The Effects of Virtual Avatar Visibility on Pointing Interpretation by Observers in 3D Environments

Brett Benda* Eric D. Ragan†

Department of Computer & Information Science & Engineering, University of Florida

Abstract

Avatars are often used to provide representations of users in 3D environments, such as desktop games or VR applications. While full-body avatars are often sought to be used in applications, low visibility avatars (i.e., head and hands) are often used in a variety of contexts, either as intentional design choices, for simplicity in contexts where full-body avatars are not needed, or due to external limitations. Avatar style can also vary from more simplistic and abstract to highly realistic depending on application context and user choices. We present the results of two desktop experiments that examine avatar visibility, style, and observer view on accuracy in a pointing interpretation task. Significant effects of visibility were found, with effects varying between horizontal and vertical components of error, and error amounts not always worsening as a result of lowering visibility. Error due to avatar visibility was much smaller than error resulting from avatar style or observer view. Our findings suggest that humans are reasonably able to understand pointing gestures with a limited observable body.

Index Terms: Human-centered computing—Human computer interaction—Interaction techniques—Pointing; Human-centered computing—Human computer interaction—Interaction paradigms—Collaborative interaction

1 Introduction

In 3D environments, it can be beneficial to provide users with a virtual body (an *avatar*) to represent themselves in a virtual space. Motion-tracked avatars also facilitate natural motion/gestural cues for social communication (e.g., waving to communicate "Hello", or pointing in mid-air to indicate a location or direction). These gestures are particularly useful in *collaborative virtual environments* (CVEs) where multiple users are co-located in the same environment by providing a means of nonverbal communication.

Avatars can be displayed in varying degrees of fidelity, with aspects like style varying based on application context (e.g., cartoony in a game compared to realistic in a workplace) or user choice. Users often customize avatars to look like themselves or someone completely different since the real world limitations do not apply. The visual design of avatars can also be influenced by the technology being used (i.e., the Oculus Quest only natively supports head/hand tracking). A minimal approach to avatar design is often taken by only including visual representations tracked body parts or to hide parts of an avatar that cannot be reasonably approximated. This lowers the avatar's *visibility*, or how much of an avatar's body is displayed in an environment. In fact, using low visibility avatars is what many games [27] and CVEs [46] currently do. While full-body avatars may be considered ideal, they may not always be able to be used. Other aspects of avatar design, such as avatar style [37] and viewing angle/distance [44] have been shown to impact first-person pointing behavior as well as the accuracy of pointing interpretation by others. These are important to consider in conjunction with visibility as these often vary based on under preference and application context. Less is known about how the visibility of a body represented in the virtual environment influences interpretation of pointing. Pointing, as a communication cue that humans commonly use in the real world, is often seen as more natural than augmented interaction techniques such as 3D pointers [47] or raycasting from a controller [42]. Because avatar display can vary, it is important to examine the interpretability and comprehension of gestural pointing in virtual environments based on an avatar's body visibility.

Prior work has shown that low visibility avatars function similarly to full-body avatars in an egocentric perspective, but in collaborative settings it is unknown if these cause observers to become inaccurate in their interpretations. It is possible that pointing interpretation may worsen with low visibility avatars due to less positional information being communicated (e.g., a lack of elbows in a hand and head only avatar may remove crucial reference points for understanding the pointing gesture).

Thusly, our research question is as follows: *How does decreasing avatar visibility affect human ability to understand collaborator pointing location?* We address this question by conducting two desktop experiments; the first utilizing a static viewing perspective and the second allowing for view changing. Due to restrictions in place because of COVID-19, these were conducted as desktop studies.

2 Related Work

We base our experiment on prior work related to general benefits of avatar inclusion in virtual environments, motivations and use cases for different visual styles of avatars, psychology of human pointing, and how avatar design impacts pointing behavior and interpretation.

2.1 Avatar Use in Virtual Environments

Avatars provide a wide range of benefits for individual users, even outside of collaborative situations. One such benefit is that self-avatars allow users to feel *embodied* in a scene by adding a visible representation of themselves in a scene [18, 28]. Embodiment is related to the *illusion of virtual body ownership* (IVBO), or the sensation that virtual body parts are actually part of oneself [40]. Embodiment and IVBO can be achieved in both third person and first person views [8] to varying degrees. These phenomena can be increased by a combination of high fidelity displays, mappings between user input and the display/environment, and support for expected interactions [17]. High amounts of embodiment and presence induced by avatar use and design can affect how a user feels or acts in virtual spaces, causing them to act older [36] or experience less bias based on race [32] or gender [22]. It is also not imperative for avatars to be humans [24] and customization options based on user wishes can also benefit a sense of embodiment [9]. Feeling embodied not only influences how a user sees themselves, but also how they act. For example, Ogawa et al. examined how avatar anthropomorphism and visibility affect tendencies to walk through virtual walls and other geometry [30]. Despite knowing that walls were not solid, full-body, realistic avatars were found to induce a

*e-mail: brett.benda@ufl.edu

†e-mail: eragan@ufl.edu

higher sense of presence and reduce rates at which users walked through walls.

While embodiment is often desired, sometimes avatars are designed with different goals in mind. For example, designers frequently want to increase the capabilities of users beyond what is possible in the real world. Less visible avatars are used because of the wider range of interactions they provide. Consider cases where users may want to grab an object in the scene but it is out of reach. By using a lower fidelity avatar with a disconnected hand, techniques such as Go-Go [35] or linearly scaling hand movement [10], can increase reach allowing users to grab distance objects. Techniques for hand redirection (e.g. [1, 16, 48]) also often use a disconnected hand since the displayed hand location differs from those in the real world. "Stretching" the arm of a full-body avatar can be used to maintain connectedness and a sense of embodiment using similar scaling techniques [11]. However, if these techniques are used in collaborative environments, it is possible that only hands may be displayed to preserve fidelity in a first person perspective without considering how this could affect the perception of onlookers.

Additionally, it is common practice to exclude full bodies in situations where users will not be able to observe themselves to save development costs; if a body (or any part of an object in an environment) is not able to be seen, it is not necessary to model and rig. Other practical factors such as tracking limitations may also influence what types of avatars are suitable for an application. For example, the Oculus Rift S[1] and Oculus Quest 2[2] are only able to natively tracks a head location and hands. This makes display of full-body and full-arm avatars unreliable. While ideal systems should be able to support reliable full body tracking, most consumer systems are not at this point. Inverse-kinematics can be applied in cases where full-body tracking is not possible, but this is not always accurate to the real body and may not be suitable for all use cases (i.e., exergames) [5].

One may expect embodiment to decrease if a lower visibility avatar is used, but prior work has found that lower visibility avatars can produce a similar sensation of ownership and perform similarly on a wide range of tasks. In comparing hand only, hand with forearm, and full arm, Tran et al. [41] found that each of these visibilities produce similar ownership ratings, but that the full arm was slower to reach targets. While embodiment may not be a design goal of low visibility avatars, it can still be felt with as little as a tracked hand [20] or with non-optimal avatar configurations [12]. Additionally, Pan et al. [31] did not find differences between hand-only and full-arm avatars on cognitive load while using each avatar, indicating that their use does not add unnecessary complexity.

Thus far, we have only discussed self-avatars and how they impact a solo experience, but avatars also can affect other users in shared environments. Social presence ("the sense of being with another [person]" [3]) often relies on visual qualities of co-located parties and communication mediums between them. A key issue with avatar design for collaborative environments is the *uncanny valley* [4, 39], a level of visual and kinematic fidelity that is "not quite real, but not quite stylized" which elicits negative emotions. Thankfully, the "uncanniness" of avatars can be minimized by properly accounting for known traits that induce it [38]. With collaborative avatars, the style and visibility of a partner's avatar have been demonstrated to influence social presence [46], with whole-body realistic avatars increasing social presence and upper-body and cartoony avatars acting as fair substitutes if needed by a collaborative context. Even avatars that are entirely different in scale from collaborators in the real world can elicit social presence. Wang et al. examined how avatar body representation level (analogous to visibility) affected usability of AR instructors [43]. In their task, avatars pointed to various objects in close proximity and users selected what was pointed to. Full body avatars were found to decrease workload and increase efficiency. Near and far views were also examined, with far views decreasing accuracy of object selection. Miniature collaborators can be leveraged to keep collaborators inside each other's field of view, but still exhibited comparable levels of social presence. [33]. Assessing social presence is not a goal of this work, but it is important to understand that including avatars in a social environment tends to increase the sensation.

[1]https://www.oculus.com/rift-s/

[2]https://www.oculus.com/quest-2/

2.2 Pointing in Virtual Environments

Of most importance to this work is how avatars facilitate gestural communication. Gestures are among the most basic means of communication, outside spoken and written communication, and allows humans to understand each other through a visual medium that carries more nuance and information than may be carried through speech alone [15]. Speech and gesture are highly related, with the two combined acting often as a single unit of communication with more impact than either alone [2]. Our primary interest is pointing, a subset of gestural communication that projects a vector from one's body that indicates a specific object, location, or direction [19]. In this section, we discuss computational models of how pointing is interpreted and how avatar design has already been shown to impact interpretation.

Prior work has found pointing to be an effective means of communication in virtual reality, despite the varying behaviors displayed by users in *how* they make their pointing gestures [21]. Different techniques of computationally predicting pointing locations have been explored, each relying on different body part locations. Perhaps the most intuitive approach is to virtually "extend" the users index finger and use the location where it first intersects an object in the scene as the target location. This is known as *index-finger ray cast* (IRFC) [7]. Other models incorporating additional joints and body parts have also been explored, including *forearm ray cast* (FARC) [29] and *eye-finger ray cast* (EFRC) [26]. Additionally, non-gestural pointing and selection methods exist when a high degree of accuracy is needed, such as 3D pointers [47] or raycasting from a controller [42].

Many aspects of avatar design and their effects on pointing interpretation in CVEs have been examined previously. Wong and Gutwin [44] compared real-world pointing and pointing in virtual environments. By varying user location and target distance, they found higher accuracy in the real world both for interpreting pointing and producing pointing gestures. However, they identified that a higher degree of accuracy may not be needed given varying task domains. Specific to desktop applications, Wong and Gutwin also examined the use of different input devices to control pointing in desktop applications in order to provide users without tracking systems the ability to point without sophisticated tracking [45]. However, a limitation of both these works is the quality of avatar examined; both utilized an avatar consisting of simple primitive geometry with a single unarticulated finger.

Later work looked more rigorously at how visual properties of avatars impacted pointing, both in egocentric and observational contexts. An analysis of egocentric pointing with six avatar styles under three pointing models (IFRC, FARC, EFRC) performed found that EFRC produced significantly less error than other models [37]. Corrective models can further reduce error, and it has also been demonstrated that accuracy does not correlate with perceived human-likeness. In social and collaborative settings, corrections to a pointer's pose have also been applied to reduce interpretation error [25]. While these correction models are outside the scope of this work, they are worth noting due to their ability to decrease pointing error without direct user awareness while still using gestural pointing.

Alternative forms of rendering an avatar were examined by Gamelin et al. [13]. In two collaborative tasks, point cloud avatars

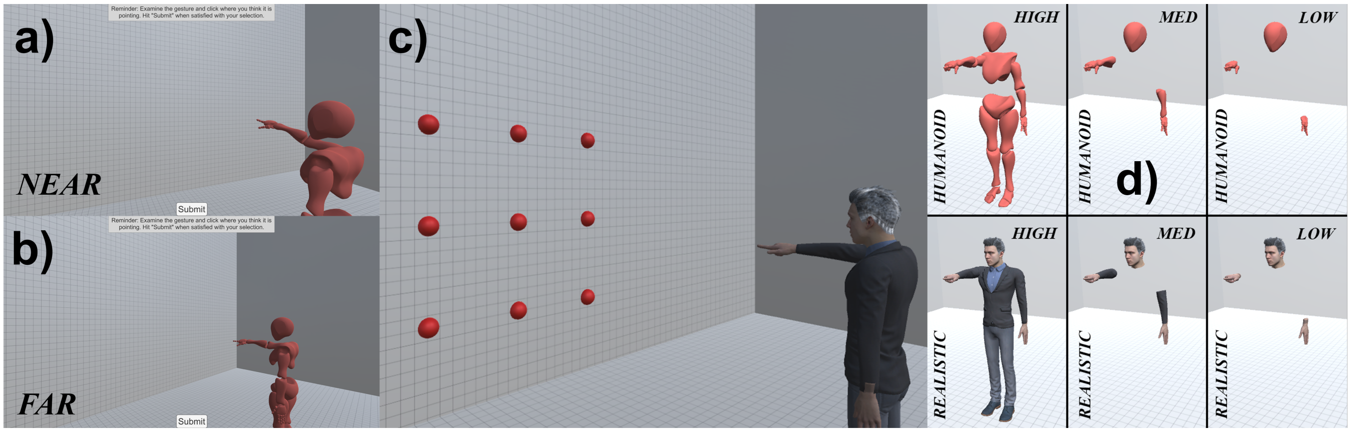

Figure 1: **a)** The *NEAR* view. **b)** The *FAR* view. **c)** A cropped view showing the nine possible targets in red (not visible to participants). **d)** The six avatars examined in the experiment. Each row represents one of the two styles. Each column shows avatars of the same visibility.

were compared against preconstructed solid avatars and were found to decrease error rates and time to complete the tasks. This was thought to be due to an increased *kinematic* fidelity of the point cloud avatar, since it could capture user movements more accurately and with more nuance compared to the premade joints used for the solid avatar. Accurately displaying motion was more important than including a visually detailed avatar.

2.3 Summary

Overall, application goals and practical factors play a role in what avatars are used. Avatars are important for providing users a sense of self in a scene and a sense of colocation with others in a shared space. While full body avatars are often considered the gold-standard for how a user's body should be displayed, lower visibility avatars are also used considerably due to their interaction flexibility, low implementation cost, and ability to maintain performance and a sense of embodiment. Gestures are also an important part of human communication, and while several avatar design factors have been investigated for effects on generation and interpretation of pointing gestures, *visibility* has not been investigated in this context.

3 Research Overview

We conducted two experiments to determine effects of avatar body visibility on the ability of collaborators to understand pointing gestures. Both experiments utilize a similar design, with view rotation and controls varying between the two. This section provides an overview of details shared between both experiments. We then cover unique design features for each experiment and results. Finally, we discuss them together and key observations from them both.

Participants take on the role of a collaborator in a desktop environment (the *observer*) interacting with a motion-tracking-enabled collaborator (the *pointer*) using pointing gestures. They are asked to visually examine different pointing gestures and then select the location in the scene the pointer is gesturing towards. Avatar visibility, style, and view are varied between instances of the task. We examine error rates and how they differ based on visibility, style, and view.

3.1 Goals and Experimental Hypotheses

Our primary goal of these experiments is to determine effects of avatar visibility on pointing interpretability by collaborators. Our hypotheses are as follows:

- **H1**: Observer error rates increase as avatar visibility decreases.
- **H2**: Observer error rates decrease as avatar style becomes more realistic.
- **H3**: Observer error rates increase as viewing distance increases.

3.2 Independent Variables

These experiments are conducted using a repeated-measures design. We have three independent variables, two of which regard the visual properties of the avatar (seen in Figure 1d) and one focused on positioning relative to the pointer (seen in Figure 1a and Figure 1b):

- Visibility: *HIGH* (full body), *MED* (head and hands with forearm), *LOW* (head and hands)
- Style: *REALISTIC* (human), *HUMANOID* (human in figure, but lacking detail)
- View: *NEAR* (observer is close in proximity to the pointer, view angles similar), *FAR* (observer is further from the pointer, view angles are not aligned)

3.3 Measures

We are interested in examining the effects of avatars on pointing interpretation error. Three error measures are examined:

- Total Error: the euclidean straight-line distance between a selected location and the target location
- Horizontal Error: the horizontal component of the Total Error
- Vertical Error: the vertical component of the Total Error

These provide insight into overall effects on error and directional effects.

3.4 Avatar Design

This section outlines implementation details for achieving each level of avatar visibility and style.

The avatar models and skeletons utilized are modified versions of Mixamo's[3] "Josh" and "X Bot" characters. "Josh" (a clothed male human with facial features) was used as the base for REALISTIC avatars, while "X Bot" (an abstract robotic female) was the base model for HUMANOID avatars. These were chosen due to their shared human figures but different visual qualities. "Josh" closely matches a real human, with clothes and hair and realistic body proportions that may be seen in professional collaborative settings. On the other hand, "X Bot" is a more stylized avatar with robotic joints, lack of face, and consistent texturing which you may see in social or gaming contexts.

[3]Mixamo, https://mixamo.com

Autodesk Maya[4] was used to edit each model into different levels of visibility. *HIGH* visibility avatars were left unmodified since they are already full-body avatars. *MED* visibility avatars had all geometry below the lowest neck bone removed, but keeping arm geometry below the elbow bone to display large part of the arm pose. Similar levels of visibility have been considered in first person contexts as a way to maintain body ownership without needed a full model for avatars (i.e., [14, 23, 41]). *LOW* visibility avatars were the same as *MED*, but with the arm geometry limited to geometry beyond the wrist bone. The resulting holes were filed and retextured to restore closed geometry. Final avatars used in the experiments can be seen in Figure 1d.

3.5 Virtual Environment

This section discusses the graphics, layout, and view perspectives included in the virtual environment.

A grid texture was applied to the floor and wall being pointed towards. This was to provide a sense of scale and depth that is missing due to simple geometry and textures being used in the rest of the environment. The scene was lit from above to provide shaded textures for the avatar, but no shadows were cast to allow users to focus purely on the visual properties of the avatar.

In the environment, one unit was approximately one meter of equivalent real-world distance. However, since screen size (and thus perceived distances) varied between participants, we report distances using the generic relative unit.

A set of targets located on a wall across the pointer and observer were examined for each treatment to provide variance in avatar pose. The wall was placed 3 units in front of the pointer. Nine targets were placed on the wall, in a 3-by-3 square grid, spanning 2 units by 2 units (see Figure 1c). The center target was located directly in front of the pointer's chest level, 1.5 units off the ground.

The pointer in this experiment was simulated and utilizes premade poses pointing towards the targets. The poses utilized a rigid right arm pointing with an extended index finger. The pointer's head was constrained to look just above their index finger which simulates pointing with ERFC. This was chosen based on prior work which has found errors under this model being less than those under IRFC or FARC [37]. To generate the pose for each target, experimenters rotated the pointing right shoulder until the view from the avatar's perspective was pointing to the target. The same pose was then applied to each avatar for all targets for consistency.

In order to asses how positioning relative to the pointer affects interpretation, we examine two viewing perspectives. Prior work (e.g., [45]) has already identified that distance and angle play a key role in interpreted pointing gestures. Given that co-located persons in a scene may not always be in close proximity, we examined two different viewing positions to provide insight into the role of distance on error. Generally, the *FAR* view had a smaller relative avatar size and less-aligned perspective than the *NEAR* view.

In the *FAR* perspective, observers examined a pointing gesture from a distance while seeing most of an avatar's body: they were positioned 2.25 units to the left and 0.75 units behind the pointing avatar viewing initially at a 45 degree angle. In the *NEAR* perspective, observers were closer to the pointer and a more aligned perspective; they were positioned 1.25 units to the left and 0.75 units behind the pointing avatar viewing initially at a 22.5 degree angle. Views from the application can be seen in Figure 1a and Figure 1b. In both cases the observer's and pointer's eye level were 1.75 units off the ground and have a 60 degree field of view.

3.6 Apparatus

Due to limitations of COVID-19, we conducted this study remotely using a desktop browser application. The study applications were developed using the Unity game engine (version 2019.4.21f1). It was deployed as a WebGL build and hosted on a server belonging to our institution. Users interacted with the application on their personal computers, with screen resolutions ranging from 960x600 to 3840x2160 Web browser choice was not restricted as WebGL functions consistently across all common browsers.

Qualtrics was utilized to administer a demographics questionnaire and collect additional information about the participant's experience in the study.

3.7 Procedure and Task

Participants were directed to the study website where informed consent is obtained and were then instructed that they would observe an avatar with different visual properties from different perspectives and make judgements about where it is pointing and to be both accurate and timely in their responses, but to focus on accuracy. They were asked to complete the study in fullscreen on a desktop or laptop computer using a mouse to select locations on the scene.

The participant was allotted as much time as they wish to observe the gesture but were instructed to provide answers quickly and accurately. They were instructed how to select a location in the scene where they believed the avatar to be pointing (see Section 4.1 and Section 5.1). An indicator was shown so they can confirm the location and see it in context of the pointing gesture. They could change their answer as many times as desired, with the indicator changing as well. When satisfied, they submitted their answer (see Section 4.1 and Section 5.1). After receiving instructions, they moved on to the application in a separate web page and begin completing the task automatically.

An list of every target/style/visibility/perspective combination was generated and randomized per participant. The first trial served as an introduction to the task and input methods for the experiment. The condition was repeated as the final trial, and the first instance was excluded from analysis. A brief menu pausing the study appeared between each trial, both to provide opportunities for breaks and to hide the changing view and avatar. Because 9 poses were observed under each treatment (3 visibilities, 2 styles, 2 views), a total of 108 data points were collected for each treatment.

After observing all avatars, they were directed to a questionnaire about demographic information, information about their experience with games and 3D environments, and open-ended questions about their experience completing the study.

3.8 Analysis Methodology

This methodology was used for each experiment.

At the trial level, error values outside one and a half of the interquartile range were removed from analysis as outliers. Values across targets were then averaged across each treatment leaving one average value for each combination of visibility, style, and view for each participant.

The remaining data met sphericity and normality assumptions for all error measures so we employed a three-way repeated-measures ANOVA to find significant effects of visibility, style, and viewing perspective on each error rate. Due to repeated testing, we applied Bonferroni correction (n=3) to limit false discoveries. Pairwise t-tests with Bonferonni correction were used to examine differences between treatments. We also present η^2 effect sizes [6] for each factor to provide insight into how impactful each factor was on the measures.

3.8.1 Interpretation of Error

Reported distances should be considered as *relative* units rather than in real world units. We choose to report distance in this manner due to the online format of our experiment. Participant screen resolution was not controlled, so there is no guarantee all participants perceived distance identically. However, we conducted a Pearson's correlation

[4]Maya, https://www.autodesk.com/products/maya/overview

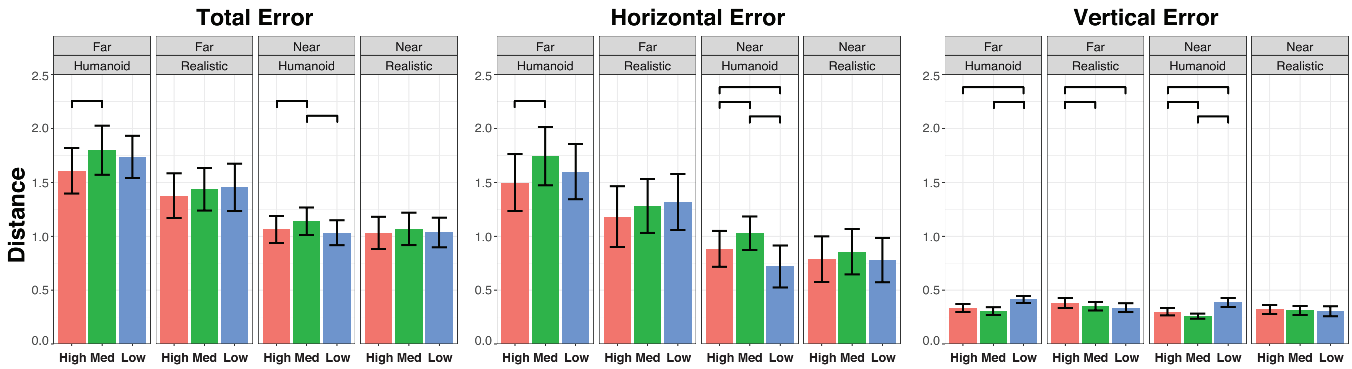

Figure 2: Results from **Experiment 1**. Plotted are average values for each measure under each treatment. Error bars are 95% confidence intervals. Lines between bars signify significant differences at $p < 0.05$ (only presented between visibilities in the same view and style treatments). Positive horizontal error represents rightward error (i.e., the user answer was more right than the intended location). Positive vertical error represents upward error (i.e., the user answer was higher than the intended location).

test to verify that resolution did not impact interpretation and found no significant correlation between resolution and errors for either experiment.

4 Experiment 1: Interpretation with a Fixed Perspective

Our first experiment sought to examine a simple case of pointing interpretation, with perspective being controlled by locking the camera to a single view for each trial. When each view is applied, the user is unable to modify the view and must make judgements about the pointing gesture from this static view. This is to provide a focus on the visual properties of the avatar and limit other factors from influencing the participant's response.

4.1 Interaction Methods

The mouse cursor is used entirely to select a location and submit a response. When the participant is ready to provide an answer, they click to select where in the environment they interpret the pointing location to be. A ray is cast from this location, and an indicator placed on where it intersects with the wall. When satisfied, they click a submit button at the bottom of the screen to move to the next avatar/pose from the randomized list.

4.2 Results

We collected data from 27 participants (22 males), ranging from ages 18-23 (median 21) from undergraduate computer science courses at our institution.

We examined total error as well as horizontal and vertical components of error. Participants averaged 14 minutes to complete the entire study, averaging 7.78 seconds per trial. This is appropriate given the instructions; users were told to make quick judgements about the gesture.

4.2.1 Total Error

First, we examined total distance error between the interpreted location and the target location. 1.9% of all trials were removed during outlier removal. All factors had significant main effects:

- Visibility: $F(2,52) = 6.59$ and $p < 0.01$, $\eta^2 = 0.005$
- Style: $F(1,26) = 29.64$ and $p < 0.001$, $\eta^2 = 0.022$
- View: $F(1,26) = 98.99$ and $p < 0.001$, $\eta^2 = 0.175$

One two-way interaction was significant, Style-View: $F(1,26) = 66.88$ and $p < 0.001$, $\eta^2 = 0.014$. No significant three-way interaction effect between our factors. Averages for each treatment as well as significant differences are presented in Figure 2.

While visibility was a significant factor, its very small effect size suggests it was not a very large contributor to error. This is reflected in Figure 2, where there are some differences between visibility levels in the same view and style treatment, but the differences are generally small (at most 0.19 units or 10.0%). In cases where there were significant differences, error did not always increase according to visibility; *LOW* visibility avatars sometimes yielded lower error than *MED* visibility avatars.

In the *FAR* view, significant differences based on style were seen between *HUMANOID* and *REALISTIC* avatars, with *REALISTIC* avatars yielding lower error. However, in the *NEAR* view errors between the two were not significantly different. Being in close proximity to avatars allowed participants better interpret the gesture regardless of the style. This explains the significant two-way interaction between style and View, but the effect of the interaction was very small.

Overall, total distance was minimally impacted by visibility. View was the dominant factor impacting total error with a large effect size (0.175). This agrees with prior work [44] which found that distance relative to the pointer hinders pointing interpretation as the distance increases.

4.2.2 Horizontal Error

For horizontal error, 0.9% of all trials were removed during outlier removal. All factors also had significant main effects:

- Visibility: $F(2,52) = 11.09$ and $p < 0.001$, $\eta^2 = 0.008$
- Style: $F(1,26) = 31.78$ and $p < 0.001$, $\eta^2 = 0.021$
- View: $F(1,26) = 130.88$ and $p < 0.001$, $\eta^2 = 0.147$

All two-way interactions were significant:

- Visibility-Style: $F(2,52) = 9.74$ and $p < 0.001$, $\eta^2 = 0.003$
- Visibility-View: $F(2,52) = 7.75$ and $p < 0.01$, $\eta^2 = 0.003$
- View-Style: $F(1,26) = 61.39$ and $p < 0.001$, $\eta^2 = 0.01$

ANOVA did not identify a significant three-way interaction effect.

Visibility observations are similar to those of total error, with visibility having a small effect size and significant differences not always increasing as visibility decreases. We found view to be the significant factor with the largest effect on horizontal error. Observations relating to style also carry over, with the two-way interaction between view and style leading to larger differences in the *FAR* view.

The remaining two-way interactions are difficult to explain, as their effect size is very small. Generally, the *MED* visibility avatar yielded larger error only in the *HUMANOID* avatar. *LOW* visibility avatars also had lower error compared to *HIGH* visibility in the

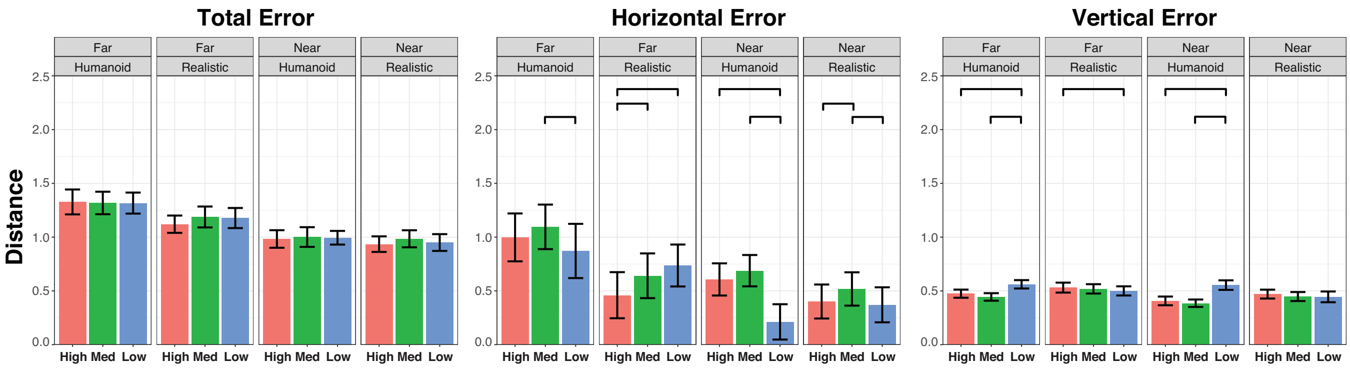

Figure 3: Results from **Experiment 2**. Error bars are 95% confidence interval Lines between bars signify significant differences at $p < 0.05$ (only presented between visibilities in the same view and style treatments). Positive horizontal error represents rightward error (i.e., the user answer was more right than the intended location). Positive vertical error represents upward error (i.e., the user answer was higher than the intended location).

NEAR view. However, these effects were minimal according to the small effect size ($\eta^2 = 0.003$ for both).

4.2.3 Vertical Error

For vertical error, 1.2% of all trials were removed during outlier removal. Two factors had significant main effects:

- Visibility: $F(2,52) = 24.21$ and $p < 0.001$, $\eta^2 = 0.032$
- View: $F(1,26) = 21.23$ and $p < 0.001$, $\eta^2 = 0.026$

One two-way interaction was found, Style-Visibility: $F(2,52) = 75.15$ and $p < 0.001$, $\eta^2 = 0.054$. Again, ANOVA did not find a significant three-way interaction effect between the factors.

View continued to be a significant factor on vertical error, with the *FAR* view producing higher error than *NEAR*. However, the effect was much smaller than those seen in total or horizontal error as seen by its small effect size.

The two-way interaction between style and visibility manifests as *LOW* visibility avatars producing higher error, but only with a *HUMANOID* style. For these avatars, *MED* visibility avatars also had lower error than *HIGH* visibility in the *NEAR* view. visibility also had an effect of producing larger errors with *HIGH* visibility avatars, but only for the *FAR REALISTIC* avatar. No differences were found between visibilities for the *NEAR REALISTIC* avatar.

4.2.4 Overview

In summary, view was the largest influencer on error, with style coming next, and visibility being the smallest contributor. Visibility effects were inconsistent; *LOW* sometimes decreased horizontal error, but also increased vertical error. Horizontal error also comprised the largest amount of total error. However, these differences due to visibility were fairly small and would likely not be impactful in cases where low-to-medium precision is needed.

5 Experiment 2: Interpretation with View Rotation

Our second experiment loosened the perspective constraint of the first. At the beginning of each trial the designated view is applied, but participants can rotate their view while staying in the same location. This change is motivated by real use cases for virtual avatars. In 3D environments, perspectives can almost always be manipulated which benefits spatial awareness.

5.1 Interaction Methods

Horizontal and vertical mouse movements rotated the observer view in the scene while leaving their position unchanged. The rotation was limited in the y-axis to ±60 degrees up and down in order to maintain proper orientation. Horizontal locks were also applied to keep participants looking in the general direction of the task in the scene. These limits were generous enough to allow participants to view the entire area of interest in the scene and were solely meant to prevent orientation issues.

Because the mouse was used to rotate the camera, the selection method was modified from Experiment 1. A crosshair was added in the middle of the camera, and clicking now selected the location pointed to by the cross-hair rather than mouse cursor. An indicator was shown at the location so participants can verify the location and see it in context of the pointing gesture. When satisfied, they pressed the space bar to move to the next avatar/pose from the randomized list.

5.2 Results

We collected data from 36 participants (28 males), ranging from ages 18-30 (median 21) from undergraduate computer science courses at our institution. These participants were unique from those participating in Experiment 1.

As in Experiment 1, we examined total error as well as horizontal and vertical components of error. Participants averaged 14 minutes and 35 seconds to complete the entire study, averaging 8.10 seconds per trial. As was the case with the first experiment, this is appropriate given the instructions; users were told to make quick judgements about the gesture.

5.2.1 Total Error

First, we examined total distance error between the interpreted location and the target location. 3.1% of all trials were removed during outlier removal. Only two factors had significant main effects:

- Style: $F(1,35) = 12.33$ and $p < 0.01$, $\eta^2 = 0.023$
- View: $F(1,35) = 80.73$ and $p < 0.001$, $\eta^2 = 0.147$

One two-way interaction was significant, Style-View: $F(1,35) = 19.18$ and $p < 0.001$, $\eta^2 = 0.009$. No significant three-way interaction effect between our factors. Averages for each treatment as well as significant differences are presented in Figure 3.

Notably, only view and style significantly impacted error. The *FAR* view increased error, with *HUMANOID* avatars yielding a larger increase. However, no differences were found between either style in a *NEAR* View. No significant effects due to visibility were observed.

5.2.2 Horizontal Error

For horizontal error, 0.7% of all trials were removed during outlier removal. All factors also had significant main effects:

- Visibility: $F(2,70) = 16.00$ and $p < 0.001$, $\eta^2 = 0.013$
- Style: $F(1,35) = 21.17$ and $p < 0.001$, $\eta^2 = 0.026$

- View: $F(1,35) = 44.58$ and $p < 0.001$, $\eta^2 = 0.057$

All two-way interactions were significant:

- Visibility-Style: $F(2,70) = 19.97$ and $p < 0.001$, $\eta^2 = 0.015$
- Visibility-View: $F(2,70) = 28.51$ and $p < 0.001$, $\eta^2 = 0.009$
- View-Style: $F(1,35) = 39.01$ and $p < 0.001$, $\eta^2 = 0.012$

ANOVA did not identify a significant three-way interaction effect.

View again had a significant main effect, with *FAR* viewing leading to increased error. Style differences were also seen, with *HUMANOID* avatars having increased error compared to *REALISTIC* avatars. Visibility also had significant main effects, but observed differences were not correlated with the level of visibility. For example, in both *FAR HUMANOID* and *NEAR HUMANOID* treatments *LOW* visibility avatars produced less error than *MED* avatars. Interestingly, for *NEAR HUMANOID* avatars error was *lowest* with a *LOW* visibility avatar.

5.2.3 Vertical Error

For vertical error, 3.0% of all trials were removed during outlier removal. Two factors had significant main effects:

- Visibility: $F(2,70) = 31.02$ and $p < 0.001$, $\eta^2 = 0.033$
- View: $F(1,35) = 74.64$ and $p < 0.001$, $\eta^2 = 0.031$

One two-way interaction was found, Style-Visibility: $F(2,70) = 92.92$ and $p < 0.001$, $\eta^2 = 0.053$. Again, ANOVA did not find a significant three-way interaction effect between the factors.

As seen with the prior two measures, *FAR* viewing produced higher error than *NEAR*. In *HUMANOID* avatars, *LOW* visibility avatars had significantly higher error than both *MED* and *HIGH* in both views, but for *FAR REALISTIC* avatars *LOW* had lower error than *HIGH*. Again, effects of visibility are inconsistent, but fairly small in magnitude.

5.2.4 Overview

The results of this experiment largely yielded the same results as Experiment 1. View and style lead to the largest differences in error, while visibility yielded very small differences. Effects of visibility were also inconsistent, sometimes producing lower error as it decreases while other times producing higher.

6 Discussion

We conducted two experiments to examine effects of view, style, and visibility on pointing interpretation error. In the first experiment ($n = 27$), subjects made judgements about static pointing gestures from a fixed observer perspective. In the second experiment ($n = 36$), subjects made judgements about static pointing gestures while able to rotate their view. Both experiments were able to find significant effects of view, style, and visibility on pointing interpretation error. We discuss implications of our findings and provide suggestions to VR designers on how avatar visibility can be used in applications.

6.1 Key Observations

Between both experiments, findings were similar for each factor. As expected, observer viewing perspective was a large contributor to error. This makes sense, as viewing from a further distance decreases the relative avatar size and increases uncertainty about its pose.

Style also produced large differences, with *REALISTIC* avatars generally outperforming *HUMANOID* avatars. During our post-questionnaire participants were asked to identify features of the study that impacted their ability to accurately judge the pointing location. Many participants acknowledged that the *REALISTIC* avatar's face was a helpful factor in interpreting the pointing despite lacking features such as eye movement or animated expressions which could benefit their interpretation. The face and eyes (as leveraged in EFRC [26]) provide information about the intended target location since pointers tend to look at their pointing location. This explains the large effect of style in the *FAR* view (where the face was more visible to users) compared to *NEAR* and agrees with Wang et al. [43]. An example of this interaction can be seen in Figure 5.

While significant effects of visibility were observed, they were not linear in their effect (e.g., *LOW* did not always produce higher error than *MED* and *HIGH*). We also did not find consistent evidence that visibility effects are always observed; *REALISTIC* avatars only saw an effect of visibility on vertical error in the *FAR* view in both experiments. However, this does follow from other work (e.g., [37,41]) which found that visual realness did not always correlate with pointing accuracy, but we demonstrate that this is also seen in observer interpretation. Together, these observations do not support our first hypothesis **H1** (error is linearly related to the level of visibility), partially support **H2** (error is larger for less-realistic avatars), and support **H3** (error is larger in a more distance viewing perspective).

Of cases where visibility effects were detected, visibility consistently had the smallest effect on errors. This is surprising, as it may be expected that seeing less of a body would limit information about the entire body pose and produce a much worse interpretation. Because of this, designers may be free to use a wider set of avatars (ranging from a detailed full-body avatar to a minimal head and hands avatar) without compromising pointing interpretation by other users.

Total errors ranged between 1.0 to 1.5 units on average suggesting that participants were not very accurate in their interpretations. Most of this error is comprised of horizontal error, with vertical error being much smaller. This is likely due to the desktop study format (where depth perception is limited), the positioning of the participants/pointers, or a combination of the two. However, when comparing average errors between the two experiments in Figure 4 lower error was observed when view rotation was supported. This mostly impacted horizontal error, with rates for all treatments decreasing by approximately 0.47 units or 42.6%. However, this appears to have come at a cost of increased vertical error with an average increase or all treatments of 0.16 units or 47.7%. The net effect on total error is a decrease in total error of approximately 0.18 units or 12.4%.

In Experiment 1, the perspective was static to require users to focus on the visual properties of the avatar, though this may have created difficulty in understanding the environment's spatial layout. We believe the addition of camera rotation in Experiment 2 allowed users to better situate themselves and the other avatar better in the environment that overall reduced error. While vertical error did increase in Experiment 2, it was less than or equal to the horizontal component. Additionally, given the layout of our targets (where each was 1.0 unit from another) a vertical error of 0.5 means that the interpreted locations were closer to the true location than other targets. In summary, the key findings of our experiments are as follows:

1. *Avatar visibility minimally affected pointing interpretation.* Observed trade-offs solely due to visibility in significant cases were very small, and likely would not severely impact gesture interpretability in general applications.High precision use cases can incorporate alternative techniques if needed.

2. *Avatar visibility does not linearly affect pointing interpretation.* Solely reducing the visibility of the avatar is not guaranteed to decrease interpretation accuracy. In cases were significant differences were found, low visibility avatars periodically resulted in lower error than high visibility avatars.

3. *View and style largely impacted pointing interpretation.* Interpretation errors were always larger when observing from a greater distance. The largest differences due to style were observed in a far viewing position, but these were reduced when observers were nearer.

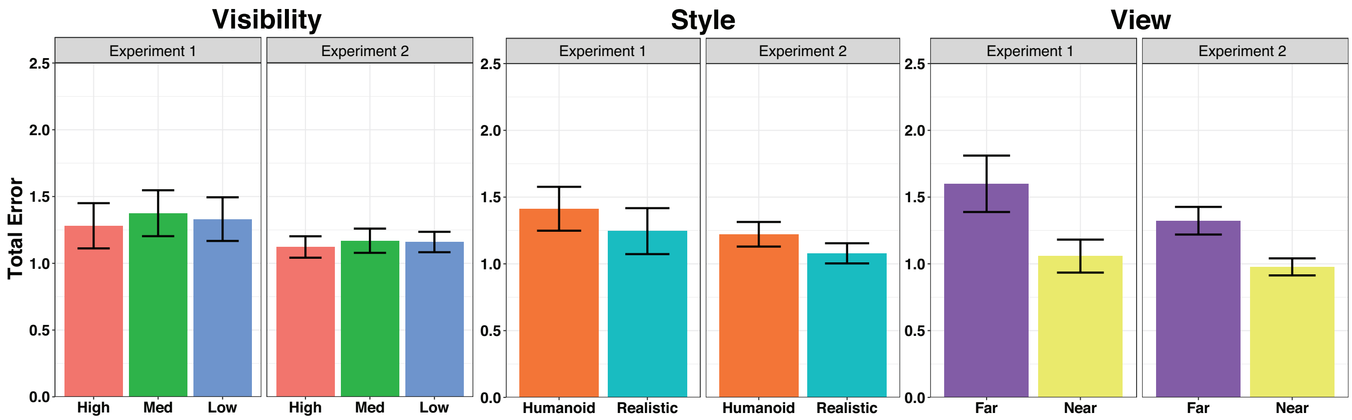

Figure 4: Average Total Error determined by individual factors for Experiments 1 and 2 with 95% confidence intervals. Small differences are seen based on visibility, while larger differences were seen based on style and view.

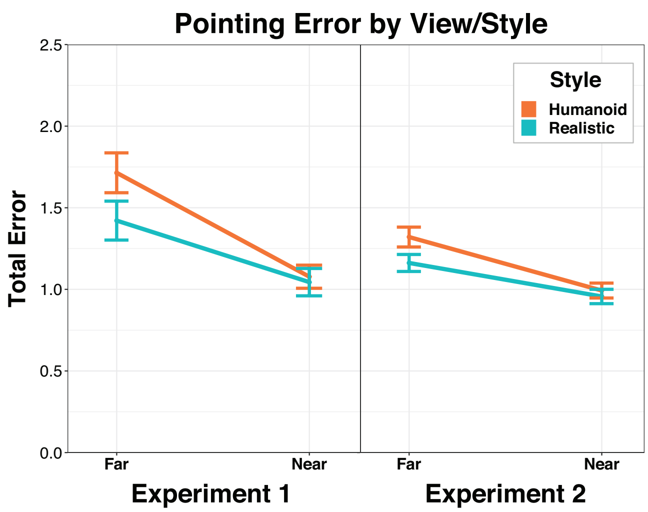

Figure 5: Interaction diagrams between style and view for Experiment 1 and 2. Presented is Total Target Error, grouped by view and style. Average values are plotted with 95% confidence intervals. *HUMANOID* avatars resulted in worse error, but only in the *FAR* view.

Facial details included in realistic avatars were more perceivable in far views, which helped with interpretation.

6.2 Future Work and Limitations

We utilized desktop applications to conduct our experiments, but other types of display devices should be examined due to the additional depth and spatial cues they afford such as stereoscopy in virtual reality experiences [34]. If these experiments were replicated in virtual reality, we expect that this would further reduce error rates (as was observed when view rotation was added in Experiment 2). Adding further ways to slightly modify the view, such as small amounts of translational head bobbling often experienced during idle standing, may also decrease error while users are positioned in a constant location.

As currently assessed, static pointing gestures limit this work since movement and animation is not included. However, when asked during the post-questionnaire about any factors that helped or hindered their ability to interpret the gesture, no participants indicated that poses appeared unnatural. Future work should examine collaborative situations with real or simulated animated avatars, as motions made by pointers while making their gestures could provide additional information which would increase accuracy. Additionally, it may be beneficial to consider left-handed pointing from the same perspective, as the arm is occluded differently than right-handed pointing. Another notable comment trend is a desire to perform the task in a more natural or realistic application. This would better reflect real application use cases.

While participants necessarily adjusted their view in Experiment 2 to select perceived pointing locations given the interaction methods, we did not conduct analysis of where participants were looking in the environment. Future work could include eye-tracking to determine what avatar features participants find most salient.

Finally, these results may only extend to cases of mid-range pointing with magnitudes of varying with different target locations. Examining target distance/clustering or with visible targets as opposed to the continuous wall of possible targets used in our experiments may yield larger effects of visibility. If known targets are visible to observers, we may expect hit rates of correct target objects to be higher if the distance between targets is sufficiently large.

7 Conclusion

Motivated by the understanding that virtual avatars are not always chosen to be displayed with a full-body, we conducted two experiments to determine how avatar visibility affects observer interpretation of pointing gestures. This work continues a thread of work that examines how low visibility avatars affect user experience, such as performance, social presence, and embodiment, in virtual environments. We provide insight into how *other* users in a scene are able to interpret natural means of communication, pointing, in shared experiences.

We examined two different avatar styles, three visibilities, and two viewing angles. In our first experiment, perspective was fixed to provide a focus on the visual properties of the avatar. In the second, users were able to rotate their view to accommodate their interpretation with varied perspectives. We observed that avatar style and viewing distance were large influencers on error while visibility was minimally impactful. However, we did discover that visibility affected directional error differently. In some cases, lower visibility avatars did not always produce higher error while in others it did; visibility, like style, is not necessarily linearly related to error.

These findings provide further evidence that low-visibility avatars are able to be used successfully in a wide range of scenarios. Observed differences in pointing interpretation error due to visibility alone were not substantial enough to warrant active consideration when choosing an avatar.

References

[1] B. Benda, S. Esmaeili, and E. D. Ragan. Determining detection thresholds for fixed positional offsets for virtual hand remapping in virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 269–278, 2020. doi: 10.1109/ISMAR50242.2020.00050

[2] P. Bernardis and M. Gentilucci. Speech and gesture share the same communication system. *Neuropsychologia*, 44(2):178–190, 2006.

[3] F. Biocca, C. Harms, and J. K. Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments*, 12(5):456–480, 2003.

[4] H. Brenton, M. Gillies, D. Ballin, and D. Chatting. The uncanny valley: does it exist. In *Proceedings of conference of human computer interaction, workshop on human animated character interaction*. Citeseer, 2005.

[5] P. Caserman, P. Achenbach, and S. Göbel. Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality. In *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)*, pp. 1–8. IEEE, 2019.

[6] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013.

[7] A. Corradini and P. R. Cohen. Multimodal speech-gesture interface for handfree painting on a virtual paper using partial recurrent neural networks as gesture recognizer. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, vol. 3, pp. 2293–2298 vol.3, 2002. doi: 10.1109/IJCNN.2002.1007499

[8] H. G. Debarba, E. Molla, B. Herbelin, and R. Boulic. Characterizing embodied interaction in first and third person perspective viewpoints. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–72. IEEE, 2015.

[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 SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, p. 1151–1160. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1518701.1518877

[10] S. Esmaeili, B. Benda, and E. D. Ragan. Detection of scaled hand interactions in virtual reality: The effects of motion direction and task complexity. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, March 2020.

[11] T. Feuchtner and J. Müller. Extending the body for interaction with reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 5145–5157. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025689

[12] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet. Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2062–2072, 2020. doi: 10.1109/TVCG.2020.2973077

[13] G. Gamelin, A. Chellali, S. Cheikh, A. Ricca, C. Dumas, and S. Otmane. Point-cloud avatars to improve spatial communication in immersive collaborative virtual environments. *Personal and Ubiquitous Computing*, pp. 1–18, 2020.

[14] B. Gao, J. Lee, H. Tu, W. Seong, and H. Kim. The effects of avatar visibility on behavioral response with or without mirror-visual feedback in virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 780–781, 2020. doi: 10.1109/VRW50115.2020.00241

[15] S. Goldin-Meadow. The role of gesture in communication and thinking. *Trends in cognitive sciences*, 3(11):419–429, 1999.

[16] D. T. Han, M. Suhail, and E. D. Ragan. Evaluating remapped physical reach for hand interactions with passive haptics in virtual reality. *IEEE transactions on visualization and computer graphics*, 24(4):1467–1476, 2018.

[17] W. A. IJsselsteijn, H. De Ridder, J. Freeman, and S. E. Avons. Presence: concept, determinants, and measurement. In *Human vision and electronic imaging V*, vol. 3959, pp. 520–529. International Society for Optics and Photonics, 2000.

[18] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[19] S. Kita. *Pointing: Where language, culture, and cognition meet*. Psychology Press, 2003.

[20] R. Kondo, M. Sugimoto, K. Minamizawa, T. Hoshi, M. Inami, and M. Kitazaki. Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity. *Scientific reports*, 8(1):1–8, 2018.

[21] D. Lala and T. Nishida. A collaborative pointing experiment for analyzing bodily communication in a virtual immersive environment. In *2012 IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing*, pp. 50–57, 2012. doi: 10.1109/ICCI-CC.2012.6311126

[22] S. Lopez, Y. Yang, K. Beltran, S. J. Kim, J. Cruz Hernandez, C. Simran, B. Yang, and B. F. Yuksel. Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[23] J. Lugrin, M. Ertl, P. Krop, R. Klüpfel, S. Stierstorfer, B. Weisz, M. Rück, J. Schmitt, N. Schmidt, and M. E. Latoschik. Any "body" there? avatar visibility effects in a virtual reality game. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 17–24, 2018. doi: 10.1109/VR.2018.8446229

[24] J.-L. Lugrin, J. Latt, and M. E. Latoschik. Avatar anthropomorphism and illusion of body ownership in vr. In *2015 IEEE Virtual Reality (VR)*, pp. 229–230. IEEE, 2015.

[25] S. Mayer, J. Reinhardt, R. Schweigert, B. Jelke, V. Schwind, K. Wolf, and N. Henze. Improving humans' ability to interpret deictic gestures in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020.

[26] S. Mayer, K. Wolf, S. Schneegass, and N. Henze. Modeling distant pointing for compensating systematic displacements. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 4165–4168, 2015.

[27] D. Murphy. Bodiless embodiment: A descriptive survey of avatar bodily coherence in first-wave consumer vr applications. In *2017 IEEE Virtual Reality (VR)*, pp. 265–266, 2017. doi: 10.1109/VR.2017.7892278

[28] C. D. Murray and J. Sixsmith. The corporeal body in virtual reality. *Ethos*, 27(3):315–343, 1999.

[29] K. Nickel and R. Stiefelhagen. Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. In *Proceedings of the 5th International Conference on Multimodal Interfaces*, ICMI '03, p. 140–146. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/958432.958460

[30] N. Ogawa, T. Narumi, H. Kuzuoka, and M. Hirose. Do you feel like passing through walls?: Effect of self-avatar appearance on facilitating realistic behavior in virtual environments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376562

[31] Y. Pan and A. Steed. Avatar type affects performance of cognitive tasks in virtual reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359996.3364270

[32] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, 22(3):779–787, 2013.

[33] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018.

[34] B. Pollock, M. Burton, J. W. Kelly, S. Gilbert, and E. Winer. The right view from the wrong location: Depth perception in stereoscopic multi-user virtual environments. *IEEE transactions on visualization and computer graphics*, 18(4):581–588, 2012.

[35] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go

interaction technique: Non-linear mapping for direct manipulation in vr. *Proc. of UIST'96*, 09 1998.

[36] R. Reinhard, K. G. Shah, C. A. Faust-Christmann, and T. Lachmann. Acting your avatar's age: effects of virtual reality avatar embodiment on real life walking speed. *Media Psychology*, 23(2):293–315, 2020.

[37] V. Schwind, S. Mayer, A. Comeau-Vermeersch, R. Schweigert, and N. Henze. Up to the finger tip: The effect of avatars on mid-air pointing accuracy in virtual reality. CHI PLAY '18, p. 477–488. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3242671.3242675

[38] V. Schwind, K. Wolf, and N. Henze. Avoiding the uncanny valley in virtual character design. *interactions*, 25(5):45–49, 2018.

[39] J. Seyama and R. S. Nagayama. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and virtual environments*, 16(4):337–351, 2007.

[40] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. V. Sanchez-Vives. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience*, 3:29, 2009.

[41] T. Q. Tran, H. Shin, W. Stuerzlinger, and J. Han. Effects of virtual arm representations on interaction in virtual environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3139131.3139149

[42] D. Vogel and R. Balakrishnan. Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pp. 33–42, 2005.

[43] T.-Y. Wang, Y. Sato, M. Otsuki, H. Kuzuoka, and Y. Suzuki. Effect of body representation level of an avatar on quality of ar-based remote instruction. *Multimodal Technologies and Interaction*, 4(1), 2020. doi: 10.3390/mti4010003

[44] N. Wong and C. Gutwin. *Where Are You Pointing? The Accuracy of Deictic Pointing in CVEs*, p. 1029–1038. Association for Computing Machinery, New York, NY, USA, 2010.

[45] N. Wong and C. Gutwin. Controlling an avatar's pointing gestures in desktop collaborative virtual environments. In *Proceedings of the 17th ACM International Conference on Supporting Group Work*, GROUP '12, p. 21–30. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2389176.2389180

[46] B. Yoon, H. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556, 2019. doi: 10.1109/VR.2019.8797719

[47] R. C. Zeleznik, A. S. Forsberg, and P. S. Strauss. Two pointer input for 3d interaction. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pp. 115–120, 1997.

[48] A. Zenner and A. Krüger. Estimating detection thresholds for desktop-scale hand redirection in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 47–55, 2019. doi: 10.1109/VR.2019.8798143