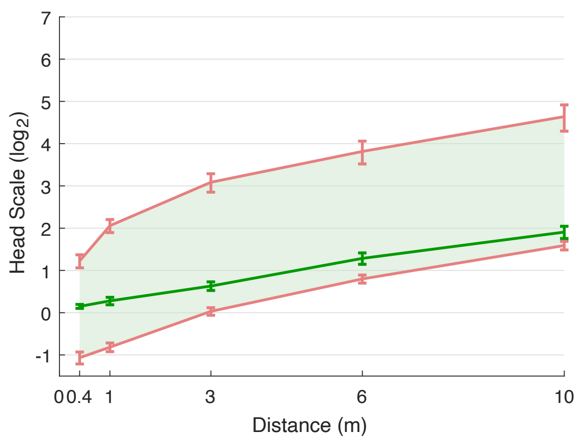
Fixed Neck Height. For the *fixed neck height* scaling method, we found a significant main effect of **distance** on head scales for the *minimum* task, $F(1.23,27.06) = 7.91$, $\boldsymbol{p = 0.006}$, $\eta_p^2 = 0.26$. Post-hoc tests showed that all pairwise comparisons were significant



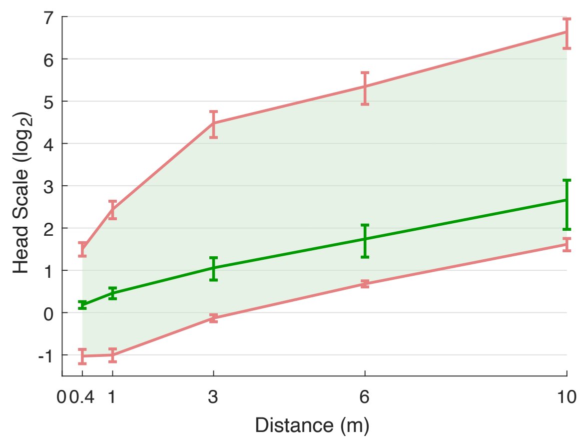
(a) Comfort Level Task — Fixed Neck Height Scaling



(b) Comfort Level Task — Fixed Eye Height Scaling



(c) Facial Expressions Task — Fixed Neck Height Scaling



(d) Facial Expressions Task — Fixed Eye Height Scaling

Figure 5: Results for the two task contexts (comfort level and facial expressions) and the two scaling methods (fixed neck height scaling and fixed eye height scaling). The x-axes show the distances to the virtual avatar; the y-axes show the indicated head scales by the participants (on a logarithmic scale with base 2). The red lines indicate the thresholds while the green line indicates ideal head scales.

($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1, 0.4 and 3, 3 and 6, 3 and 10, as well as 6 and 10.

We also found a significant main effect of **distance** on head scales for the *maximum* task, $F(1.11, 14.31) = 4.64$, $\boldsymbol{p = 0.038}$, $\eta_p^2 = 0.17$. All pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1, 3 and 6, as well as 6 and 10.

We found no significant main effect of **distance** on head scales for the *ideal* task, $F(1.15, 25.29) = 2.27$, $p = 0.142$, $\eta_p^2 = 0.09$.

Fixed Eye Height. For the *fixed eye height* scaling method, we found a significant main effect of **distance** on head scales for the *minimum* task, $F(2.14, 47.08) = 9.85$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.31$. Post-hoc tests showed that all pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1, 1 and 3, 1 and 10, 3 and 6, 3 and 10, as well as 6 and 10.

We also found a significant main effect of **distance** on head scales for the *maximum* task, $F(1.45, 31.82) = 14.19$, $\boldsymbol{p = 0.038}$, $\eta_p^2 = 0.39$. All pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 3 and 6.

Moreover, we found a significant main effect of **distance** on head scales for the *ideal* task, $F(1.73, 38.09) = 8.25$, $\boldsymbol{p = 0.002}$, $\eta_p^2 = 0.27$. All pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1, 3 and 6, as well as 6 and 10.

4.1.2 Facial Expression Recognition

Fixed Neck Height. For the *fixed neck height* scaling method, we found a significant main effect of **distance** on head scales for the *minimum*, $F(1.26, 27.60) = 115.75$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.84$. Post-hoc tests showed that all pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1.

We also found a significant main effect of **distance** on head scales for the *maximum* task, $F(1.07, 22.55) = 14.94$, $\boldsymbol{p = 0.001}$, $\eta_p^2 = 0.42$. Post-hoc tests showed that all pairwise comparisons were significant ($\boldsymbol{p < 0.05}$).

Moreover, we found a significant main effect of **distance** on head scales for the *ideal* task, $F(1.32, 29.08) = 48.26$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.69$. All pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1.

Fixed Eye Height. For the *fixed eye height* scaling method, we found a significant main effect of **distance** on head scales for the *minimum* task, $F(1.18, 25.98) = 52.46$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.71$. Post-hoc tests showed that all pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1.

We also found a significant main effect of **distance** on head scales for the *maximum* task, $F(1.28, 28.21) = 14.21$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.39$. Post-hoc tests showed that all pairwise comparisons were significant ($\boldsymbol{p < 0.05}$).

Moreover, we found a significant main effect of **distance** on head scales for the *ideal* task, $F(1.01, 22.22) = 4.43$, $\boldsymbol{p = 0.047}$, $\eta_p^2 = 0.17$. All pairwise comparisons were significant ($\boldsymbol{p < 0.05}$), except between distances 0.4 and 1.

4.2 Preference of Scaling Method

We asked participants to indicate on a 7-point Likert scale which of the two scaling methods they preferred (1 = fixed neck height scaling; 7 = fixed eye height scaling). For the task recognizing facial expressions, participants indicated a slight preference of the fixed eye height method ($M = 4.57$, $SD = 2.27$); specifically, eight were in favor of the fixed neck height method, 14 were in favor of the fixed eye height method, and one was undecided. For the task examining the sense of comfort, participants also indicated a slight preference of the fixed eye height method ($M = 4.48$, $SD = 2.45$); specifically, eight were in favor of the fixed neck height method, 13 were in favor of the fixed eye height method, and two were undecided.

5 Discussion

In this section, we summarize the main findings and discuss implications for the use of "big head" avatars in social VR, while also addressing limitations of our experiment. Overall, our results show that participants generally increased the avatar's head scale both for the facial expression recognition and the comfort level task context.

5.1 Avatar Head Scales Increased Over Distance for Recognizable Facial Expressions

For the facial expression recognition task context, in line with our Hypothesis **H1**, we found that participants increased the avatar's head scale when the avatar was moved farther away. As expected, our debriefing confirmed that participants shared a common strategy for the head scaling with respect to the distances—they scaled the head sizes up because they could not make out sufficient details on the avatar's face when they were too distant. In other words, the HMD's limited screen resolution took social cues away for longer distances, while increasing the head scale could bring them back.

Also, based on the *minimum* head scale results shown in Figures 5 (c) and (d), we noticed that there was a clear decision moment when participants changed their scaling direction from decreasing to increasing. Participants were able to see the facial expressions even with a decreased head scale when the avatar was closer than 3 m, while they started to increase the head scale when the avatar was moved beyond 3 m. If mainly caused by the screen resolution of the HMD, this result implies that doubling the screen resolution in the next generation of HMDs could push this distance to 6 m, but in the foreseeable future it will persist as a limiting factor compared to natural human vision and social interaction.

As both the *ideal* and *minimum* head scale results indicate the importance of scaling the head sizes up in order to be able to perceive facial expressions, we propose the values in Figures 5 (c) and (d) as a reference for practitioners in this field. While the values might be shifted a bit due to different screen resolutions and related factors depending on the particular HMD (or other immersive display), we believe that they provide a useful guideline for the effective communication of such social cues in VR.

For completion, we also asked participants to indicate the *maximum* head scale in the experiment. The results are mainly governed by participants enlarging the avatar's head until it filled the HMD's field of view (see Section 3.2). Here, the benefits of the *fixed eye height* head scaling method showed over the *fixed neck height* method, as the head scale could be increased more if the head was located in front of the participant instead of looming over them.

5.2 Avatar Head Scales Increased Over Distance for Comfortable Proportions

Interestingly, in contrast to our Hypothesis **H2**, our results show that participants still increased the avatar's head scale even for the comfort level task, in which they were asked to focus on the avatar's most comfortable head scales and proportions. Given the *ideal* head scale results shown in Figures 5 (a) and (b), participants did not scale up the head as much as they did for the facial expression recognition task, but still increased the head scales quite a bit throughout all the tested distances. In particular, it is interesting that a head scale with a factor of two was perceived as more comfortable than a natural head scale at a distance of 10 m. One potential explanation might be the HMD's limited screen resolution, which, at this long distance, might have affected the participants' spatial perception [4, 31], in particular their ability to estimate the avatar's size or proportions.

However, we were further surprised to see that participants even increased the head scales slightly for very close distances, such as within intimate space or personal space. It appears that participants did not perceive a "big head" as a critical factor that could ruin their sense of comfort with the virtual avatar, even at such close distances.

Our Hypothesis **H2** was based on the consideration of potential *Uncanny Valley* [26] effects caused by an avatar that has human-like body parts but with unrealistic proportions, so we expected that participants would not change the head scale much to keep the avatar's proportions similar to those of a real human. However, our expectation about the *Uncanny Valley* effect in the experiment might have played in a different way, i.e., participants might have perceived the normal-scale avatar as a bit uncanny as it was realistic but "not real enough" to be a real human, so they might have decided to change the head scale to feel more comfortable with a more abstract (cartoonish) character with a big head. It is an interesting vista for practitioners in this field that "big head" avatars (or agents) might actually be perceived as more comfortable than traditional representations.

5.3 Preferences of Scaling Methods for Big Head Avatars

We found a slight preference of the *fixed eye height* scaling method compared to the *fixed neck height* scaling method among our participants, in line with our Hypothesis **H3**, although the difference was less pronounced than we expected. Contrary to our expectations, maintaining the same eye height as the virtual avatar was not the dominating factor for our participants. In the following, we summarize some of the comments that participants made for both scaling methods with their pros and cons, and reasoning for their preference.

The participants who preferred the fixed neck height method stated that it was easy and familiar:

> P3: *"Because there is only one variable changing. With a fixed body only one thing is changing (the head) when there isn't a fixed body, it is very awkward very quickly."*

> P21: *"I like the idea of virtual reality feeling as real as possible opposed to cartoonish."*

> P23: *"I am used to it."*

Whereas, the participants who preferred the fixed eye height method considered its benefits for the convenience of seeing the avatar's face, maintaining eye contact, and as a tool that might be used in an entertainment context:

> P10: *"Since the body shrinks I do not need to move my head to up or down but the head size is big as body is small. but in fact while looking at facial expressions we need to look at the body. I mostly didn't give that much importance to body."*

> P12: *"I felt more comfortable with someone that has natural body features, but the larger head can be more acceptable depending on the situation and if the body's layout is purely for entertainment and not a professional setting."*

> P18: *"I wanted to keep the eye contact because it was more comfortable to me in this kind of situation."*

The comments reinforce the importance of eye-contact and gaze behavior in social contexts [10], but also emphasize the importance of the rest of the body.

5.4 Limitations and Future Work

Our study showed interesting effects of avatar head scale on the perception of facial expression and comfort with respect to the distance considering common social VR settings. However, there are also a few limitations of the current work, which can lead us to additional study ideas we can further investigate in the future.

First, the current study used a single female virtual human character as the visual stimulus to create a consistent VR experience. We understand that different avatar appearances and rendering styles could influence a user's perception of an avatar [23, 35]. The range of a participant's choice of an avatar's head scale at which facial expressions can be observed might differ depending on these factors. Future work may investigate different appearances, such as the avatar's race, skin color, age, and gender, as well as rendering styles, such as photorealistic or abstracted cartoonish characters.

Second, our results are dependent on the current-state HMD (HTC Vive Pro) that we tested in this experiment. We expect that higher-resolution displays will enable users to perceive facial expressions over a longer distance in the future. However, we would like to point out that a threshold for the perception of facial expressions exists even with 20/20 visual acuity, which can be overcome by using the approach presented in this paper.

Third, the avatar in our study setting was static with idle standing animations at different distances. Avatars in social VR would usually not remain at a static distance but move around freely, which may affect the scaling of head sizes and the use of big head avatars. Future work may investigate how dynamic avatar movement and changes in head scales affect a users' perception of social avatars.

6 Conclusion

In this paper, we presented a human-subject study, where we investigated the impact of a virtual avatar's increased or decreased head scale in social VR on a participant's ability to recognize facial expressions and their sense of comfort with the proportions of the avatar. Our results show an important effect of the distance to the virtual avatar on the chosen head scales, indicating that it is imperative to up-scale head sizes over longer distances in order to be able to make out differences in facial expressions on current-state VR HMDs. Our results also showed that even when not concerned with recognizing facial expressions, participants felt more comfortable with head sizes that were up-scaled over their natural size at longer distances. We further compared two alternative head scaling methods and found that maintaining the eye height of the virtual avatar was generally slightly preferred over the other method, where the head was scaled upwards from the neck. We discussed the implications and guidelines for practitioners aiming to leverage "big head" avatars in social VR. For future work, we propose looking into other body parts that could also be important for conveying non-verbal cues in social VR, such as an interlocutor's virtual hands. Moreover, as AR becomes more popular and is more practical as a social platform [17, 20, 33], we believe that social cues could also be similarly enhanced in AR, where such AR-enhanced cues could facilitate more effective collaboration at farther distances [18, 19].

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 Number N00014-17-1-2927 (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] 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.

[2] 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.

[3] J.-F. Bélisle and H. O. Bodur. Avatars as information: Perception of consumers based on their avatars in virtual worlds. *Psychology & Marketing*, 27(8):741–765, 2010.

[4] G. Bruder, A. Pusch, and F. Steinicke. Analyzing effects of geometric rendering parameters on size and distance estimation in on-axis stereographics. In *Proceedings of the ACM Symposium on Applied Perception (SAP)*, pp. 111–118, 2012.

[5] A. Cafaro, H. H. Vilhjálmsson, T. Bickmore, D. Heylen, K. R. Jóhannsdóttir, and G. S. Valgarsson. First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *Proceedings of the International Conference on Intelligent Virtual Agents*, pp. 67–80, 2012.

[6] A. Colburn, M. F. Cohen, and S. Drucker. The role of eye gaze in avatar mediated conversational interfaces. Technical report, Microsoft Research, 2000.

[7] C. M. de Melo, P. Carnevale, and J. Gratch. The effect of virtual agents' emotion displays and appraisals on people's decision making in negotiation. In *Proceedings of the International Conference on Intelligent Virtual Agents*, pp. 53–66, 2012.

[8] C. F. DiSalvo, F. Gemperle, J. Forlizzi, and S. Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the ACM Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, pp. 321–326, 2002.

[9] N. Ducheneaut, M.-H. Wen, N. Yee, and G. Wadley. Body and mind: a study of avatar personalization in three virtual worlds. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1151–1160, 2009.

[10] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 529–536, 2003.

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

[12] J. C. Hager and P. Ekman. Long-distance of transmission of facial affect signals. *Ethology and Sociobiology*, 1(1):77–82, 1979.

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

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

[15] T. Iachini, Y. Coello, F. Frassinetti, and G. Ruggiero. Body space in social interactions: a comparison of reaching and comfort distance in immersive virtual reality. *PloS One*, 9(e111511), 2014.

[16] D. P. Kennedy, J. Gläscher, J. M. Tyszka, and R. Adolphs. Personal space regulation by the human amygdala. *Nat Neurosci*, 12:1226–1227, 2009.

[17] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch. Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 24(11):2947–2962, 2018.

[18] K. Kim, G. Bruder, and G. Welch. Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pp. 1–7, 2017.

[19] K. Kim, D. Maloney, G. Bruder, J. N. Bailenson, and G. F. Welch. The effects of virtual human's spatial and behavioral coherence with

physical objects on social presence in AR. *Computer Animation and Virtual Worlds*, 28(3–4):e1771, 2017.

[20] K. Kim, N. Norouzi, T. Losekamp, G. Bruder, M. Anderson, and G. Welch. Effects of patient care assistant embodiment and computer mediation on user experience. In *Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 17–177, 2019.

[21] K. Kiyokawa, H. Takemura, and N. Yokoya. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 48–53, 1999.

[22] E. Langbehn, G. Bruder, and F. Steinicke. Scale matters! Analysis of dominant scale estimation in the presence of conflicting cues in multi-scale collaborative virtual environments. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 211–220, 2016.

[23] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pp. 39:1–10, 2017.

[24] M. Lee, G. Bruder, and G. F. Welch. Exploring the effect of vibrotactile feedback through the floor on social presence in an immersive virtual environment. *Proceedings of IEEE Virtual Reality (VR)*, pp. 105–111, 2017.

[25] M. Lee, N. Norouzi, G. Bruder, P. Wisniewski, and G. Welch. Mixed reality tabletop gameplay: Social interaction with a virtual human capable of physical influence. *IEEE Transactions on Visualization and Computer Graphics*, 24(8):1–12, 2019.

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

[27] N. Norouzi, G. Bruder, B. Belna, S. Mutter, D. Turgut, and G. Welch. *Transactions on Computational Science and Computational Intelligence: Artificial Intelligence in IoT*, chap. A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things, pp. 1–24. Springer, Cham, 2019.

[28] N. Norouzi, A. Erickson, K. Kim, R. Schubert, J. LaViola, G. Bruder, and G. Welch. Effects of shared gaze parameters on visual target identification task performance in augmented reality. In *ACM Symposium on Spatial User Interaction (SUI)*, pp. 1–11, 2019.

[29] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G. Lee, and M. Billinghurst. Exploring enhancements for remote mixed reality collaboration. In *Proceedings of ACM SIGGRAPH Asia Mobile Graphics & Interactive Applications*, p. 16, 2017.

[30] T. Piumsomboon, Y. Lee, G. A. Lee, A. Dey, and M. Billinghurst. Empathic mixed reality: Sharing what you feel and interacting with what you see. In *Proceedings of the IEEE International Symposium on Ubiquitous Virtual Reality*, pp. 38–41, 2017.

[31] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments - a review. *ACM Comput. Surv.*, 46(2):23:1–23:40, 2013.

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

[33] G. Welch, G. Bruder, P. Squire, and R. Schubert. Anticipating Widespread Augmented Reality: Insights from the 2018 AR Visioning Workshop. Technical report, University of Central Florida and Office of Naval Research, August 2019.

[34] Will Cherry. Oculus Connect 6, All In This Together - Live Blog. `https://noproscenium.com/oculus-connect-6-all-in-this-together-1ad86529e62b`, Sep 25, 2019.

[35] E. Zell, C. Aliaga, A. Jarabo, K. Zibrek, D. Gutierrez, R. McDonnell, and M. Botsch. To stylize or not to stylize? the effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)*, 34(6):1–12, 2015.

[36] X. Zhang and G. W. Furnas. MCVEs: Using cross-scale collaboration to support user interaction with multiscale structures. *Presence*, 14(1):31–46, 2005.