Virtual Big Heads in Extended Reality: Estimation of Ideal Head Scales and Perceptual Thresholds for Comfort and Facial Cues

ZUBIN CHOUDHARY, AUSTIN ERICKSON, and NAHAL NOROUZI, University of Central Florida
KANGSOO KIM, University of Calgary
GERD BRUDER and GREGORY WELCH, University of Central Florida

4

Extended reality (XR) technologies, such as virtual reality (VR) and augmented reality (AR), provide users, their avatars, and embodied agents a shared platform to collaborate in a spatial context. Although traditional face-to-face communication is limited by users' proximity, meaning that another human's non-verbal embodied cues become more difficult to perceive the farther one is away from that person, researchers and practitioners have started to look into ways to accentuate or amplify such embodied cues and signals to counteract the effects of distance with XR technologies. In this article, we describe and evaluate the *Big Head* technique, in which a human's head in VR/AR is scaled up relative to their distance from the observer as a mechanism for enhancing the visibility of non-verbal facial cues, such as facial expressions or eye gaze. To better understand and explore this technique, we present two complimentary human-subject experiments in this article. In our first experiment, we conducted a VR study with a head-mounted display to understand the impact of increased or decreased head scales on participants' ability to perceive facial expressions as well as their sense of comfort and feeling of "uncannniness" over distances of up to 10 m. We explored two different scaling methods and compared perceptual thresholds and user preferences. Our second experiment was performed in an outdoor AR environment with an optical see-through head-mounted display. Participants were asked to estimate facial expressions and eye gaze, and identify a virtual human over large distances of 30, 60, and 90 m. In both experiments, our results show significant differences in minimum, maximum, and ideal head scales for different distances and tasks related to perceiving faces, facial expressions, and eye gaze, and we also found that participants were more comfortable with slightly bigger heads at larger distances. We discuss our findings with respect to the technologies used, and we discuss implications and guidelines for practical applications that aim to leverage XR-enhanced facial cues.

CCS Concepts: • **Human-centered computing → Virtual reality**; **User studies**; • **Computing methodologies → Virtual reality;**

Additional Key Words and Phrases: Virtual environments, social virtual reality, outdoor augmented reality, non verbal communication

ACM Reference format:
Zubin Choudhary, Austin Erickson, Nahal Norouzi, Kangsoo Kim, Gerd Bruder, and Gregory Welch. 2023. Virtual Big Heads in Extended Reality: Estimation of Ideal Head Scales and Perceptual Thresholds for Comfort and Facial Cues. *ACM Trans. Appl. Percept.* 20, 1, Article 4 (January 2023), 31 pages.
https://doi.org/10.1145/3571074

This material includes work supported in part by the National Science Foundation under Award Number 1564065 (Dr. Ephraim P. Glinert, IIS) and Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University, respectively; the Office of Naval Research under Award Numbers N00014-21-1-2578 and N00014-21-1-2882 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch).
Authors' addresses: Z. Choudhary, A. Erickson, N. Norouzi, G. Bruder, and G. Welch, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816, USA; emails: {zubinchoudhary, ericksona}@knights.ucf.edu, {nahal.norouzi, gerd.bruder, welch}@ucf.edu; K. Kim, University of Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada; email: kangsoo.kim@ucalgary.ca.

1544-3558/2023/01-ART4 $15.00
https://doi.org/10.1145/3571074

1 INTRODUCTION

Recent advances in **extended reality (XR)** technologies, including **virtual reality (VR)** and **augmented reality (AR)**, have given rise to new opportunities in fields such as simulation, training, and healthcare. For many of these, collaboration between multiple users, their avatars, and/or embodied agents is an important aspect of the experience. Collaborative environments aim to facilitate effective communication by leveraging different types of embodied human signals and cues, which includes verbal and non-verbal communication channels. Researchers have introduced and investigated a wide range of methods to support and accentuate such channels with XR technologies [42, 51, 56, 57, 62, 64]. However, there are still many challenges we face in collaborative environments [44]. For instance, a crucial factor in face-to-face communication is the dependency on the proximity of interlocutors. The closer one is to another human, the easier it is to perceive non-verbal embodied cues. However, if one moves farther away, details such as facial features or expressions become more difficult to perceive. In this article, we explore how we can leverage VR/AR technologies to accentuate or amplify some of the non-verbal embodied cues that are reduced or lost in face-to-face communication due to distance.

We hypothesize that for longer distances, the *Big Head* technique [14], which we introduced in VR/AR as a mechanism for up-scaling heads of interlocutors relative to the rest of their body, can be an effective (subtle or overt) mechanism for accentuating and enhancing the visibility and transmission of embodied cues. We further hypothesize that humans have a tolerance for seeing body parts at different sizes, and up-scaling certain parts can improve their respective effectiveness in conveying social cues. Originally, a variant of this *Big Head* technique was introduced by video game developers who were looking for solutions to compensate for the low pixel resolution of traditional TV and desktop screens, which made it hard to make out features of their game characters at real scale [33]. Resolution has progressed tremendously on TV and desktop screens; however, we are seeing similar challenges with VR/AR displays today.

Although there are similarities in VR and AR technologies, there are also important differences that make it impossible to just transfer results from VR to AR or vice versa in the scope of collaborative environments [27, 47]. For instance, although limited resolution is also a driving challenge in AR environments, **optical see-through (OST)** displays particularly are further limiting the perception of embodied cues due to their low contrast [26] and additive light model [31], which causes the user's real-world physical environment to be visible behind any superimposed virtual imagery [28, 29]. This is especially true if the imagery is rendered in dark colors, such as when a rendered human interlocutor has a darker skin color or wears dark clothes [60]. These differences are particularly noticeable in outdoor AR environments, and may cause rendered humans to be more difficult to perceive than similar imagery displayed in VR, which may affect how embodied cues are perceived at real scale and other scales.

In this article, we present two human-subject studies aimed at understanding the importance and influence of an increased or decreased head scale in social VR and outdoor AR environments on a user's perception and their sense of comfort. In our first experiment, which was originally published at IEEE VR 2020 [14], we identified preferences and thresholds for head scales in a VR environment over a range of distances up to 10 m in two task contexts with two alternative head scaling methods. In this extended journal paper, we further present a second experiment, where we investigated distances up to 90 m in outdoor AR in four task contexts. Our work contributes to the research fields on collaborative XR environments with scaled human representations by providing empirical data especially for the following research questions:

- *RQ1*: How does a virtual human's head scale in VR/AR correlate with one's ability to perceive their *facial expressions*?
- *RQ2*: How does it correlate with one's ability to perceive their *eye gaze directions*?
- *RQ3*: How does it correlate with one's ability to *recognize or identify* the virtual human?
- *RQ4*: How does it correlate with one's *sense of comfort* or "uncanniness"?
- *RQ5*: Can we determine meaningful perceptual thresholds for head scales, and how do these thresholds change with respect to one's distance to a virtual human, from intimate space to public space?

This article is structured as follows. Section 2 discusses related work. Section 3 describes Experiment E1 (social VR). Section 4 describes Experiment E2 (outdoor AR). Section 5 summarizes and discusses our main findings in the two experiments. Section 6 concludes the article and discusses future opportunities for research.

2 RELATED WORK

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

2.1 Collaboration in XR

Collaborative environments in XR [61] enable users to meet and share experiences together in the same environment. Social VR, such as VRChat,[1] or Meta Horizon,[2] in particular, has increased in popularity, and its consumer market also continues to grow. These environments allow vocal and textual communication among users, and visual aspects of communication, such as appearance, gestures, or expressions, also play an integral 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 [21]. Virtual avatars can have an effect on other users' perception of some aspects of the real user behind the avatar, such as the user's personality or lifestyle [7]. It can even impact one's self-perception [80], which may in turn affect one's behavior during interactions. Since users have the ability to customize their virtual representation in social VR, there may be benefits in scaling body parts, like the *Big Head* technique, to improve the visual aspects of communication.

Similar to collaborative VR spaces, the field of AR is experiencing a growing demand for collaborative applications by fields such as simulation, training, and healthcare. Some popular use cases being investigated with AR **head-mounted displays (HMDs)** are spatial problem solving and remote guidance, allowing experts to guide a user when distance is a factor [1, 17, 59]. Such AR use cases may benefit from body scaling approaches, like the *Big Head* technique, in a similar fashion as collaborative VR spaces. In collaborative AR interactions, augmenting scaled body parts can be useful for sharing embodied cues—for example, police officers who carry out tactical operations need to quickly identify all involved humans and understand their teammates' state and intentions at a distance. Not only can a person be quickly identified, but visual nonverbal signals, like facial expressions and eye gaze, can be enhanced for a clearer message during more difficult interactions, such as when people are standing farther away from each other.

2.2 Non-Verbal Embodied Cues and Distance as a Factor

Non-verbal cues are highly important in interpersonal interactions, and it has been studied that when such cues are reproduced in an immersive virtual environment by a virtual avatar, a higher sense of social presence can be expected [74]. 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. [11] 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. 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 [18]. Likewise, accurately simulating realistic eye gaze behavior for a virtual avatar can increase the level of participation in a conversation with the avatar [2, 3, 16, 66].

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

In real life, by observing such cues on the people around us, our behaviors are consciously and sometimes unconsciously changed. For example, the way in which we navigate environments has been shown to be heavily influenced by the facial expressions of the nearby individuals within the environment. Bailenson et al. [5] showed that participants tended to allocate extra space when passing individuals that engaged in mutual gaze. Research by Schrammel et al. [68] also showed that participants tended to fixate more predominantly on cues that suggested a threat rather than non-threatening cues. Seidel et al. [69] showed that although participants were more likely to approach people with happy expressions and avoid people with angry expressions, their behavior was more complicated for expressions such as disgust or sadness. Such studies have investigated the way in which facial expressions affect the behavior of the observer in close proximity; however, due to the small size of facial features, it is difficult to accurately distinguish expressions at long distances [35, 73].

The ability of a person to perceive a feature at a distance d is dependent on the size s of the feature. The relationship between distance d, feature size s, and the visual angle θ is $tan(\theta) = s/d$. The minimum visual angle θ a person is capable of resolving is thought to be approximately 1 arc minute (1/60) of a degree [78]. Assuming the size of the region containing a person's facial features is roughly $s = 0.1$ m, the lower limit of visual acuity is approximately $d = 340$ m in distance. Around that distance, one's ability to resolve individual features on a person's face, such as to determine their facial expression, starts to become problematic. Further, if one wants to estimate a distant human's eye gaze direction, one needs to be able to resolve their eyeball. Assuming the average size of a human eye is about 0.025 m [6], the maximum distance (calculated in the same way) is approximately 85 m. Since the eyes are some of the smaller features used when detecting facial expressions, this estimate of 85 m should be reasonably close to the farthest distance at which an average person is capable of identifying the facial expression of another human.

Several previous works presented empirical studies that investigated human perception of facial expressions at long distances in the real world (without XR). One of these is the work by Hager and Ekman [35] in 1979, where they found that participants were able to resolve facial expressions at distances of up to between 30 to 45 m, where participants' accuracy declined but they still showed some manner of success. Their work also showed that the ability of an observer to correctly resolve facial features depends on the gender of the observed person as well as the type of expression being made, with happiness and surprise being easier to distinguish on females compared to fear, disgust, sadness, and anger, whereas expressions of happiness, surprise, and anger were easier to identify on males compared to fear, sadness, and disgust. Smith and Schyns [73] simulated facial expressions at distances of up to 100 m in their 2009 work, where they found that participants were sensitive, in decreasing order, to expressions of happiness, surprise, disgust, fear, anger, and sadness. Their work also showed that participants' performance decreased rapidly for distances between 7 and 50 m, before leveling out for distances greater than 50 m.

These works demonstrate that humans have difficulty identifying facial features and expressions at long distances. In the next section, we discuss how through the use of XR technology, it is possible to change the appearance of a human by scaling embodied features, such as faces, to be able to better perceive embodied social cues and signals over long distances.

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 [55, 70]. Along these lines, with respect to head proportions for a humanoid robot, DiSalvo et al. [19] 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 not detract from the effectiveness of a humanoid character.

Our sense of comfort with other real humans or their representations in a social situation is also dependent on *distance*: Hall [36] 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), and *social* space (up to 3.7 m) to *public* space. 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 [4, 50]. People in the real world but also in an immersive virtual environment generally feel uncomfortable if another person, particularly a stranger, invades their personal space or even their intimate space [38]. In such cases, humans generally experience an activation of the amygdala in response to this violation of social norms [39]. Bailenson et al. [5] 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. A breach of such social norms can reduce a person's sense of social presence—particularly comfort—with the entity.

2.4 Head Scaling in XR

XR technologies are capable of replicating spatial appearances of objects or humans in the real world, and they further provide the opportunity to go beyond realism [53] and potentially enhance a user's perception of real or virtual entities by making them easier to see, such as by making them, in their entirety or in part, appear larger.

Some recent works in VR have examined how head magnifications can be used to alter distance perception of virtual human avatars. Choudhary et al. [13] investigated the effects of avatar head scaling on distance perception in VR, which revealed a significant effect of big heads on distance judgments only if heads were presented as floating objects in VR, but not when spatially anchored and attached to a human body at true scale.

Although these studies were conducted in VR, there is also a growing interest in the field of AR to leverage such embodied scaling methods for applications such as social AR, in which users interact with virtual human avatars in a shared physical space, or when interacting with computer-controlled human agents in AR simulation or training contexts [34, 75, 76, 79]. In particular, OST-HMD technologies have made major advances over the past years and are becoming more widely available for indoor and outdoor AR applications [8, 9, 30, 40, 41, 81]. Although state-of-the-art OST-HMDs such as the Microsoft HoloLens have comparatively small fields of view, they make up for it with a high angular resolution, which far exceeds that of comparable video see-through or VR HMDs [45, 46, 54]. As these displays are hence capable of presenting more visual details per arc second, the demand for up-scaling of embodied cues is less pressing. However, the *additive light model* of the currently available consumer OST-HMDs is based on the concept that light emitted by the display can be added to a view of the real world, but light cannot be "taken away" [31]. This causes the effect that display areas that are rendered with a light color will occlude the physical background, whereas dark rendered areas appear *transparent*, through which the background is visible. Hence, with OST-HMDs, the perceived contrast of virtual imagery can vary considerably depending on the lighting conditions [26].

3 EXPERIMENT E1: BIG HEADS IN VR

In this section, we present our first user study, in which we evaluate the effects of a virtual human's head scale on participants' perception of the virtual human in a social VR context with an immersive VR HMD. We use the term *virtual human* as a blanket term, as we believe that the results of this and the next experiment are useful for all applications that use human characters under human control (avatars) or computer control (agents), as well as humans that are virtually augmented.

3.1 Participants E1

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

Fig. 1. Examples of virtual humans with different head scales: normal head scale (1x) (a), big head (2x) with the *fixed neck height* scaling method, in which the head protrudes upward from the neck (b), small head (0.5x) with the *fixed neck height* scaling method (c), big head (2x) with the *fixed eye height* scaling method, in which the body is scaled down to maintain the same eye height (d), and small head (0.25x) with the *fixed eye height* scaling method, in which the body is scaled up to maintain the eye height (e).

between 18 and 31 years, $M = 23.96$, $SD = 4.11$). All of the participants had normal or corrected-to-normal vision, whereas 8 participants wore glasses and 3 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. Seventeen participants had used a VR HMD before, and 8 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 a monetary compensation of $15 for their participation.

3.2 Materials E1

To investigate the participants' perception of different virtual human head scales in a shared VR environment, we prepared a virtual space containing a virtual human in which participants could view as well as adjust the scale of the virtual human's head. In this section, we describe the details of the virtual human 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 human for our study (Figure 1). The model was imported into the Unity game engine (version 2019.2.9.f1) and placed in a virtual hallway environment. For that environment, we simulated a hallway with the dimensions 5 m (width) × 3 m (height) × 15 m (length) as shown in Figure 2 so that participants could see the virtual human 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. Among different possible facial expressions, we decided to use four specific expressions: happy, sad, angry, and skeptical (Figure 3), and the virtual human 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 virtual human's body height was adjusted to match that of the participant. The scaling methods were applied relative to the body height.

The virtual human and the surrounding virtual environment were displayed through an immersive VR HMD, HTC Vive Pro. 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 HMD was connected to a host PC (Intel Core i7-7820HK CPU @ 2.90 GHz, 32 GB of 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

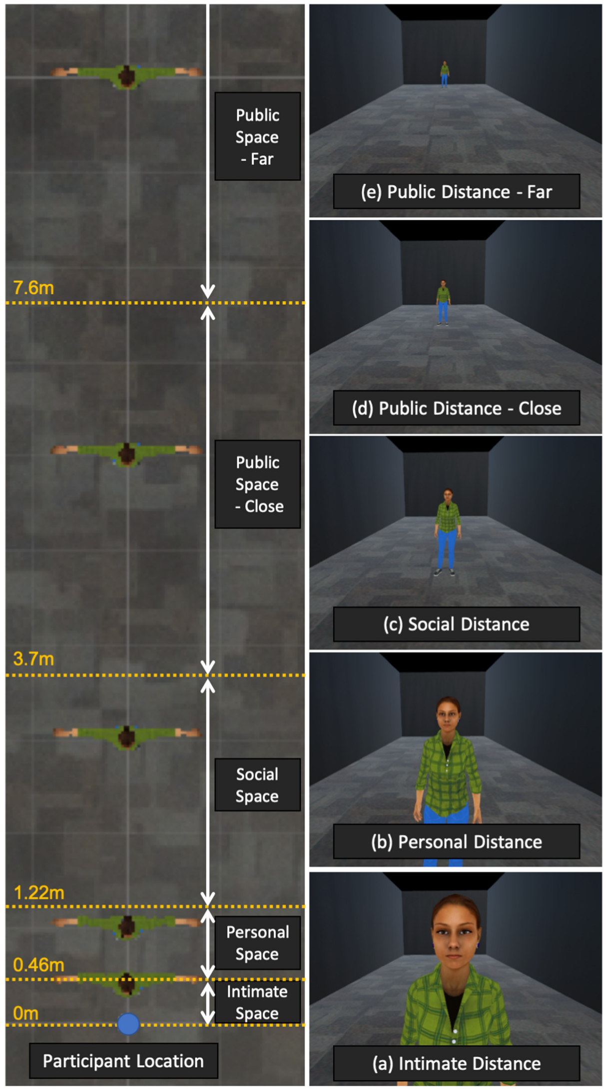


Fig. 2. Illustration of the tested distances and proxemics in Experiment E1. Screenshots on the right show the virtual human at distances of 0.4 m (intimate space) (a), 1 m (personal space) (b), 3 m (social space) (c), 6 m (close public space) (d), and 10 m (far public space) (e).

the HMD, and could adjust the scale of the virtual human head using the Vive Pro Controller (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 human. After an initial familiarization period, participants were asked to remain standing at a fixed position in the lab. The corresponding angular resolution was 15 pixels per degree.

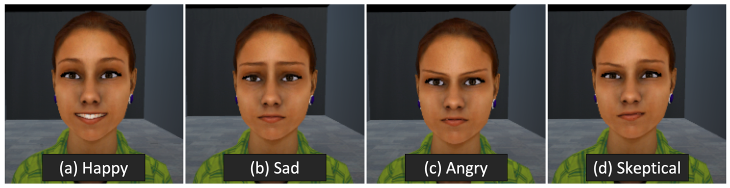


Fig. 3. The virtual human's four facial expressions used for Experiment E1: happy (a), sad (b), angry (c), and skeptical (d).

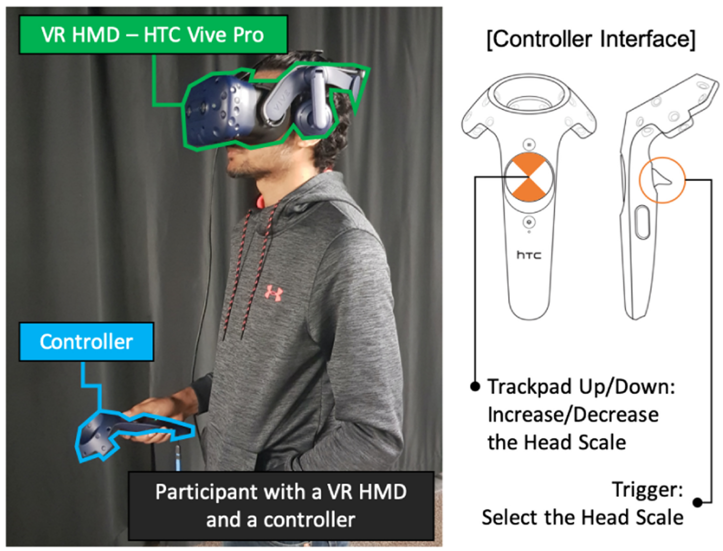


Fig. 4. Annotated photo showing a participant in Experiment E1 wearing the HMD and holding the controller during the task. The controller interface to adjust the scale of the virtual human's head is shown on the right.

3.3 Method E1

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 next:

- *Scaling method (two levels)*: We developed two mechanisms to adjust the virtual human's head scale relative to the rest of the virtual human's body (see Figure 1):
  - *Fixed neck height*: In our first scaling method, the virtual human's head scale was changed upward from the virtual human's neck height so that participants had to look up to the virtual human's head when the head scale was increased. As a result, the eye height of the virtual human changed depending on the head scale, meaning that with an increased head scale, the virtual human would look down at the participant, whereas a reduced head scale meant that the virtual human would look up at the participant.
  - *Fixed eye height*: Considering the potential influence of a mismatch in the eye height of the virtual human 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, whereas 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 virtual human always maintains the same eye height at all times.

- *Task context (two levels)*: We decided to test two separate task contexts in the virtual human head scaling tasks:
  - *Facial expression*: In this task context, participants were asked to adjust the scale of the virtual human's head while focusing on their performance in recognizing the virtual human'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 virtual human's head while focusing on their perception of the virtual human and their feeling of comfort when interacting with the virtual human 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 a virtual human in social VR.
- *Distance (five levels)*: For the experiment, the virtual human appeared at five different distances, considering the proxemics in social interactions based on the proxemics theory of Hall [37] (see Figure 2):
  - *Intimate*: Intimate space is considered for touching or whispering within a distance of 0.46 m, so we located the virtual human 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 virtual human 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 virtual human 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 virtual human 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 virtual human 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 earlier:
  - *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 virtual human 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 virtual human'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 ideal 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. Afterward, 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 human's head for two different task contexts in an immersive virtual environment. The experimenter explained all 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 human standing in a virtual hallway facing toward them. The virtual human was introduced to participants as a typical representation of a real human interlocutor in a social VR environment. Before the actual experiment, there was a practice session so that the participants could have an idea about the virtual human's facial expressions that they should look out for and how to change the virtual human's head scale using the 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 recognizing the facial expressions, the virtual human 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 2-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 human conditions as well as the reasoning behind their answers. Finally, the experiment ended with a monetary compensation.

3.4 Measures and Hypotheses E1

In this section, we describe the measures and hypotheses that we used for the experiment to understand our research questions outlined in Section 1.

3.4.1 Virtual Human Head Scale. As described in Section 3.3, participants experienced the virtual human 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 virtual human'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 as follows:

- *H1* For the facial expression task context, participants will increase the virtual human's head scale as the distance between the virtual human and themselves increases.
- *H2* For the comfort level task context, participants will maintain the virtual human's head scale at the original scale even if the distance between the virtual human 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 virtual human 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 human and the head scale chosen by our participants. In other words, if the facial expressions can be well recognized at a distance of 1 m with a certain amount of screen space and pixels, participants would up-scale the head to twice its size if the virtual human is presented at a distance of 2 m, three times its size at 4 m, and so on.

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 virtual human's face at the same eye level as the participant's eyes, which makes it easier to see the virtual human'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 human 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.

3.5 Results E1

The results are shown in Figure 5, in which Figure 5(a) and 5(b) show the results when tasked with identifying a comfortable range of head scales, whereas Figure 5(c) and (d) show the results when tasked with identifying head scales so that participants can still perceive the virtual human's facial expressions. The left column shows results for the *fixed neck height* scaling method, in which only the head size is scaled upward/downward, whereas 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 human. The y-axes show the head scales defined by the participants; the results are presented on a logarithmic (base 2) scale on that axis. This is only done for presentation purposes, as we did not use a log transform in the statistical analysis. 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.

3.5.1 Virtual Human 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, $F(1, 21) = 7.93$, $p = 0.01$, $\eta_p^2 = 0.27$, 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, $F(1, 21) = 22.44$, $p < 0.001$, $\eta_p^2 = 0.52$, indicating that the task to identify *facial expressions* resulted in larger head scales than the task to identify *comfort levels*.

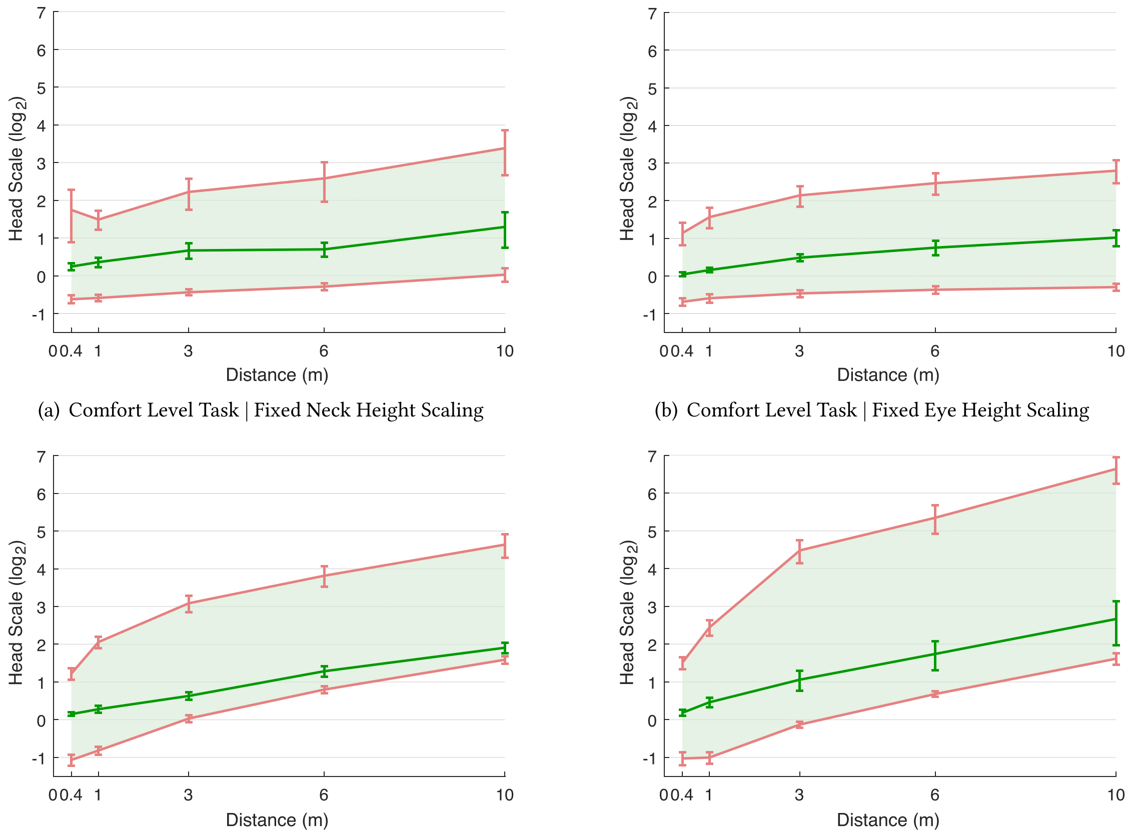
Comfort Level.

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, $F(1.23, 27.06) = 7.91$, $p = 0.006$, $\eta_p^2 = 0.26$. Post hoc tests showed that all pairwise comparisons were significant ($p < 0.05$), 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, $F(1.11, 14.31) = 4.64$, $p = 0.038$, $\eta_p^2 = 0.17$. All pairwise comparisons were significant ($p < 0.05$), 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, $F(2.14, 47.08) = 9.85$, $p < 0.001$, $\eta_p^2 = 0.31$. Post hoc tests showed that all



(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

Fig. 5. Results of Experiment E1 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 human; the y-axes show the indicated head scales by the participants (on a logarithmic scale with base 2). The red lines indicate the thresholds, whereas the green line indicates ideal head scales.

pairwise comparisons were significant ($p < 0.05$), 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, $F(1.45, 31.82) = 14.19$, $p = 0.038$, $\eta_p^2 = 0.39$. All pairwise comparisons were significant ($p < 0.05$), except between distances 3 and 6.

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

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*, $F(1.26, 27.60) = 115.75$, $p < 0.001$, $\eta_p^2 = 0.84$. Post hoc tests showed that all pairwise comparisons were significant ($p < 0.05$), except between distances 0.4 and 1.

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

Moreover, we found a significant main effect of *distance* on head scales for the *ideal* task, $F(1.32, 29.08) = 48.26$, $p < 0.001$, $\eta_p^2 = 0.69$. All pairwise comparisons were significant ($p < 0.05$), 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, $F(1.18, 25.98) = 52.46$, $p < 0.001$, $\eta_p^2 = 0.71$. Post hoc tests showed that all pairwise comparisons were significant ($p < 0.05$), except between distances 0.4 and 1.

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

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

3.5.2 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, 8 participants were in favor of the fixed neck height method, 14 were in favor of the fixed eye height method, and 1 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, 8 participants were in favor of the fixed neck height method, 13 were in favor of the fixed eye height method, and 2 were undecided.

3.6 Discussion and Limitations E1

In this section, we summarize the main findings and discuss implications for the use of the *Big Head* technique in social VR, and we also address limitations of our experiment. Overall, our results show that participants generally increased the virtual human's head scale both for the facial expression recognition and the comfort level task context.

3.6.1 Virtual Human 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 virtual human's head scale when the virtual human 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 virtual human's face when they were too distant. In other words, the HMD's limited screen resolution took social cues away for longer distances, whereas increasing the head scale could bring them back.

In addition, based on the *minimum* head scale results shown in Figure 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 virtual human was closer than 3 m, whereas they started to increase the head scale when the virtual human 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 to be able to perceive facial expressions, we propose the values in Figure 5(c) and (d) as a reference for practitioners in this field. Although 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 virtual human'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.

3.6.2 Virtual Human Head Scales Increased over Distance for Comfortable Proportions. Interestingly, in contrast to our Hypothesis H2, our results show that participants still increased the virtual human's head scale even for the comfort level task, in which they were asked to focus on the virtual human's most comfortable head scales and proportions. Given the *ideal* head scale results shown in Figure 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 tested distances. In particular, it is interesting that a head scale with a factor of 2 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 [10, 65], particularly their ability to estimate the virtual human'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 human, even at such close distances.

Our Hypothesis H2 was based on the consideration of potential *Uncanny Valley* [55] effects caused by a virtual human 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 virtual human'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 virtual human 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" virtual humans might actually be perceived as more comfortable than traditional representations.

3.6.3 Preferences of Scaling Methods for Big Head Virtual Humans. 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 human 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."

However, the participants who preferred the fixed eye height method considered its benefits for the convenience of seeing the virtual human'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 [32], but they also emphasize the importance of the rest of the body.

3.6.4 Limitations and Future Work. Our study showed interesting effects of virtual human 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, this study used a single female virtual human character as the visual stimulus to create a consistent VR experience. We understand that different virtual human appearances and rendering styles could influence a user's perception of a virtual human [48, 83]. The range of a participant's choice of a virtual human'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 virtual human'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 article.

Third, the virtual human in our study setting was static with idle standing animations at different distances. Virtual humans 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 virtual humans. Future work may investigate how dynamic virtual human movement and changes in head scales affect a users' perception of social virtual humans.

Last but not least, Experiment E1 looked at virtual humans in an indoor VR setting using an immersive VR-HMD, where we scaled the virtual head of a virtual human character. In contrast, in our effort to gain a broader understanding of the *Big Head* technique, our Experiment E2 (which follows) was designed to provide further data points in an outdoor AR setting (instead of indoor VR) using an OST-HMD (instead of VR-HMD), where we scaled the virtual head of a physical human stand-in (instead of a virtual human character).

4 EXPERIMENT 2: BIG HEADS IN OUTDOOR AR

In this section, we present our second experiment, in which we evaluate the effects of head scale on participants' perception of an AR head with an OST-HMD. For this study, as indicated in Section 3.6.4, we decided to use a physical human stand-in (mannequin) for experimental purposes and as a conservative setup, where we matched the AR head as an overlay over the physical body (see Figure 6). We consider such magnified heads (or other body parts) a useful technique for future AR use cases, whether based on purely virtual AR humans, mixed physical-virtual representations, or overlays of a real-time 3D reconstructed head over the view of a real person. In comparison to Experiment E1 (refer Section 3), we decided to add eye gaze detection and face detection tasks. In addition, we decided to investigate an outdoor AR environment compared to our previous VR environment. According to Hager and Ekman [35], humans are capable of accurately detecting facial expressions up to 45 m (not using XR). So we decided to test the effect of head scales over three far-away distances (30, 60, and 90 m). We will now discuss Experiment E2 in detail in the next sections.

4.1 Participants E2

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 20 participants from our university community (14 male and 6 female; ages between 19 and 31 years, $M = 25.6$, $SD = 3.3$). All participants had normal or corrected-to-normal vision, whereas 8 participants wore glasses and 3 wore contact lenses. None of the participants reported known visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance. The participants were

Fig. 6. Illustration of *Big Head* scaling in outdoor AR. From left to right, the images show an unaugmented view of two humans at different distances, an AR head that was up-scaled to be able to barely identify the face, the same head up-scaled more to perceive the facial expression (smiling), and even more up-scaled to make out the gaze direction (looking right).

either students or non-student members of our university, who responded to open calls for participation, and received a monetary compensation of $15 for their participation.

4.2 Materials E2

To investigate human perception of different AR human head scales in an outdoor AR environment, we set up our experiment outdoors (Figure 7) with a mannequin that indicated the three distances at which participants then adjusted the scale of an AR human's head with a HoloLens and an Xbox Controller (Figure 8). In this section, we describe the details of the AR human and the study settings for the experiment.

We used AR humans in this experiment that were life-size 3D female virtual human models obtained from Adobe Mixamo. The models were cropped using Blender, to use their heads only, and imported into the Unity game engine (version 2019.2.9.f1). Figure 9(c) shows the different 3D characters we used. We used a real-life 1.8 m tall female mannequin (see Figure 8) to indicate the distances in the experiment at which participants had to scale a virtual head that we aligned with the mannequin's neck.

As mentioned in Section 3.6.4, we decided to use a physical mannequin stand-in and a virtual head to help us broaden our understanding of the *Big Head* technique in terms of full or partial AR human representations. We started off planning this experiment with a real human standing at different distances; however, for practical purposes and to ensure that the experiment trials were sufficiently controlled and could be compared, we opted for a mannequin as a human stand-in. Although recent work by Choudhary et al. [15] presented a first demonstration of the *Big Head* technique for real humans in AR implemented using live head-mounted camera capture with a computer vision algorithm, due to our choice of a static mannequin with a static face, we opted to use an animated virtual head overlaid over the head of the mannequin instead, i.e., not a live video capture of the mannequin's head. We believe that this was an interesting and informative choice since our experiment focuses on the technique more than the single use case of magnifying real heads captured by a camera [15].

The 3D heads we used were programmed to perform various facial expressions using blendshapes on their faces such as lip and eye brow movements. Among different possible facial expressions, we decided to use four specific expressions: happy, sad, angry, and skeptical (see Figure 9). We could have tested for different stylized facial expressions [77], but we made this decision based on similar facial expressions being used in Experiment E1 (refer to Section 3). Instead of testing only one facial expression per condition, we presented them in a continuous loop, which allowed participants to align their responses based on their perception of all of them, including the dynamics of changing facial expressions; arguably, it would be more difficult to only see one static facial expression and there are some expressions that are harder to see (including micro-expressions), but as in Experiment E1, we were mainly interested in understanding typical expressions in dynamic contexts.

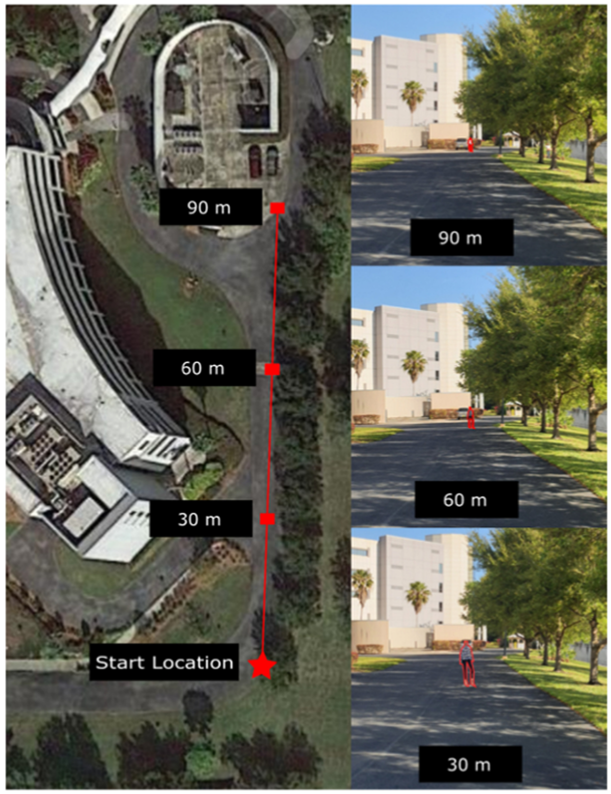


Fig. 7. In Experiment E2, we tested three AR human distances (30, 60, and 90 m) in the shown outdoor environment.

We further decided to test four different eye gaze directions. Eye gaze is an important cue in interpersonal communication, which ranges from eye contact to deictic interacting with the environment. As discussed in Section 2.2, eye gaze corresponds to some of the smallest facial features that are regularly employed by humans. The 3D character in our experiment looked left, right, up, and down (see Figure 9). We presented them in a loop, similar to the loop we used for facial expressions to make sure that participants will receive not only static gaze cues but typical dynamic gaze behaviors.

Last, we also decided to test three 3D character faces. Although it is arguably more difficult to read facial expressions or gaze directions than to identify a person or distinguish one from another based on the overall appearance of their head, we nonetheless decided to include them in our experiment, as we believe that it is important to understand how large heads have to be to allow us to identify/distinguish them. To be consistent with the other conditions, we also looped the heads.

The AR human was displayed through an OST-HMD, a Microsoft HoloLens 1. The HMD provided a 30-degree horizontal and 17.5-degree vertical field of view, and had a resolution of 1280 ×720 pixels per eye at a refresh rate of 60 Hz. The angular resolution was 43.6 pixels per degree. The HMD was connected to a backpack PC (MSI, Intel Core i7-7820HK 2.9 GHz CPU, 16 GB of RAM, Nvidia GTX 1070 graphics card, Windows 10 Pro). We utilized a cellular mobile hot-spot to stream imagery to the HoloLens via Unity in Holographic Remoting mode. As the experiment was conducted over multiple days in a sunny outdoor environment, two stacked sheets of neutral density filters, one with an optical density of 0.6 and another with a density of 0.9, were added to the HoloLens. This could block 37.5% of all incoming light and ensured a good contrast between the virtual imagery and the

Fig. 8. Annotated photos showing on the left side a participant in Experiment E2 with the HoloLens and XBox controller and on the right side the mannequin we placed at the three physical distances with an AR *Big Head* overlay.

participants' physical environment. During the experiment, participants were standing on a marked location on a private street in an outdoor environment, wearing the HMD. They could align and adjust the scale of the AR human head using a handheld XBox controller (see Figure 8). Specifically, they could tilt the right joystick up or down to choose and fine-tune the virtual head scale. When they felt confident about the chosen head scale and decided to select it, they could press the "A" button on the controller, which saved their chosen head scale for the current condition.

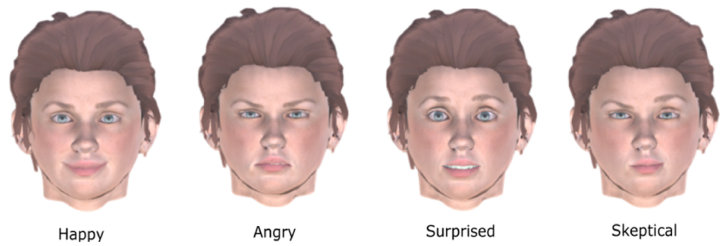
4.3 Methods E2

We used a $4 \times 3 \times 3$ full-factorial within-subjects design with three factors (*task context*, *distance*, and *target scale*) as described in the following.

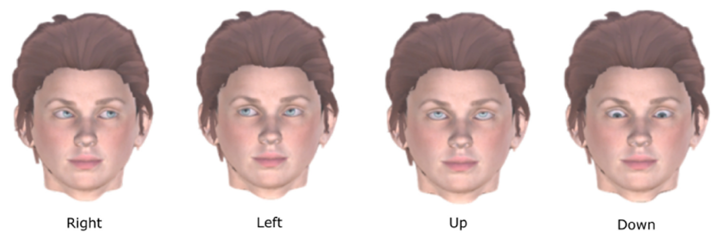
4.3.1 Study Design.

Task Context (Four Levels). Participants were instructed to perform the head scaling task in four different contexts (see Figure 9):

- *Facial expression detection*: In this task context, participants were asked to adjust the scale of the AR head while focusing on their ability to detect the AR human's facial expressions. If the head scale was too small or too large, they would not be able to detect the facial expressions. We prepared this task context considering the importance of understanding human social signals, particularly through facial expressions. We showed four facial expressions in a loop, which are described in Section 4.2.



(a) Facial Expressions



(b) Eye Gaze

(c) Faces

Fig. 9. Visual stimuli used in Experiment E2.

- *Eye gaze detection*: In this task context, participants were asked to adjust the scale of the AR human's head while focusing on their ability to detect the AR human's eye gaze. If the head scale was too small or too large, they would not be able to detect the eye gaze. We showed four gaze directions in a loop, which are described in Section 4.2.
- *Face detection*: In this task context, participants were asked to adjust the scale of the AR human's head while focusing on their ability to detect/identify the different virtual heads. If the head scale was too small or too large, they would not be able to detect the different heads. We showed three virtual heads in a loop, which are described in Section 4.2.
- *Comfort level*: In this task context, participants were asked to adjust the scale of the AR human's head while focusing on their perception of the AR human and their feeling of comfort when seeing the AR human. If the head scale is too small or too large, a feeling of "uncanniness" (or "creepiness" [55]) tends to arise, which indicates that participants would not feel comfortable interacting with such an AR human.

Distance (Three Levels). As discussed in Section 2.2, in the work by Hager and Ekman [35], every observer was able to label different human facial expressions accurately up to 45 m in the real world (not using XR), although accuracy declined as distance increased. They tested four distances (30, 35, 40, and 45 m).

Hence, in our study, we decided to start at 30 m but go beyond these distances. We tested three distances: 30, 60, and 90 m. The distances are shown in Figure 7 from a bird's eye view and from a first-person view to show the effects of these distances on the visibility of an AR human. These distances are all located well behind the focal depth of the HoloLens, which is set at 2 m.[3] As such, these conditions induce small amounts of vergence-accommodation conflict [43, 82] of 0.47, 0.48, and 0.49 diopters, respectively. These values each fall within the zone of comfort explored by Shibata et al. [71] and therefore should induce minimal visual discomfort from users. Although there is little difference in the amount of vergence-accommodation conflict induced between conditions, it is possible that users will need to compensate for somewhat blurrier imagery at further depths by choosing slightly larger scalings.

Target Scales (Three Levels). We asked the participants to perform the task for the following target scales. During each trial, participants were asked to either identify the *minimum*, *maximum*, or *ideal* head scale in the task contexts mentioned previously:

- *Minimum*: Participants were asked to identify and select the minimum head scale in the particular task context. For instance, 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 AR human 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 instance, 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.

4.3.2 Procedure. Once participants arrived, an informed consent form was provided, and they were asked to give their verbal consent to participate in the experiment. Afterward, the experimenter verbally explained the study details with the task instructions, and made sure that the participants understood the tasks through a demonstration before the experiment. Participants then donned the HoloLens 1 on which they could see the AR human's head facing toward them. In the demonstration, the participant would first make sure the virtual head is aligned with the mannequin and then adjust the scale of the virtual head for the different task contexts, which we described in Section 4.3.1. They familiarized themselves with the AR human's facial expressions, eye gazes, and the different faces that they should look out for as described in Section 4.2. They could increase or decrease the head scale with the right joystick on the XBox controller, and make the final decisions for the conditions by pressing the "A" button, which also advanced the experiment to the next trial condition.

Once participants were familiar with the interface and tasks, we started the experimental trials. For detecting facial expressions, the AR human exhibited four expressions in an infinite loop—happy, angry, surprised, and skeptical (see Figure 9). For detecting eye gaze, the AR human exhibited four gaze directions in an infinite loop—left, right, up, and down. For detecting virtual faces, the virtual head alternated between three different 3D character heads. To make sure that the participants understood what they had to do in each trial, the current target scale and task context were shown to participants via AR labels on the HoloLens.

We randomized the distances per participant. We first placed the mannequin at the target distance and confirmed with the participants that the AR human head was well aligned. All task contexts for that distance were then tested in randomized order. For each distance and task context, the target scales (minimum, maximum, and

[3]https://docs.microsoft.com/en-us/windows/mixed-reality/design/comfort.

ideal) were then tested in fixed order as our previous study did not reveal any carryover effects but a sense that randomizing these three tasks would confuse participants and promote errors.

When the experiment was halfway completed, participants took a 2-minute break. After completing all conditions, they completed a post-questionnaire, assessing their demographics and prior experience, and we asked them to indicate their general perception and preference of the AR human conditions as well as the reasoning behind their answers. Finally, the experiment ended with a monetary compensation of $15.

4.4 Measures and Hypotheses E2

As described in Section 4.3, participants experienced the AR human head scaling task with 36 different conditions in terms of three factors: task context, distance, and target scale. We collected the participants' selections on the AR human's head scale during the tasks to investigate how the three factors influence the participants' selection of head scales. For all task contexts, we expected to find minimum and maximum thresholds showing a range of head scales over which AR human head scales can be varied. Related to our research questions in Section 1, the hypotheses we established were as follows:

- *H1*: For the facial expressions, eye gaze, and face detection task contexts, participants will increase the AR human's head scale as the distance to the AR human increases.

We further considered that due to the size of facial features, it would likely be the easiest task to detect entire faces, followed by facial expressions, and then gaze directions, which we hypothesized would be reflected in the results for the head scale minima:

- *H2*: For the minimum target scale, participants will set the AR human's head scales at the lowest level for the face detection task, at a larger level for the facial expression detection task, and at the largest level for the eye gaze task.

If this hypothesis holds true, the results would indicate that different head scales may be chosen depending on different application demands. For instance, those applications that only need their users to be able to distinguish faces may use smaller head scales than those applications that require their users to be able to read facial expressions and/or gaze directions.

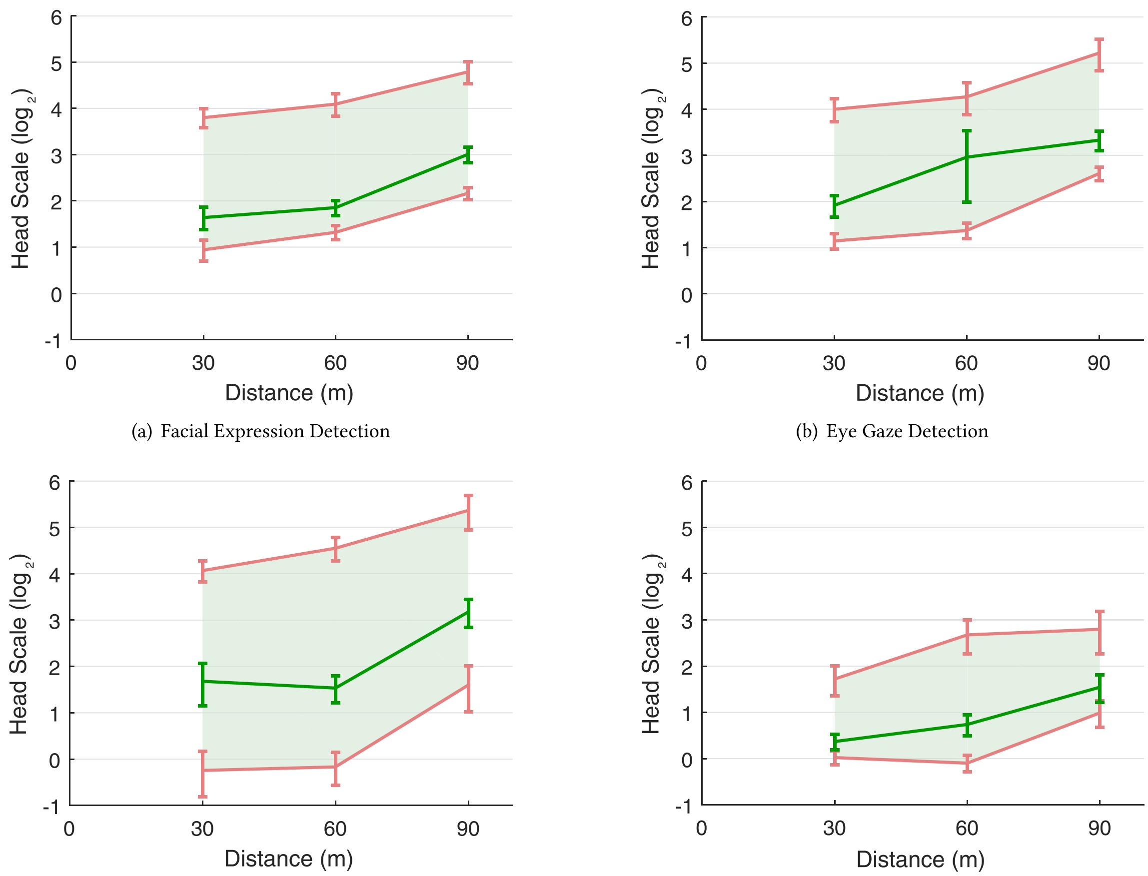
Further, we expected to see that the ideal AR head scales in terms of comfort would remain at the level of a normal human head:

- *H3*: For the comfort level task context, participants will maintain the AR human's head scale at the original scale even if the distance between the AR human and themselves increases.

Although Experiment E1 indicated that participants generally preferred slightly bigger heads in VR, we assumed that this would not carry over to this AR context, e.g., due to the higher number of real-world size cues in the physical outdoor environment, among other factors.

4.5 Results E2

Our pooled results are shown in Figure 10. The figure shows the results when participants were tasked with identifying head scales so that they could detect the AR human's (a) facial expressions, (b) eye gaze, (c) virtual faces, and (d) a comfortable range of head scales. The x-axes show the distances to the AR human. The y-axes show the head scales set by the participants; the results are presented on a logarithmic (base 2) scale on that axis. As in Experiment E1 (refer Section 3), we only used the log scale for presentation purposes and did not use the log transform for our statistical analysis. 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.



(a) Facial Expression Detection
(b) Eye Gaze Detection
(c) Face Detection
(d) Comfort Level

Fig. 10. Results of the four experimental tasks. The x-axes show the distances to the AR human; 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.

We analyzed the responses with repeated-measures ANOVAs and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. As in Experiment E1, all assumptions of the parametric statistical tests were confirmed or corrected for. We only report the significant effects.

4.5.1 Facial Expression Detection.

Distance. We found a significant main effect of *distance* on head scales, $F(1.88, 35.73) = 36.22$, $p < 0.001$, $\eta_p^2 = 0.74$. Post hoc tests showed that all pairwise comparisons between target scales were significant ($p < 0.05$) except between 30 and 60 m.

Target Scales. We found a significant main effect of *target scales* on head scales, $F(1.52, 28.85) = 185.92$, $p < 0.001$, $\eta_p^2 = 0.91$. Post hoc tests showed that all pairwise comparisons between target scales were significant.

Distance and Target Scales. We found a significant main effect of *distance* on head scales for the *minimum* task, $F(1.48, 28.06) = 43.99$, $p < 0.001$, $\eta_p^2 = 0.70$, for the *maximum* task, $F(1.89, 35.97) = 7.44$, $p < 0.001$, $\eta_p^2 = 0.43$, and for the *ideal* task, $F(1.7, 32.3) = 32.42$, $p < 0.001$, $\eta_p^2 = 0.63$. Post hoc tests for the three target scales showed that all pairwise comparisons were significant, except between distances 30 and 60 ms, and between 60 and 90 m for maximum target scale.

4.5.2 Eye Gaze Detection.

Distance. We found a significant main effect of *distance* on head scales, $F(1.68, 31.84) = 27.80$, $p < 0.001$, $\eta_p^2 = 0.60$. Post hoc tests showed that all pairwise comparisons between target scales were significant except between 30 and 60 m.

Target Scales. We found a significant main effect of *target scales* on head scales, $F(1.56, 29.61) = 94.31$, $p < 0.001$, $\eta_p^2 = 0.83$. Post hoc tests showed that all pairwise comparisons between target scales were significant.

Distance and Target Scales. We found a significant main effect of *distance* on head scales for the *minimum* task, $F(1.79, 34.04) = 45.20$, $p < 0.001$, $\eta_p^2 = 0.70$, for the *maximum* task, $F(1.64, 31.13) = 15.29$, $p < 0.001$, $\eta_p^2 = 0.45$, and for the *ideal* task, $F(1.60, 30.43) = 14.50$, $p < 0.001$, $\eta_p^2 = 0.43$. Post hoc tests for the three target scales showed significant pairwise comparisons, except between 30 and 60 m.

4.5.3 Face Detection.

Distance. We found a significant main effect of *distance* on head scales, $F(1.68, 31.99) = 53.29$, $p < 0.001$, $\eta_p^2 = 0.66$. Post hoc tests showed that all pairwise comparisons between target scales were significant except between 30 and 60 m.

Target Scales. We found a significant main effect of *target scales* on head scales, $F(1.71, 32.54) = 138.02$, $p < 0.001$, $\eta_p^2 = 0.88$. Post hoc tests showed that all pairwise comparisons between target scales were significant.

Distance and Target Scales. We found a significant main effect of *distance* on head scales for the *minimum* task, $F(1.85, 35.13) = 18.16$, $p < 0.001$, $\eta_p^2 = 0.49$, for the *maximum* task, $F(1.79, 34) = 7.44$, $p = 0.003$, $\eta_p^2 = 0.28$, and for the *ideal* task, $F(1.77, 33.59) = 15.67$, $p < 0.001$, $\eta_p^2 = 0.45$. Post hoc tests for the three target scales showed that all pairwise comparisons were significant, except between 30 and 60 m.

4.5.4 Comfort Level.

Distance. We found a significant main effect of *distance* on head scales, $F(1.75, 33.19) = 5.73$, $p = 0.01$, $\eta_p^2 = 0.23$. Post hoc tests showed that the only pairwise comparison that was significant was between 30 and 90 m.

Target Scales. We found a significant main effect of *target scales* on head scales, $F(1.28, 24.25) = 47.73$, $p < 0.001$, $\eta_p^2 = 0.72$. Post hoc tests showed that all pairwise comparisons between target scales were significant.

Distance and Target Scales. We found a significant main effect of *distance* on head scales for the *minimum* task, $F(1.60, 30.44) = 7.63$, $p = 0.004$, $\eta_p^2 = 0.29$. Post hoc tests showed that all pairwise comparisons were not significant.

We found no significant main effect of *distance* on head scales for the *maximum* task, $F(1.3, 24.6) = 2.8$, $p = 0.09$, $\eta_p^2 = 0.13$.

We found a significant main effect of *distance* on head scales for the *ideal* task, $F(1.95, 37.1) = 8.23$, $p = 0.001$, $\eta_p^2 = 0.30$. Post hoc tests showed that the only pairwise comparison that was significant was between 30 and 90 m.

4.6 Discussion and Limitations E2

In this section, we summarize the main findings and discuss implications for the use of “big head” humans in AR, while also addressing qualitative feedback we received from our participants, and limitations of our experiment.

4.6.1 AR Human Head Scales Increased over Distance for Detection Tasks. Overall, for the facial expressions detection, eye gaze detection, and face detection tasks, in line with our Hypothesis H1, we found significant effects that participants increased the AR human's head scale when the human was moved farther away. When we asked the participants in our post-questionnaire about their reasoning behind their selected head scales at different distances, some of them stated (non-verbatim reproduction; corrected for language):

> *P9*: "When the human was far away, I was forced to choose a bigger head scale, so that I can see it better."
> *P10*: "[...] when the virtual human was far away, I choose the head size a little bigger so that I can read the expressions on the face."
> *P19*: "I wanted to see the human's facial expression, so the size had to be larger."

Our participants shared the common strategy to make the heads bigger if they were otherwise too small to make out sufficient facial features. By scaling up the heads to different degrees, they were able to recover the facial features that deteriorated over distance.

Practitioners aiming to create a similar user experience with the *Big Head* technique in AR may choose to go with the ideal head scales and their relative progression for larger distances we observed in this experiment; however, as discussed earlier, the specific magnitudes in our experiment are dependent on the OST-HMD and may depend on the stimuli and outdoor lighting conditions.

Facial Expressions Detection. From Figure 10(a), based on the *minimum* head scale results, we can see that participants chose a larger head than normal at 30 m and their head scales increased as the distance increased. Without up-scaling the AR human's head, the facial expressions were not discernible even at our shortest distance of 30 m. In comparison, normal human vision limits our ability to detect facial expressions to about 45 m [35].

For the *ideal* head scales, participants adjusted the head scales closer to their minimum selections rather than their maximum selections. However the slope of ideal head scales increased more than that of the minimum head scales from 60 to 90 m. One participant explained this by stating the following:

> *P17*: "[...] to differentiate the slightest movement of expressions without straining my eyes too hard."

The results from both *ideal* and *minimum* head scales indicate that the *Big Head* technique is effective for scaling AR human heads up to perceive facial expressions.

Eye Gaze Detection. As can be seen in Figure 10(b), in line with Hypothesis H2, participants selected larger heads for the eye gaze detection and facial expressions detection tasks compared to the face detection task for the *minimum* and *ideal* target scales, and from our post-questionnaire we found that eye gaze was rated the toughest detection task followed by facial expressions and face detection. We can also see that participants selected the largest head scale at 90 m among all other tasks. Most participants' reasoning was the size of the details, since eye gaze movements were the smallest compared to facial expressions and facial features in the face detection task (non-verbatim reproduction; corrected for language):

> *P3*: "Eyes were hard to detect since what I am trying to recognize is smaller."
> *P18*: "Eye gaze is the smallest part amongst the three, so it was the toughest one."

Another interesting possible reason stated by P1 could be the color contrast of the eye:

> *P1*: "I think the eye gaze was the most difficult. Partly because it is a very small feature and partly because the color of the pupil was slightly difficult to discern."

As discussed in Section 2.4, the additive light model of OST-HMDs makes it difficult to perceive dark pupils as the physical background is visible through these rendered screen areas [60]. Specifically, if the background is bright, it can wash out the pupil.

Face Detection. In line with our Hypothesis H2, the face detection task was rated the easiest among the three detection tasks on our post-questionnaire. From Figure 10(c), we can even see values below normal head scales for the *minimum* task at 30 and 60 m. This means that when selecting the *minimum* head scales to detect the different faces, participants reduced the head scales even below their normal size. However, at 90 m, we can see a steep increase in head scales. We can say that the minimum threshold distance beyond which participants have to increase head scale to detect faces is between 60 and 90 m.

Some common reasons given for this apparent ease to detect faces by our participants were hairstyles, eye socket depth, and styles of ears, among others. A few example comments by our the participants were the following (non-verbatim reproduction; corrected for language):

> *P1*: "Virtual faces were the easiest because all the features were changing so I didn't need to look for a specific feature."
> *P16*: "Virtual faces were easiest because all of them had different features and different hairstyles. It made it easier to recognize them at any distance."
> *P20*: "For virtual faces it was easy to recognize from far away due to their hairstyle, eyebrow style, eye socket depth. It allowed me to see more of the face from far away and even closer."

This detection task showed the largest range between *minimum* and *maximum* head scales. Although the *maximum* head scales for this task are consistent with the other detection tasks, it has the lowest *minimum* head scales among all tasks.

4.6.2 AR Human Head Scales Increased over Distance for Comfortable Proportions. In Hypothesis H3, we predicted that participants would keep the head scales constant over the three distances for the comfort level task, assuming that normal human head scales will appear the most comfortable independent of distance. We based this hypothesis on the *Uncanny Valley* [55] effect, where a disproportionate AR human would appear "uncanny" and uncomfortable. However, again, similar to Experiment E1, our results do not match this prediction. In this task, where we asked participants to select the AR human's head scale and proportions that feel the most comfortable to them, we can see an increase in head scales as distance increased in Figure 10. We believe this may be related to the lighting conditions, low resolution, and additive light model of the OST-HMD (see Section 2.4) that may have caused the AR human's head to appear less than natural and comfortable to start with, which could be "lessened" by increasing their scale.

Experiment E1 found a similar effect, although with a large difference in magnitude. For ideal head scales, participants in Experiment E1 increased the VR heads to twice their size beyond 10 m distance, whereas participants in Experiment E2 increased the AR heads to twice their size only after 60 to 90 m. Our results show that the participants' head scale judgments are influenced by the factors that differed between our Experiments E1 and E2, suggesting that future research may try to isolate the corresponding factors in controlled comparative studies.

Our results also show a large range from *minimum* to *maximum* head scales, giving developers room to explore different head scales on AR humans without participants feeling uncomfortable. Specifically, the *minimum* head scales for the three detection tasks for facial expressions, eye gaze, and faces fall below the *maximum* head scales indicated for the range of comfortable head scales.

4.6.3 Applications and Guidelines. We asked our participants in our post-questionnaire where they believe this kind of scaling method in AR would be most useful, and we received some very interesting answers. The most common application proposed by our participants was the identification of friends and family in a crowded area. Another application proposed was directed at finding or identifying criminals for the police. One participant also suggested that this would be a fun game mechanic for AR games. A few statements from our participants were the following (non-verbatim reproduction; corrected for language):

> *P10*: "It could be used by the Army to identify the facial expression of the enemy."
> *P16*: "Identifying suspects or any person of interest in crowds, beneficial for people with vision problems."
> *P19*: "[...] finding your loved one in a crowd or a theme park or to look for a lost child in a crowd or to look for a possible suspect/criminal who has committed a felony."

One participant expressed a potential concern about this technique in AR that we found particularly interesting. This participant talked about privacy issues that may arise due to this technique, if it involves scaling of other peoples' heads without them being aware of it. Another interesting concern mentioned by the same participant was about information overload, as bigger heads may convey more information than one would normally receive:

> *P12*: "Magnifying heads feels a little like an invasion of privacy if not everyone is aware that it is happening. Like, I shouldn't be able to see if someone who is far away from me is sad. If we are friends, we would move closer together to interact and share those emotions. If we are not friends, I would feel like I know too much about them. More than privacy, any place with many people in it would be overwhelming. Seeing a lot of facial information (eye gazes, expressions, etc.) would cause an information overload for me. I would just want to tell people apart, but not have to think about where they are looking or how they are feeling."

In particular, for real-time high-resolution camera-based magnifications of other people's heads, such as supported by the AR prototype presented by Choudhary et al. [15], those are important considerations to keep in mind for future work in this direction.

4.6.4 Limitations and Future Work. The *Big Head* technique has shown some interesting effects of AR human head scales on the perception of facial expressions, eye gaze, faces, and comfort at different distances in an outdoor AR setting. However, there are also limitations of our current work, which can lead to additional study ideas that may be investigated in future work.

First, since our experiment was set in an outdoor environment, sunlight exposure, shadows, and so forth could not be fully controlled. To fully understand these effects, we believe it will be important to *triangulate* [67] our results in future studies in different environments, from urban streets to empty fields or the desert.

Second, our experiment only used female humans with a specific set of facial expressions and basic eye gaze movements, which are representative for a subset of human faces and facial features, but of course we make no claims that these results will hold true for all situations that may occur in practical applications. We refer to several works on facial expressions [23–25, 58, 72], different rendering styles for humans [48, 83], and appearances based on gender, age, race, skin color, and so on [22, 49, 84] that may be important topics for future work.

Third, our experiment did not give participants any time limit for the detections. This might be an interesting topic to explore in time-critical situations, where a time limit may affect their accuracy or cognitive load [20, 52].

Fourth, we used the HoloLens 1 OST-HMD, and our results are dependent on this device. Different displays, such as the HoloLens 2, provide a resolution of 2048 × 1080 pixels per eye at a 54-degree field of view. Future cross-device studies may determine how much of the variance in the head scales we may be able to explain by the characteristics of these devices, such as their resolution or contrast.

Fifth, our experiment used a static spatial configuration, whereas other tasks may involve linear and/or rotational movements. Future work may study similar detection tasks at different rotations of the virtual heads and at different motion states.

Finally, we propose looking into possible effects of scaling other body parts. Possible scenarios could be scaling one's hands for conveying gestures over longer distances, or a dynamic scaling system that may scale body parts based the type of non-verbal signal one tries to convey. Scenarios may also involve verbal communication, such as up-scaled heads for lip reading applications.

5 GENERAL DISCUSSION

In this section, we summarize some of our main findings from both experiments and discuss possible use cases for the *Big Head* technique.

Our two experiments were designed to explore a broad range of XR scenarios from indoor VR to outdoor AR. Even though VR and AR are different display technologies, we found some commonalities along with differences in our main findings from both experiments.

First, distance is an integral factor for humans to discern facial cues from a virtual human, whether we use VR or AR. This falls in line with the results of Hager and Ekman [35], where humans were able to discern facial expressions (in real life) up to 30 to 45 m, although their accuracy declined as distance increased.

In Experiment E1 (see Section 3), we investigated distance as a factor to discern facial expressions in VR and to test the *Big Head* technique for fully virtual humans. Participants were able to discern facial expressions up to only 3 m. Beyond 3 m, they had to up-scale the virtual human's head to sufficiently make out facial details (see Figure 3). Although there may be more factors, we believe the VR-HMD's limited screen resolution was the primary factor that took social cues away for longer distances, whereas increasing the head scale could "bring them back."

In Experiment E2 (see Section 4), we investigated human perception of facial features in an outdoor setting with an AR-HMD and tested the *Big Head* technique for partial (hybrid) virtual humans. We generally observed similar correlations, although with largely different magnitudes. Participants up-scaled heads to twice their size after 10 m in Experiment E1; however, they did that only beyond 60 to 90 m in Experiment E2. This could have been heavily influenced by the higher AR display resolution but may also have been affected by the outdoor lighting conditions and the additive light model of the OST-HMD, warranting future work to isolate the contributing factors.

Second, participants felt overall more comfortable with head sizes that were up-scaled over their natural size at longer distances. This was observed in both VR and outdoor AR. Although identifying the causes of this preference of slightly bigger-than-normal heads needs further exploration, it is an interesting and promising observation that application developers in XR are not limited by true-to-scale human proportions. Further, if slightly bigger heads are not perceived as entirely uncomfortable, the *Big Head* technique could indeed be a stepping stone for the near future in the field of XR before the display resolution of commercial devices can be increased to reach the visual acuity of the human eye.

Third, in VR (see Section 3), we 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 upward from the neck. This is specifically interesting with hybrid mixed reality applications—for example, if we use computer vision algorithms to detect and magnify heads in real time with real humans [15], this real-world dynamic AR scaling could be either fixed eye height, fixed neck height, or more. Since fixed eye height was preferred (see Section 3.5), we would suggest developers to use this technique over fixed neck height; however, future work is needed to gain a better understanding about a wider range of possible scaling methods.

Last, in outdoor AR (see Section 4), among three facial cue detection tasks (facial expressions, eye gaze, and face detection), detecting eye gaze was rated the toughest, followed by facial expressions and finally face detection. This can be observed in the lower minimum head scales selected by our participants for the face detection task (see Figure 10). In applications where facial details are important to understand and discern, it is important to understand that different facial details, like facial expressions or eye gaze, have different levels of difficulty.

Potential Applications and Use Cases of the Big Head Technique. For both experiments, the *Big Head* technique has shown to be useful in enhancing facial information at longer distances. After Experiment E2 (see Section 4), we asked participants what possible applications or use cases the *Big Head* technique may have. We received some very interesting feedback in Section 4.6.3. The most common answers were about identification of humans in XR situations where it may be important, such as identifying suspects or friends in crowds, and it could also benefit people with vision problems. This technique could be used along with computer vision algorithms to detect, segment, and magnify certain faces in real time. One may then display the magnified head over the location of the real head on an AR HMD or even in 2D video conferencing [12, 15]. Another interesting application could be in virtual events, like virtual conferences or meetings. In such events, if speakers are trying to reach out to their audience, the *Big Head* technique could help express their facial information better, and it could even shift the audience's attention to the person with the "biggest head."

Alternative Methods. There are other alternative methods to enhance facial information specifically when the source is at long distances. Another technique could be showing an icon above the avatar. This icon could have a larger image of their face, or it could have immediate information about the facial information, like their mood or their facial expression. Another alternate could be using the mini-me approach [63]. A mini-me or a small 3D visualization of the avatar up close to the user could provide similar enhanced facial information. Last, one could walk closer to the person or teleport the person closer to the user. Teleportation may work with virtual humans; however, this would not be possible with a real human with AR. These techniques (including the *Big Head* technique) may be more useful in certain situations over others. To fully understand the differences between these techniques, further research is required.

6 CONCLUSION

In this article, we present two human-subject experiments in indoor VR and outdoor AR, where we investigated the impact of an increased or decreased virtual human's head scale on participants' ability to perceive facial details, and to assess their sense of comfort. Our results from both experiments suggest that as distance increased between participants and virtual humans, they increased the virtual human's head scale to improve their perception of facial details. This indicates that the *Big Head* technique can be an effective means for recovering a user's perception of a virtual human's facial features, which deteriorates over longer distances. When we tested participants' sense of comfort with respect to the *Big Head* technique, our results even indicated that they felt more comfortable with head sizes that were slightly up-scaled over their natural size at longer distances. We discussed potential explanations, implications, and applications to guide practitioners aiming to leverage this technique in VR and AR, and we discussed limitations of our experiments and future avenues for research.

REFERENCES

[1] Matt Adcock and Chris Gunn. 2010. Annotating with 'sticky' light for remote guidance. In *ACM SIGGRAPH ASIA 2010 Posters*. 1.

[2] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. 2012. Designing effective gaze mechanisms for virtual agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 705–714.

[3] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. 2012. A head-eye coordination model for animating gaze shifts of virtual characters. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*. 1–6.

[4] Ferran Argelaguet Sanz, Anne Helene Olivier, Gerd Bruder, Julien Pettre, and Anatole Lecuyer. 2015. Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments. In *Proceedings of IEEE Virtual Reality (VR'15)*. 75–80.

[5] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin* 29, 7 (2003), 819–833.

[6] Inessa Bekerman, Paul Gottlieb, and Michael Vaiman. 2014. Variations in eyeball diameters of the healthy adults. *Journal of Ophthalmology* 2014 (2014), 503645.

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

[8] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook—Moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications* 21, 3 (2001), 6–8.

[9] Philipp Breuss-Schneeweis. 2016. "The speaking celt" augmented reality avatars guide through a museum–case study. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 1484–1491.

[10] Gerd Bruder, Andreas Pusch, and Frank Steinicke. 2012. 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'12)*. 111–118.

[11] Angelo Cafaro, Hannes Högni Vilhjálmsson, Timothy Bickmore, Dirk Heylen, Kamilla Rún Jóhannsdóttir, and Gunnar Steinn Valgarðsson. 2012. First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *Proceedings of the International Conference on Intelligent Virtual Agents*. 67–80.
[12] Zubin Choudhary, Gerd Bruder, and Gregory Welch. 2021. Scaled user embodied representations in virtual and augmented reality. In *Proceedings of the Workshop on User-Embodied Interaction in Virtual Reality (UIVR'21)*.
[13] Zubin Choudhary, Matthew Gottsacker, Kangsoo Kim, Ryan Schubert, Jeanine Stefanucci, Gerd Bruder, and Gregory F. Welch. 2021. Revisiting distance perception with scaled embodied cues in social virtual reality. In *Proceedings of IEEE Virtual Reality (VR'21)*. 1–10.
[14] Zubin Choudhary, Kangsoo Kim, Ryan Schubert, Gerd Bruder, and Gregory F. Welch. 2020. Virtual big heads: Analysis of human perception and comfort of head scales in social virtual reality. In *Proceedings of IEEE Virtual Reality (VR'20)*. 425–433.
[15] Zubin Choudhary, Jesus Ugarte, Gerd Bruder, and Greg Welch. 2021. Real-time magnification in augmented reality. In *Proceedings of the ACM Symposium on Spatial User Interaction (SUI'21)*. 1–2.
[16] Alex Colburn, Michael F. Cohen, and Steven Drucker. 2000. *The Role of Eye Gaze in Avatar Mediated Conversational Interfaces*. Technical Report. Microsoft Research.
[17] Matthew Christopher Davis, Dang D. Can, Jonathan Pindrik, Brandon G. Rocque, and James M. Johnston. 2016. Virtual interactive presence in global surgical education: International collaboration through augmented reality. *World Neurosurgery* 86 (2016), 103–111.
[18] Celso M. de Melo, Peter Carnevale, and Jonathan Gratch. 2012. 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*. 53–66.
[19] Carl F. DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. 2002. 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*. 321–326.
[20] Shichuan Du and Aleix M. Martinez. 2013. Wait, are you sad or angry? Large exposure time differences required for the categorization of facial expressions of emotion. *Journal of Vision* 13, 4 (2013), 13.
[21] Nicolas Ducheneaut, Ming-Hui Wen, Nicholas Yee, and Greg Wadley. 2009. 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*. 1151–1160.
[22] Robert Andrew Dunn and Rosanna E. Guadagno. 2012. My avatar and me—Gender and personality predictors of avatar-self discrepancy. *Computers in Human Behavior* 28, 1 (2012), 97–106.
[23] Paul Ekman. 1989. The argument and evidence about universals in facial expressions. In *Handbook of Social Psychophysiology*. Wiley, 143–164.
[24] Paul Ekman. 1993. Facial expression and emotion. *American Psychologist* 48, 4 (1993), 384.
[25] Rosenberg Ekman. 1997. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press.
[26] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Gregory Welch. 2020. Exploring the limitations of environment lighting on optical see-through head-mounted displays. In *Proceedings of the ACM Conference on Spatial User Interfaces (SUI'20)*. 1–8.
[27] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Gregory F. Welch. 2020. A review of visual perception research in optical see-through augmented reality. In *Proceedings of the International Conference on Artificial Reality and Telexistence and the Eurographics Symposium on Virtual Environments*. 1–9.
[28] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Gregory F. Welch. 2020. Effects of dark mode graphics on visual acuity and fatigue with virtual reality head-mounted displays. In *Proceedings of IEEE Virtual Reality (VR'20)*. 434–442.
[29] Austin Erickson, Kangsoo Kim, Alexis Lambert, Gerd Bruder, Michael P. Browne, and Greg Welch. 2021. An extended analysis on the benefits of dark mode user interfaces in optical see-through head-mounted displays. *ACM Transactions on Applied Perception* 18, 3 (2021), 1–22.
[30] Antonio Fernández-Caballero, Elena Navarro, Patricia Fernández-Sotos, Pascual González, Jorge J. Ricarte, José M. Latorre, and Roberto Rodriguez-Jimenez. 2017. Human-avatar symbiosis for the treatment of auditory verbal hallucinations in schizophrenia through virtual/augmented reality and brain-computer interfaces. *Frontiers in Neuroinformatics* 11 (2017), 64.
[31] Joseph Gabbard, J. Edward Swan, Jason Zedlitz, and Woodrow W. Winchester. 2010. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In *Proceedings of IEEE Virtual Reality (VR'10)*. 79–86.
[32] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. 2003. 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*. 529–536.
[33] Giant Bomb. 2021. Big Head Mode. Retrieved November 12, 2022 from https://www.giantbomb.com/big-head-mode/3015-403/.
[34] Sinem Guven, Ohan Oda, Mark Podlaseck, Harry Stavropoulos, Sai Kolluri, and Gopal Pingali. 2009. Social mobile augmented reality for retail. In *Proceedings of the 2009 IEEE International Conference on Pervasive Computing and Communications*. IEEE, Los Alamitos, CA, 1–3.
[35] Joseph C. Hager and Paul Ekman. 1979. Long-distance of transmission of facial affect signals. *Ethology and Sociobiology* 1, 1 (1979), 77–82.
[36] Edward Hall. 1963. A system for the notation of proxemic behavior. *American Anthropologist* 65, 5 (1963), 1003–1026.

[37] Edward Hall. 1969. *The Hidden Dimension: Man's Use of Space in Public and in Private*. Anchor Books.
[38] Tina Iachini, Yann Coello, Francesca Frassinetti, and Gennaro Ruggiero. 2014. Body space in social interactions: A comparison of reaching and comfort distance in immersive virtual reality. *PloS One* 9 (2014), e111511.
[39] Daniel P. Kennedy, Jan Gläscher, J. Michael Tyszka, and Ralph Adolphs. 2009. Personal space regulation by the human amygdala. *Nature Neuroscience* 12 (2009), 1226–1227.
[40] Kangsoo Kim, Mark Billinghurst, Gerd Bruder, Henry Been-Lirn Duh, and Gregory F. Welch. 2018. Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2947–2962.
[41] Kangsoo Kim, Nahal Norouzi, Tiffany Losekamp, Gerd Bruder, Mindi Anderson, and Gregory Welch. 2019. 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'19)*. 17–177.
[42] Kiyoshi Kiyokawa, Haruo Takemura, and Naokazu Yokoya. 1999. 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*. 48–53.
[43] Gregory Kramida. 2015. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics* 22, 7 (2015), 1912–1931.
[44] Veronika Krauß, Alexander Boden, Leif Oppermann, and René Reiners. 2021. Current practices, challenges, and design implications for collaborative AR/VR application development. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.
[45] Bernard Kress, Ehsan Saeedi, and Vincent Brac-de-la Perriere. 2014. The segmentation of the HMD market: Optics for smart glasses, smart eyewear, AR and VR headsets. In *Photonics Applications for Aviation, Aerospace, Commercial, and Harsh Environments V*, Vol. 9202. International Society for Optics and Photonics, 92020D.
[46] Bernard C. Kress. 2019. Digital optical elements and technologies (EDO19): Applications to AR/VR/MR. In *Digital Optical Technologies 2019*, Vol. 11062. International Society for Optics and Photonics, 1106222.
[47] Ernst Kruijff, J. Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR'10)*. 3–12.
[48] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 1–10.
[49] Jong-Eun Roselyn Lee, Clifford I. Nass, and Jeremy N. Bailenson. 2014. Does the mask govern the mind?: Effects of arbitrary gender representation on quantitative task performance in avatar-represented virtual groups. *Cyberpsychology, Behavior, and Social Networking* 17, 4 (2014), 248–254.
[50] Myungho Lee, Gerd Bruder, and Gregory F. Welch. 2017. Exploring the effect of vibrotactile feedback through the floor on social presence in an immersive virtual environment. In *Proceedings of IEEE Virtual Reality (VR'17)*. 105–111.
[51] Myungho Lee, Nahal Norouzi, Gerd Bruder, Pamela Wisniewski, and Gregory Welch. 2019. Mixed reality tabletop gameplay: Social interaction with a virtual human capable of physical influence. *IEEE Transactions on Visualization and Computer Graphics* 24, 8 (2019), 1–12.
[52] Michael D. Lee, Robyn L. Vast, and Marcus A. Butavicius. 2006. Face matching under time pressure and task demands. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. 1675–1680.
[53] Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics* 31, 4 (2012), 1–11.
[54] Mark McGill and Stephen A. Brewster. 2017. I am the passenger: Challenges in supporting AR/VR HMDs in-motion. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. 251–251.
[55] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.
[56] Nahal Norouzi, Gerd Bruder, Brandon Belna, Stefanie Mutter, Damla Turgut, and Gregory Welch. 2019. A systematic review of the convergence of augmented reality, intelligent virtual agents, and the Internet of Things. In *Artificial Intelligence in IoT*. Transactions on Computational Science and Computational Intelligence. Springer, 1–24.
[57] Nahal Norouzi, Austin Erickson, Kangsoo Kim, Ryan Schubert, Joseph LaViola, Gerd Bruder, and Greg Welch. 2019. Effects of shared gaze parameters on visual target identification task performance in augmented reality. In *Proceedings of the ACM Symposium on Spatial User Interaction (SUI'19)*. 1–11.
[58] Soo Youn Oh, Jeremy Bailenson, Nicole Krämer, and Benjamin Li. 2016. Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PloS One* 11, 9 (2016), e0161794.
[59] Riccardo Palmarini, John Ahmet Erkoyuncu, Rajkumar Roy, and Hosein Torabmostaedi. 2018. A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing* 49 (2018), 215–228.
[60] Tabitha C. Peck, Jessica J. Good, Austin Erickson, Isaac Bynum, and Gerd Bruder. 2022. Effects of transparency on perceived humanness: Implications for rendering skin tones using optical see-through displays. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2179–2189.

[61] Catlin Pidel and Philipp Ackermann. 2020. Collaboration in virtual and augmented reality: A systematic overview. In *Proceedings of the International Conference on Augmented Reality, Virtual Reality, and Computer Graphics*. 141–156.

[62] Thammathip Piumsomboon, Arindam Day, Barrett Ens, Youngho Lee, Gun Lee, and Mark Billinghurst. 2017. Exploring enhancements for remote mixed reality collaboration. In *Proceedings of ACM SIGGRAPH Asia Mobile Graphics and Interactive Applications*. 16.

[63] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[64] Thammathip Piumsomboon, Youngho Lee, Gun A. Lee, Arindam Dey, and Mark Billinghurst. 2017. Empathic mixed reality: Sharing what you feel and interacting with what you see. In *Proceedings of the IEEE International Symposium on Ubiquitous Virtual Reality*. 38–41.

[65] Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. 2013. The perception of egocentric distances in virtual environments—A review. *ACM Computing Surveys* 46, 2 (2013), Article 23, 40 pages.

[66] Kerstin Ruhland, Christopher E. Peters, Sean Andrist, Jeremy B. Badler, Norman I. Badler, Michael Gleicher, Bilge Mutlu, and Rachel McDonnell. 2015. A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. *Computer Graphics Forum* 34 (2015), 299–326.

[67] Neil J. Salkind. 2010. Triangulation. In *Encyclopedia of Research Design*. SAGE, Thousand Oaks, CA, 1.

[68] Franziska Schrammel, Sebastian Pannasch, Sven-Thomas Graupner, Andreas Mojzisch, and Boris M. Velichkovsky. 2009. Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. *Psychophysiology* 46, 5 (2009), 922–931.

[69] Eva-Maria Seidel, Ute Habel, Michaela Kirschner, Ruben C. Gur, and Birgit Derntl. 2010. The impact of facial emotional expressions on behavioral tendencies in women and men. *Journal of Experimental Psychology: Human Perception and Performance* 36, 2 (2010), 500.

[70] Jun'ichiro Seyama and Ruth S. Nagayama. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence* 16, 4 (2007), 337–351.

[71] Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision* 11, 8 (2011), 11.

[72] Mel Slater, D.-P. Pertaub, and Anthony Steed. 1999. Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications* 19, 2 (1999), 6–9.

[73] Fraser W. Smith and Philippe G. Schyns. 2009. Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distances. *Psychological Science* 20, 10 (2009), 1202–1208.

[74] Harrison Jesse Smith and Michael Neff. 2018. Communication behavior in embodied virtual reality. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 1–12.

[75] Lucía Vera, Jesús Gimeno, Inmaculada Coma, and Marcos Fernández. 2011. Augmented mirror: Interactive augmented reality system based on kinect. In *Proceedings of the IFIP Conference on Human-Computer Interaction*. 483–486.

[76] Michael E. Walker, Daniel Szafir, and Irene Rae. 2019. The influence of size in augmented reality telepresence avatars. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR'19)*. IEEE, Los Alamitos, CA, 538–546.

[77] Christian Wallraven, Heinrich H. Bülthoff, Douglas W. Cunningham, Jan Fischer, and Dirk Bartz. 2007. Evaluation of real-world and computer-generated stylized facial expressions. *ACM Transactions on Applied Perception* 4, 3 (2007), 16–es.

[78] Colin Ware. 2019. *Information Visualization: Perception for Design*. Morgan Kaufmann.

[79] Gregory F. Welch, Gerd Bruder, Peter Squire, and Ryan Schubert. 2019. *Anticipating Widespread Augmented Reality: Insights from the 2018 AR Visioning Workshop*. Technical Report. University of Central Florida and Office of Naval Research.

[80] Nick Yee, Jeremy N. Bailenson, Mark Urbanek, Francis Chang, and Dan Merget. 2007. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *Cyberpsychology & Behavior* 10, 1 (2007), 115–121. https://doi.org/10.1089/cpb.2006.9984

[81] Boram Yoon, Hyung-Il Kim, Gun A. Lee, Mark Billinghurst, and Woontack Woo. 2019. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR'19)*. IEEE, Los Alamitos, CA, 547–556.

[82] Roberts Zabels, Krišs Osmanis, Mārtiņš Narels, Uģis Gertners, Ainārs Ozols, Kārlis Rūtenbergs, and Ilmārs Osmanis. 2019. AR displays: Next-generation technologies to solve the vergence–accommodation conflict. *Applied Sciences* 9, 15 (2019), 3147. https://www.mdpi.com/2076-3417/9/15/3147.

[83] Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To stylize or not to stylize? The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics* 34, 6 (2015), 1–12.

[84] Yulei Gavin Zhang, Yan Mandy Dang, Susan A. Brown, and Hsinchun Chen. 2017. Investigating the impacts of avatar gender, avatar age, and region theme on avatar physical activity in the virtual world. *Computers in Human Behavior* 68 (2017), 378–387.

Received 5 May 2022; revised 12 September 2022; accepted 2 November 2022