



Affordable But Not Cheap: A Case Study of the Effects of Two 3D-Reconstruction Methods of Virtual Humans

Andrea Bartl^{1*}, Stephan Wenninger², Erik Wolf¹, Mario Botsch² and Marc Erich Latoschik¹

¹HCI Group, University of Würzburg, Würzburg, Germany, ²Computer Graphics Group, TU Dortmund University, Dortmund, Germany

OPEN ACCESS

Edited by:

Philippe Guillotel,
InterDigital, France

Reviewed by:

Cédric Fleury,
Université Paris-Sud, France
Tapio Takala,
Aalto University, Finland

*Correspondence:

Andrea Bartl
andrea.bartl@uni-wuerzburg.de

Specialty section:

This article was submitted to
Technologies for VR,
a section of the journal
Frontiers in Virtual Reality

Received: 13 April 2021

Accepted: 08 September 2021

Published: 24 September 2021

Citation:

Bartl A, Wenninger S, Wolf E, Botsch M and Latoschik ME (2021) Affordable But Not Cheap: A Case Study of the Effects of Two 3D-Reconstruction Methods of Virtual Humans. *Front. Virtual Real.* 2:694617. doi: 10.3389/frvir.2021.694617

Realistic and lifelike 3D-reconstruction of virtual humans has various exciting and important use cases. Our and others' appearances have notable effects on ourselves and our interaction partners in virtual environments, e.g., on acceptance, preference, trust, believability, behavior (the Proteus effect), and more. Today, multiple approaches for the 3D-reconstruction of virtual humans exist. They significantly vary in terms of the degree of achievable realism, the technical complexities, and finally, the overall reconstruction costs involved. This article compares two 3D-reconstruction approaches with very different hardware requirements. The high-cost solution uses a typical complex and elaborated camera rig consisting of 94 digital single-lens reflex (DSLR) cameras. The recently developed low-cost solution uses a smartphone camera to create videos that capture multiple views of a person. Both methods use photogrammetric reconstruction and template fitting with the same template model and differ in their adaptation to the method-specific input material. Each method generates high-quality virtual humans ready to be processed, animated, and rendered by standard XR simulation and game engines such as Unreal or Unity. We compare the results of the two 3D-reconstruction methods in an immersive virtual environment against each other in a user study. Our results indicate that the virtual humans from the low-cost approach are perceived similarly to those from the high-cost approach regarding the perceived similarity to the original, human-likeness, beauty, and uncanniness, despite significant differences in the objectively measured quality. The perceived feeling of change of the own body was higher for the low-cost virtual humans. Quality differences were perceived more strongly for one's own body than for other virtual humans.

Keywords: virtual humans, 3D-reconstruction methods, avatars, agents, user study

1 INTRODUCTION

The 3D-reconstruction of virtual humans is highly relevant for a variety of use cases. While the use of realistic, lifelike virtual humans is common in many areas, e.g., video games and films, it is especially interesting for Virtual Reality (VR) applications (Gonzalez-Franco et al., 2020; Freeman and Maloney, 2021). Here, virtual humans can serve as self-avatars, embodying users by representing their real body in the virtual world (Kiltner et al., 2012). It is also possible for users to meet other

virtual humans, experiencing social situations similar to the real world (Gonzalez-Franco and Lanier, 2017).

The appearance of these virtual humans has notable effects on ourselves and our interaction partners (see, e.g., Praetorius and Görlich (2020); Ratan et al. (2020)). Previous work found realistic self-avatars used for embodiment to be superior to abstract self-avatars in terms of user acceptance (Latoschik et al., 2017). Others found personalized realistic-looking self-avatars to be even more superior, enhancing the illusion of virtual body ownership as well as the feeling of presence (Waltemate et al., 2018). Comparable interesting effects occur for other-avatars (the virtual representations of other users) and virtual agents (embodied entities controlled by artificial intelligence). For example, the appearance of virtual others impacts their perceived trustworthiness (McDonnell et al., 2012; Seymour et al., 2019), approachability (Freeman and Maloney, 2021), affinity (Seymour et al., 2019), and co-presence (Bailenson et al., 2005). Given the continuous technological advances in the reconstruction of virtual humans, research on their realism is still ongoing (Slater et al., 2020). For example, there is still debate about whether realistic-looking virtual humans are prone to facilitate the uncanny valley effect (e.g., Tinwell and Grimshaw, 2009; Lugin et al., 2015a; Kätsyri et al., 2015; Wang et al., 2015), which describes the phenomenon that close-to-real looking artificial humans sometimes strike as eerie (Mori et al., 2012; Ho and MacDorman, 2010, 2017).

Today, multiple reconstruction approaches for realistic, lifelike virtual humans exist. They significantly vary in terms of the degree of achievable realism, the technical complexities, and finally, the overall reconstruction costs involved. So far, rather complex and expensive multi-camera rigs achieve the highest quality by using high-quality image sensors, e.g., as described by Feng et al. (2017) or Achenbach et al. (2017). However, approaches for reconstructing virtual humans from input data produced by more affordable consumer hardware, e.g., single 2D images (Alldieck et al., 2019a) or smartphone videos (Ichim et al., 2015; Wenninger et al., 2020), become more popular and elaborate. Most of these low-cost approaches share the vision to make it possible for everyone to generate a digital alter ego quickly and inexpensively without a complex hardware setup. Such low-cost approaches would drastically leverage the possibilities for research, industry, and overall users of embodiment systems. Researchers can create and use personalized, realistic virtual humans in their work by simply utilizing consumer-level hardware, e.g., a 600 \$ smartphone instead of a camera rig costing tens of thousands of dollars. Smaller development teams can afford life-like virtual humans, for example, in their games and social VR applications, and users would benefit from a much more personalized experience using their realistic look-alike avatars.

Recent work suggests that some of these low-cost approaches can compete with the more elaborate, high-cost approaches (Wenninger et al., 2020). However, the comparisons so far have focused primarily on objective criteria. Equally important, though, is the subjective perception of these low-cost virtual humans. Hence, in this work, we address the following research questions:

RQ₁ Can low-cost approaches for generating realistic virtual humans keep up with high-cost solutions regarding their perception by users in embodied VR?

RQ₂ Is the quality difference more noticeable for the own virtual body compared to the virtual body of others?

Contribution: For the investigation of our research questions we conducted a user study to compare a low- and a high-cost 3D-reconstruction approach for virtual humans. For each category, we chose a state-of-the-art representative that produces 1) realistically looking and 2) ready-to-animate virtual humans 3) in a time frame that is compliant with common study procedures, i.e., within minutes. Both methods build on recent advancements in photogrammetric reconstruction. They are tailored to different input material and vary heavily in their hardware requirements. One method uses a complex, elaborate, and expensive camera rig including 94 DSLR cameras to capture images of a person (Achenbach et al., 2017). The second method uses a simple smartphone camera to capture videos including multiple views of a person (Wenninger et al., 2020). We scanned participants by both methods. Then they embodied the resulting self-avatars in an interactive, immersive virtual environment and encountered pre-scanned virtual others of both reconstruction methods. We report on the sense of embodiment for the self-avatars and the perceived similarity, uncanniness, and preference for both the self-avatars and the virtual others. We further look at objective differences between the two methods and investigate whether these differences are more noteworthy for the self-avatar than someone else's body. Our results indicate that the avatars from the low-cost approach are perceived similarly to the avatars from the high-cost approach. This is remarkable since the quality differed significantly on an objective level. The perceived change of the own body was more significant for the low-cost avatars than for the high-cost avatars. The quality differences were more noticeable for the own than for other virtual bodies.

2 RELATED WORK

2.1 Perception of Virtual Humans

Virtual humans are part of a great variety of applications. They serve as avatars (representations of real people in digital worlds), virtual trainers, assistants, companions, game characters, and many more. Often, developers strive to make them as realistic as possible. The perceived realism of virtual humans depends on their appearance and their behavior (Magenat-Thalmann and Thalmann, 2005; Steed and Schroeder, 2015). While we acknowledge the importance of behavioral realism, our work focuses on the appearance of virtual humans. Our appearance and the appearance of others in a virtual environment have notable effects on our perception (Hudson and Hurter, 2016; Freeman and Maloney, 2021).

2.1.1 The Own Virtual Appearance

When it comes to using virtual humans as avatars, i.e., digital representations of persons in a virtual world, the Proteus effect (Yee and Bailenson, 2006, 2007) is a prominent research topic. It

describes the phenomenon that the avatar appearance can influence users' attitudes and behavior based on stereotypical beliefs. For example, in previous research, participants who embodied a child associated more child-like attributes with themselves (Banakou et al., 2013), attractive avatars increased intimacy (Yee and Bailenson, 2007), strong-looking avatars improved physical performance (Kocur et al., 2020), and taller avatars led to more confidence (Yee and Bailenson, 2007). Wolf et al. (2021) recently showed that the embodiment of an avatar can potentially alter its body weight perception relating to the user's body weight.

For many VR applications, the Proteus effect is desirable. Users can slip into a body with different size, shape, look, age or gender, enabling experiences one could not easily create in real life. Exploiting this effect potentially even helps to reduce negative attitudes, such as racial bias (Peck et al., 2013; Banakou et al., 2016), negative stereotypical beliefs about older people (Yee and Bailenson, 2006) or misconceptions of the own body image (Döllinger et al., 2019). It could also promote positive attitudes and behavior, e.g., motivation to exercise (Peña et al., 2016). However, what if the use case requires the users just to be themselves? For example, experiments often assume a user's unbiased evaluation without taking the potential bias of the virtual body into account. Other exemplary scenarios might focus on a person's actual body shape, e.g., virtual try-on rooms, therapy applications, or specific physical training scenarios that prepare people for real-life situations.

In previous work, the self-similarity of the avatar influenced the users' perception in the virtual environment. Personalized realistic-looking avatars enhanced the illusion of body ownership and the feeling of presence in first-person (Waltemate et al., 2018) and third-person (Gorisse et al., 2019) immersive VR. Self-similarity enhanced negative attitude changes when embodying a self-similar but sexualized avatar (Fox et al., 2013) and impacted body weight perception (Thaler et al., 2018). Having a self-similar body in VR promoted creativity (de Rooij et al., 2017) and increased presence and social anxiety levels in VR (Aymerich-Franch et al., 2014). In a fitness application with a full-body virtual mirror, having an avatar that was self-similar in terms of gender enhanced the illusion of body ownership and increased performance compared to a not self-similar one (Lugrin et al., 2015c). Especially in social VR applications, people very deliberately choose to look or not look like they do in real life (Freeman and Maloney, 2021). Realistic avatar representations used for embodiment have been superior to abstract avatar representations in user acceptance (Latoschik et al., 2017). Nevertheless, the role of realism in avatars is still in debate. Other work could not reproduce this superiority (Lugrin et al., 2015b) and even found realistic avatars to be less accepted than abstract representations (Lugrin et al., 2015a). The context of the experience might be an important factor when it comes to the influence of the own avatar's appearance. The impact seems to be less significant in game-like or overall more stressful scenarios that strongly engage the user in a superordinate task that only marginally focuses on the body (e.g., Lugrin et al., 2015a,c). But it might be of greater importance for social scenarios (e.g., Aymerich-Franch et al., 2014; Freeman and

Maloney, 2021) or experiences where the user and his body is the center of attention.

2.1.2 The Virtual Appearance of Others

In virtual environments, users can also encounter virtual humans as computer-controlled virtual agents or embodied other, real users. Previous work showed that a virtual agent's appearance influenced co-presence (Bailenson et al., 2005). Nelson et al. (2020) found that virtual agents' appearance influences users' movement speed and their interpersonal distance to the agents. In social VR applications, another user's avatar's appearance influences whether and how others approach this user (Freeman and Maloney, 2021). A realistic appearance of a virtual agent impacts its perceived appeal and friendliness (McDonnell et al., 2012). Other previous work looked at the impact of realistic-looking interaction partners on perceived trustworthiness (McDonnell et al., 2012; Jo et al., 2017; Seymour et al., 2019). Seymour et al. (2019) found a preference for realistic virtual agents, which also increased the users' place illusion (Zibrek et al., 2019). Zibrek et al. (2018) investigated the impact of virtual agents' realism in virtual reality games and found complex interactions between the virtual agents' personality and appearance.

A recurring debate about the realism of virtual characters is the uncanny valley effect. Initially described by Mori et al. (2012) for human-robot interactions in the 1970s and later transferred to virtual characters, the uncanny valley effect refers to the phenomenon that close-to-real looking artificial humans sometimes strike as eerie. The original work sets human-likeness in correlation with familiarity. It proposes a drop in familiarity when the artificial character looks close to but not entirely like a human. Research on this effect is not at all consensus. Some argue that the uncanny valley effect might only occur under specific circumstances that are yet to be defined (Kättsyri et al., 2015). Some explain that the phenomenon is a wall rather than a valley since people adapt to the technical advances and therefore, the uncanny valley is untraversable (Tinwell and Grimshaw, 2009). Others argue that the key to overcoming the uncanny valley with realistic-looking characters lies in their behavior (Seymour et al., 2017, 2019). And finally, some question the existence of the uncanny valley effect as a whole (Wang et al., 2015).

In summary, research on the realism of virtual humans has been controversial for decades and is still ongoing. However, it is especially relevant today as methods for creating virtual humans are improving drastically along with the overall evolution of technology, creating new and reviving old research questions (Slater et al., 2020).

2.2 Creation Methods for Virtual Humans

Methods to create and use realistic virtual humans are essential for research in this area. With applications like Epic Games' MetaHuman Creator (Epic Games, 2021) it is now possible for everyone to create virtual humans that are state-of-the-art regarding their realism, especially regarding their hair, faces, and facial expressions. However, while it is possible to create generic virtual humans with this approach, the customization and personalization still relies on extensive manual work. For the 3D-reconstruction of a person, various techniques exist that differ in

terms of the degree of achievable realism, the technical complexities, and the overall reconstruction costs involved. Hardware requirements for current virtual human reconstruction methods range from immensely involved light stage systems (Guo et al., 2019) to single-shot multi-camera photogrammetry rigs (Feng et al., 2017; Achenbach et al., 2017) to a single RGB(-D) camera (Alldieck et al., 2018a,b, 2019a; Loper et al., 2015). Recent approaches lower the hardware requirements even further and rely on a single RGB input image only (Alldieck et al., 2019b; Weng et al., 2019). Wenninger et al. (2020) proposed a low-cost pipeline for generating realistic virtual humans from only two smartphone videos. They follow the high-cost approach of Achenbach et al. (2017) and combine photogrammetric reconstruction with a template fitting approach, using the same virtual human template model. Both methods (Achenbach et al., 2017; Wenninger et al., 2020) allow for fast reconstruction of virtual humans and require minimal user intervention. The characters are ready to be used in standard XR simulation and game engines such as Unreal or Unity. The authors compared the reconstruction fidelity between the high-cost and the low-cost method by computing the geometric difference between the resulting avatars and the reprojection error resulting from rendering the textured avatars back onto their respective input images. The evaluation shows that there still is a difference in both measures but that their low-cost approach can almost reach the same fidelity as the high-cost approach. However, Wenninger et al. (2020) did their evaluation on a purely objective basis. The authors did not address how the still existing differences affect users' perception of the virtual humans.

Based on the presented literature, we specify our research goal: We build on the purely objective comparison of Wenninger et al. (2020) and focus on the user perception of the resulting virtual humans. In a user study, we compare a high-cost method to create virtual humans to a low-cost method. The methods differ in their hardware requirements (low-cost vs. high-cost), the input material (multiple images vs. two smartphone videos), and software parameters for tailoring the approach to the specific input material. We investigate whether the differences in the quality of low- and high-cost reconstructions of virtual humans produce differences in the users' perception. The evaluation includes one's reconstructed self-avatars and virtual others, here, computer-controlled reconstructions of other real persons. We compare the users' perception in terms of the similarity of the virtual humans to the original, the sense of embodiment (only for the self-avatars), their uncanniness, and the overall preference for one of the approaches. We also investigate if differences between the high- and low-cost virtual humans are more noticeable for one's self-avatar than for virtual others. Finally, we compare the low- and high-cost virtual humans using objective measures, i.e., the reprojection error and the geometrical error.

3 STUDY

To investigate our research questions, we designed a user study that focuses on the perception of the low-cost and high-cost

virtual humans. Regarding RQ_1 , we compared the subjectively perceived quality of two 3D-reconstruction methods for realistic virtual humans. In particular, we compared one method using a high-cost photogrammetry rig containing 94 DSLR cameras with a low-cost method processing two smartphone videos. For this purpose, we scanned participants twice and created one personalized self-avatar with each generation method. In a virtual environment, participants embodied both self-avatars and observed themselves in virtual mirrors. They also encountered and evaluated other virtual humans originating from both scan processes, observing them on virtual monitors. The independent variable for RQ_1 was the reconstruction method (low-cost vs. high-cost) that we investigated for self-avatars and virtual others separately. To answer RQ_2 , participants could adjust the distance between themselves and the mirrors or monitors. The task was to set the distance at which they could no longer tell that one version was better than the other. We assumed that there would be a difference in the distance that participants set for the mirrors (self) compared to the distance they set for the monitors (other) if the quality discrepancy between methods was more noticeable for one's own or for another virtual body (RQ_2). Therefore, the independent variable for RQ_2 was the virtual human (self vs. other). The study followed a repeated-measures design.

3.1 Virtual Humans

3.1.1 High-Cost and Low-Cost Method

Figure 1 displays both the high-cost and the low-cost scan processes, including example results for a sample participant. A photogrammetry rig that contains 94 DSLR cameras generates the input for the high-cost avatars. In contrast to Achenbach et al. (2017), we did not use a separate face scanner. Instead, 10 of the 94 cameras of the body scanner are zoomed in on the scan subject's face, therefore, capturing more detail in this area. The scanner includes four studio lights with diffuser balls (see **Figure 1**, first row, first picture). For generating the avatars from these images, we follow the method of Achenbach et al. (2017), who combine photogrammetric reconstruction with a template fitting approach. The set of images produced by the camera rig is processed with the commercial software Agisoft Metashape (Agisoft, 2020), yielding dense point clouds of the scanned subjects. The subsequent template fitting process is guided by 23 landmarks which are manually selected on the point clouds. A statistical, animatable human template model is then fitted to the point clouds by first optimizing the template's alignment, pose, and shape in a non-rigid ICP manner (Bouaziz et al., 2014). Then, allowing a fine-scale deformation to match the point cloud more closely refines the initial registration. The method uses a fully rigged template model provided by Autodesk Character Generator (Autodesk, 2014), which is also equipped with a set of facial blendshapes, thus making the resulting avatars ready for full-body and facial animation. For more details about this process, we refer to the work of Achenbach et al. (2017). The pipeline for generating the high-cost avatars operated on a PC containing an Intel Core i7-7700k, a GeForce GTX 1080 Ti, and $4 \times 16\text{GB}$ DDR4 RAM. The generation took approximately 10 min per avatar.

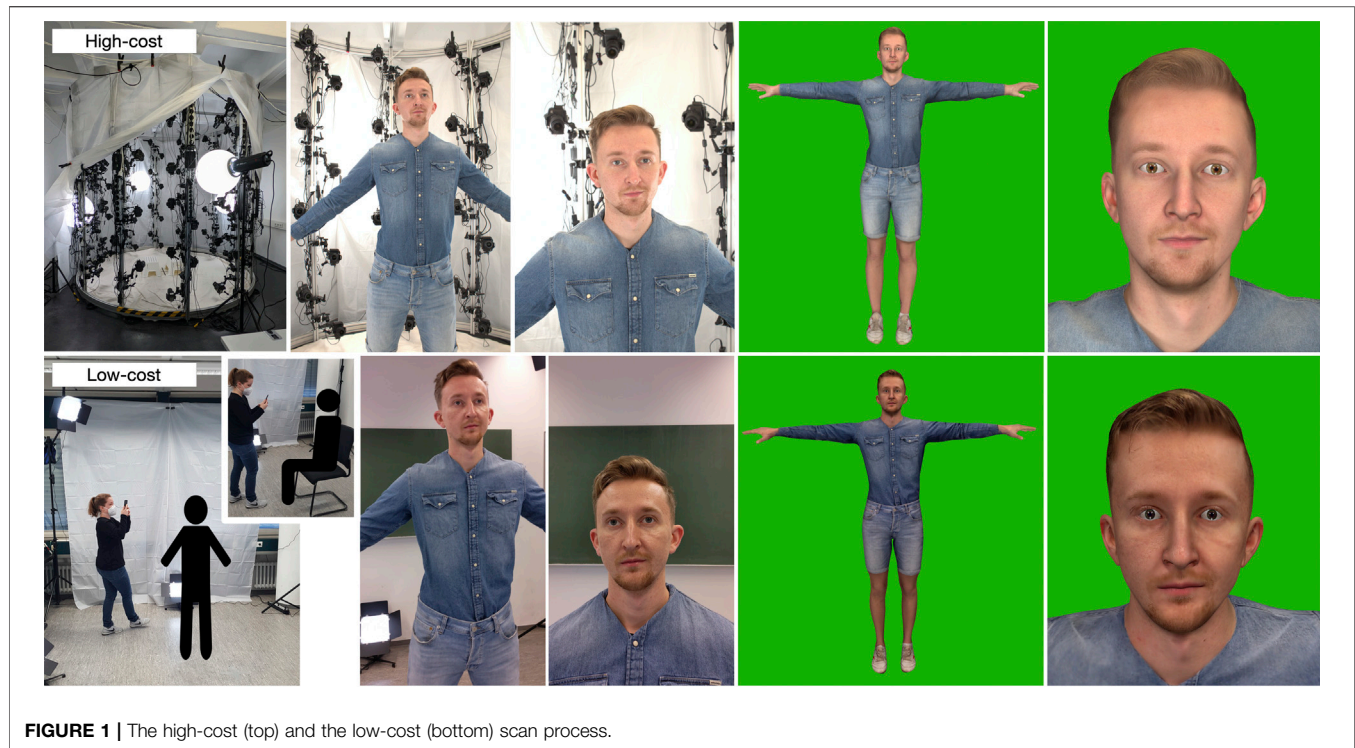


FIGURE 1 | The high-cost (top) and the low-cost (bottom) scan process.

To provide the video input for the low-cost avatar method, we used a Google Pixel 5 smartphone. We used the camera application OpenCamera because it allows for a non-automatic white balance and exposure. The smartphone captured the videos with 4K (3840×2160) resolution and 30 fps. We filmed in a room with covered windows, using the installed ceiling lights and eight additional area lights placed on the floor and on tripods around the participant (see **Figure 1**, second row, first picture). The additional lighting is not necessarily required for generating the low-cost avatars. However, it was added to brighten up the resulting low-cost avatars decreasing the brightness difference between the low-cost and the high-cost variant. After taking two videos of each subject, one capturing the whole body and the other capturing the head in a close-up fashion, the videos are processed with the method of Wenninger et al. (2020). They build on the work of Achenbach et al. (2017), i.e., use photogrammetric reconstruction and template fitting with the same template model, but extend it in several ways to deal with the difference in input modality and quality. One great advantage of stationary photogrammetry rigs is that all cameras trigger simultaneously and thereby capture the scan subject from multiple views at the same moment in time. Using video input induces motion artifacts, since the subject cannot hold perfectly still for the duration of the scan. This contradicts the multi-view-stereo assumption, leading to a decrease in point cloud quality which Wenninger et al. (2020) compensate through several adaptations of the pipeline proposed by Achenbach et al. (2017). First, an optical-flow-based frame extraction method ensures uniform coverage of the scan subject and provides suitable frames for the photogrammetry step. The template fitting process then relies on a stronger regularization towards

the statistical human body shape model in order to deal with uncertainties in the input data. Lastly, the method employs a graph-cut-based texture generation in order to deal with misalignments in the photogrammetry step, which result from motion artifacts in the video input. Again, we refer to the work of Wenninger et al. (2020) for more details about the avatar generation. The pipeline for generating the low-cost avatars operated on a PC containing an Intel Core i7-7820x, a GeForce GTX 1080 Ti, and $6 \times 16\text{GB}$ DDR4 RAM. The generation took approximately 20 min per avatar.

To summarize, the two pipelines for generating virtual humans that we compare in this study are based on the same reconstruction methodology, i.e., photogrammetric reconstruction and template fitting using the same template model. They differ, however, in hardware costs, input modality and quality, necessary preprocessing steps, template fitting parameters, and texture generation.

3.1.2 Self-Avatar Animation

The generated low- and high-cost self-avatars were both imported to our Unity application. For the avatar animation, we oriented towards the system architecture introduced by Wolf et al. (2020) and adapted their implementation. During the experiment, the two imported avatars were simultaneously animated in real-time according to the users' movements by using HTC Vive Trackers (see **Section 3.2** for details about the VR setup). To this end, we used the calibrated tracking targets of the head, left hand, right hand, pelvis, left foot, and right foot to drive an inverse kinematics (IK) animation approach realized by the Unity plugin FinalIK version 2.0 (Rootmotion, 2020).



FIGURE 2 | The female (left) and male (right) virtual others. The left virtual monitor of each pair displays the high-cost version; the right virtual monitor displays the low-cost version.

3.1.3 Virtual Others

For the virtual others, we scanned one male and one female person. **Figure 2** displays both versions of the virtual other. Male participants observed and evaluated the male other, female participants the female other. Both virtual others wore identical grey t-shirts and blue jeans. We recruited two persons who do not represent extremes in terms of their appearance. The male other was 1.72 m tall; the female other was 1.66 m tall. Both persons stated that they do not know any of the students belonging to the study's participant pool. The virtual others were animated using a pre-recorded idle animation. The animation showed a basic idle standing animation including small movements, e.g., slightly moving from one foot to the other. We also added random eye movements and blinking using an existing asset of the Unity Asset Store¹ to increase the virtual others' realism.

3.2 Virtual Reality System

We implemented our study system using the game engine Unity version 2019.4.15f1 LTS (Unity Technologies, 2019) running on Windows 10. The VR hardware explained in the following was integrated with SteamVR version 1.16.10 (Valve, 2020) and its corresponding Unity plugin version 2.6.1. As high-immersive VR display system, we used a Valve Index HMD (Valve Corporation, 2020), providing the user a resolution of $1,440 \times 1,600$ pixels per eye with a total field of view of 120 degrees running on a refresh rate of 90 Hz. For capturing the user's motions, participants held the two Valve Index controllers in their hands, wore one HTC Vive Tracker 2.0 on a belt around the hips, and one fixed on each shoe's upper side with a velcro strap. Three SteamVR 2.0 base stations braced the spacious tracking area. The system ran on a high-end, VR-capable PC composed of an Intel Core i7-9700K, an Nvidia RTX2080 Super, and 16 GB RAM. We determined the motion-to-photon latency of our system by frame-counting (He et al., 2000). For this purpose, the graphics card's video signal

output was split into two signals using the Aten VanCryst VS192 display port splitter. One signal led to the HMD and the other to the low-latency gaming monitor ASUS ROG SWIFT PG43UQ. A high-speed camera of an iPhone 8 recorded the user's motions and the corresponding reactions on the monitor screen at 240 fps. Counting the recorded frames between the user's motions and the corresponding reactions on the screen, we determined the latency for the HMD and limb movements separately. For HMD and limb movements, we repeated the measurements ten times each. The motion-to-photon latency for the HMD averaged 14.56 ms ($SD = 2.94$ ms) and therefore matched the refresh rate of the HMD closely. The motion-to-photon latency for the limb movements averaged 42.85 ms ($SD = 5.2$ ms) and was considered low enough for real-time avatar animation (Waltemate et al., 2016).

3.2.1 Virtual Environment and Task

The virtual environment consisted of one large virtual room. In the room, two virtual mirrors were mounted on a track system to allow for a direct comparison of the self-avatars and to induce the feeling of embodiment by visuomotor coherence (Slater et al., 2010; Latoschik and Wienrich, 2021). We told participants that they would see two different mirrors before they saw their self-avatars. The track system was supposed to increase coherence with the users' expectations, making the scenario more plausible (Latoschik and Wienrich, 2021). For the evaluation of the virtual other, the mirrors were exchanged with similar-looking, portal-like, virtual monitors (see **Figure 2**). A stencil buffer masks the area inside the monitor to make the virtual others visible only in this area. This setup preserved a stereoscopic view and ensured a spatial distance to the participants. To help the feeling that the virtual others are in a different place and that the monitors were no mirrors, we added textures to the surrounding walls and floor that were different from the main room. Participants received the audio information that they would see two different broadcasts of another person on these two monitors. This information served the purpose of making the scenario more plausible and less intimidating than directly encountering two similar-looking versions of a person in a virtual room that would not react in

¹<https://assetstore.unity.com/packages/tools/animation/realistic-eye-movements-29168>

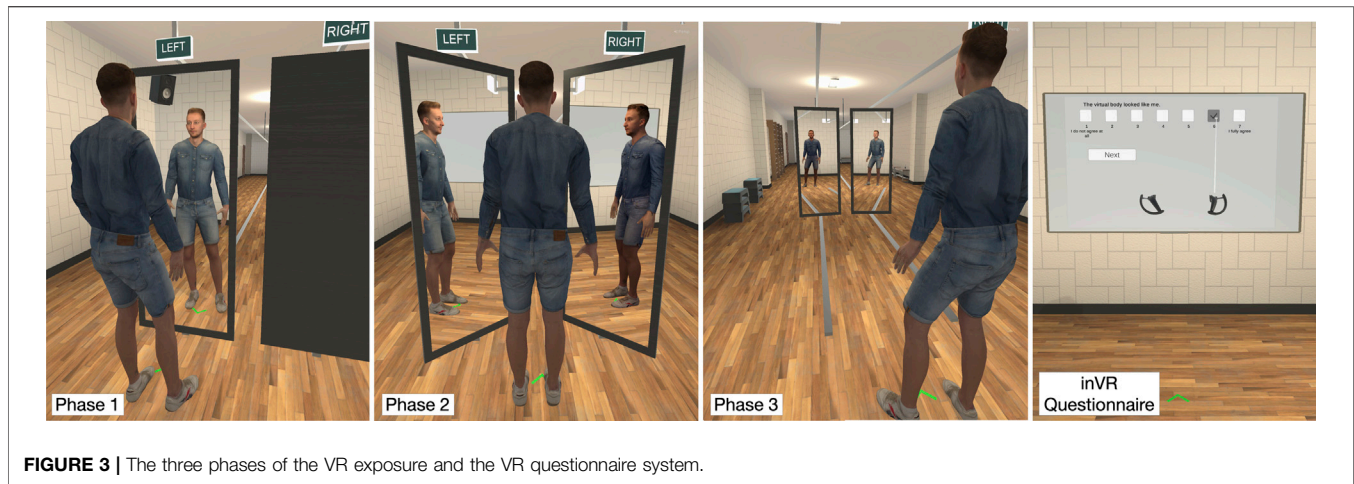


TABLE 1 | Instructions that participants received in Phase 1 while they had to inspect the virtual human in the mirror or monitor. In Phase 2, when participants saw both self-avatars or both virtual others at the same time, they received each instruction twice; first for the left mirror, then for the right mirror, e.g., “Look at your head in the left mirror.” – 5 s duration – “Look at your head in the right mirror.” – 5 s duration.

No.	Instructions Phase 1 - Self	Duration
1	Look at your reflection in the mirror. Please remain standing on the marker. You may move your arms and legs freely.	10 s
2	Look at your head in the mirror.	5 s
3	Swing your arms back and forth while looking at your torso.	5 s
4	Let your arms hang relaxed and slowly shift your weight from your left leg to your right leg and back again. Repeat this a few times while looking at your lower body.	5 s
5	Stand relaxed. Wave your dominant hand at your reflection while observing yourself in the mirror.	5 s

No.	Instructions Phase 1 - Other	Duration
1	Look at the person.	5 s
2	Look at the head of the person.	5 s
3	Look at the torso of the person.	5 s
4	Look at the lower body of the person.	5 s
5	Now look at the whole person again.	5 s

any way to the user (Slater, 2009; Latoschik and Wienrich, 2021). Study participants would encounter these virtual others for the first time. To enable them to evaluate the virtual humans' similarity to the real person, they needed to see reference material first. We displayed a photo of the real person for 10 s before the virtual other appeared on the monitor(s) and asked participants to memorize it. For the self-avatar similarity assessment, we did not show a photo of the person. Instead, we relied on familiarity with the person's own appearance. Mirrors and monitors turned automatically according to the study phases. **Figure 3** shows the virtual environment throughout the phases of the experiment.

The first two phases of the experiment concentrated on the perception of the virtual humans and the participants' preferences. In Phase 1, participants saw and evaluated the high-cost and the low-cost virtual humans one after another. In Phase 2, they saw both at the same time next to each other. Then they evaluated the left one first. After that, they again saw both at the same time and consecutively evaluated the right one. The photo of the real other person was displayed for 10 s before

every virtual other observation phase. For a controlled exposure, participants received audio instructions on where to look and what movements to perform. **Table 1** lists all instructions and the observation duration. In Phase 2, before and after the instructions, the participants got the information which virtual human they will have to rate (left or right). During the self-avatar observation, the participant always embodied the self-avatar to be rated after the observation. Analogously, participants embodied the high-cost self-avatar when viewing the high-cost virtual other and the low-cost self-avatar when viewing the low-cost version of the virtual other.

In Phase 3, participants could adjust the distance between themselves and the mirrors or monitors. The task was to increase the distance until they could no longer tell which virtual human was better. Participants could move the mirrors and monitors using the controllers' touchpads. One controller increased the distance; one decreased it. When moved back and forth, mirrors and monitors automatically rotated on the track system to always face the user. This ensured that the reflections and the virtual others were always visible to the participants.

3.2.2 Virtual Reality Questionnaire

Participants evaluated the virtual humans directly in VR. The right image in **Figure 3** shows the VR questionnaires from a third-person perspective. Following the guidelines of Alexandrovsky et al. (2020), our VR questionnaire was world-anchored and participants used a controller to operate the questionnaire using a laser pointer. A virtual display presented the VR questionnaire in the virtual environment. It was positioned on the wall left to the user. The integration into the scene's context was supposed to make it more diegetic and thus more plausible (Salomoni et al., 2016). The virtual display was approximately 1.2 m high and 2 m wide. The user stood approximately 1.5 m away from the display. This size and distance allowed the participants to read the questions comfortably without having to move the head. To keep the exposure time with each self-avatar the same for every participant, their embodiment while answering the questions only consisted of visible controllers.

3.3 Measurements

Before and after the experiment, participants answered questionnaires on a computer in the experiment room. During the experimental phases, participants answered VR questionnaires. We used German translations of all questions and questionnaires.

3.3.1 Perception of the Virtual Humans

In Phase 1 and 2, participants rated their self-avatar regarding the perceived similarity, their sense of embodiment (Kilteni et al., 2012; Roth and Latoschik, 2020), and possible uncanny valley effects (Ho et al., 2008; Ho and MacDorman, 2017). The questions regarding the virtual other were the same, only omitting the embodiment questions since they did not apply in this condition.

Similarity: For the measurement of perceived similarity, we adapted the item used by Waltemate et al. (2018). Participants rated their agreement to the statement “*The virtual body looked like me/the person on the image*” on a scale ranging from 1 (*I do not agree at all*) to 7 (*I fully agree*).

Embodiment: For measuring the sense of embodiment, we used the Virtual Embodiment Questionnaire (Roth et al., 2017; Roth and Latoschik, 2020). It consists of three subscales with four items each: Body Ownership, Agency, and Change. Participants rate their agreement to each of the twelve statements on a scale ranging from 1 (*I do not agree at all*) to 7 (*I fully agree*). High values indicate a high sense of embodiment.

Uncanny Valley: Regarding the uncanny valley effect, we built three items based on the original uncanny valley questionnaire's subscales of Ho et al. (2008); Ho and MacDorman (2017). Participants rated their agreement on the three statements: “*The virtual body looked human.*”, “*The virtual body looked eerie.*”, “*The virtual body looked beautiful.*”. Participants rated their agreement to all of the statements on a scale ranging from 1 (*I do not agree at all*) to 7 (*I fully agree*).

Preference: At the end of Phase 2, we directly asked participants which self-avatar/virtual other they preferred using the item: “*Which virtual body was better?*” with the

answer options left or right. We asked if they found the left virtual body to be *much worse*, *worse*, *neither worse nor better*, *better* or *much better* than the right virtual body, with a second item. Note that due to the randomization left and right meant different versions for different participants. This was re-coded in the analysis later.

Qualitative Feedback: Between the scan and the experiment, we asked them how they perceived the two scan processes overall. After the whole experiment, we asked them to write down reasons for their preference regarding the version of the self-avatar and the virtual other.

3.3.2 Distance

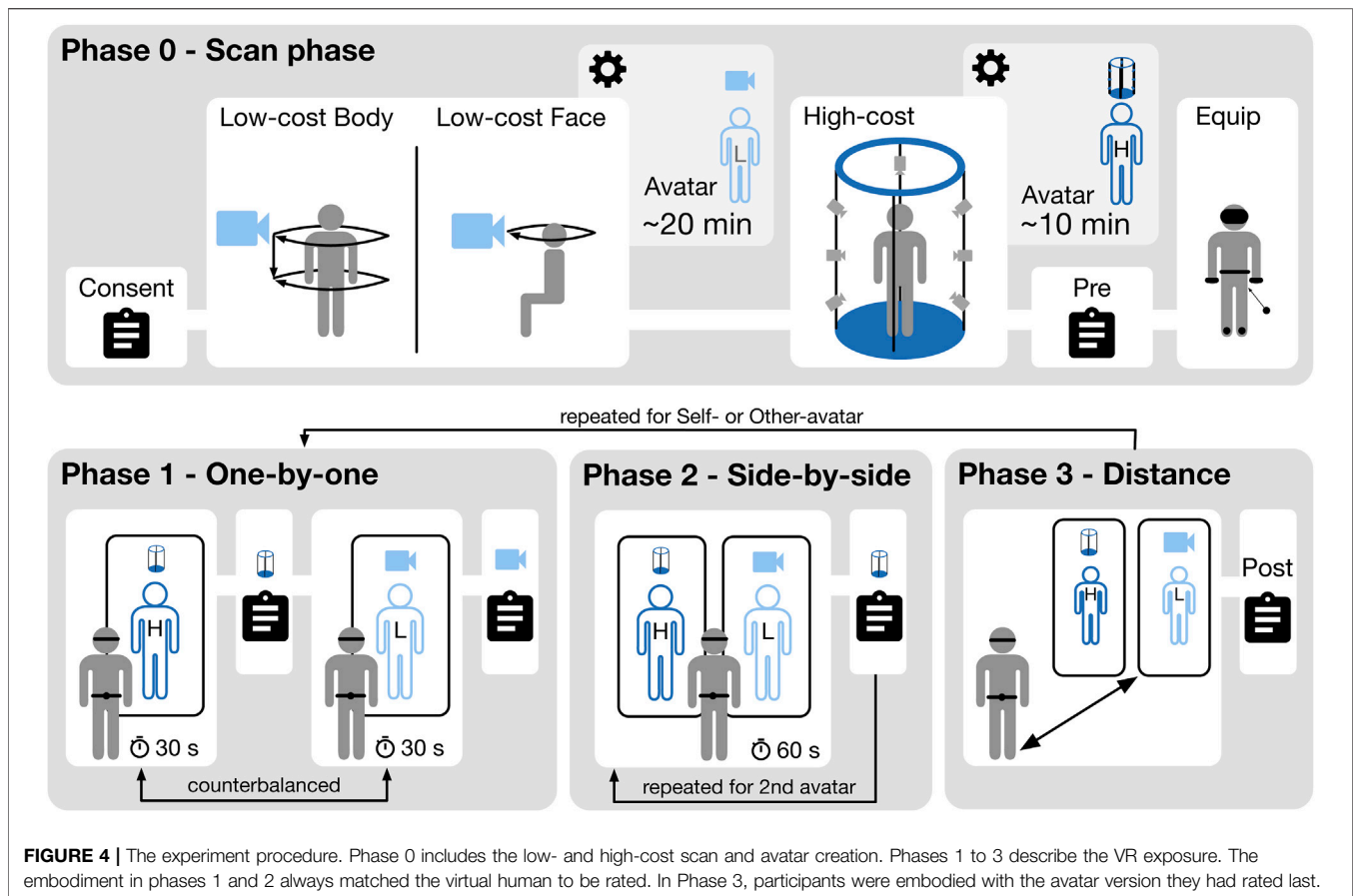
In Phase 3, we asked participants to increase the distance between the virtual bodies and themselves until they no longer can say if one of the virtual humans is better than the other one. We measured the distance in meters between the HMD and the two mirrors (or monitors in the other-condition). For the self-condition, when the participant moved the mirrors away, the reflection logically also moved away. Therefore we multiplied the measurement by two to get the actual distance between the participant and the self-avatars. For the other-condition, we added the distance between the virtual other and the monitor frame (0.5 m) to the distance the participant set. The maximum possible distance between the participant and the monitors was 18 m.

3.3.3 Objective Measures

For comparing the high-cost and the low-cost scans on an objective level, we calculate 1) the reprojection error and 2) the modified Hausdorff distance (Dubuisson and Jain, 1994) between our two reconstruction methods, as also done by Wenninger et al. (2020). The reprojection error is computed by projecting the textured avatar onto each of the cameras as estimated during the avatar generation process. We then calculate the average root-mean-square error (RMSE) between the rendered images and the actual input images in CIELAB color space, giving us a way to measure the reconstruction methods' faithfulness objectively. The modified Hausdorff distance measures the difference in shape between the two reconstruction methods on a purely geometric level.

3.3.4 Control Measures

Before and after the experiment, we used the simulator sickness questionnaire (Kennedy et al., 1993) to measure virtual reality sickness as described by Kim et al. (2018). The questionnaire includes 16 symptoms of simulator sickness. The participants rated how much they experienced each symptom on a scale ranging from 0 (*none*) to 4 (*severe*). We added three items to check for disturbances in the perceived place and plausibility illusion (Slater, 2009). At the beginning of each VR question phase, we asked participants how present they felt in the virtual environment. For this, similar to Bouchard et al. (2008) and Waltemate et al. (2018), we used one item, namely “*How present do you feel in the virtual environment right now?*” with a scale ranging from 1 (*not at all*) to 7 (*completely*). At the end of each questionnaire phase, we added two items focusing on the overall plausibility: “*The environment made sense.*” and “*The virtual body matched the virtual environment.*”. These items served the



purpose of measuring the environment's plausibility by checking for any unwanted incoherence in the experience caused by the environment (Skarbez et al., 2017; Latoschik and Wienrich, 2021). Participants rated their agreement on scales ranging from 1 (*I do not agree at all*) to 7 (*I fully agree*).

3.3.5 Demographics and User Traits

Participants answered a demographic questionnaire including items for age, gender, educational attainment, occupation, language familiarity, problems with telling left from right, visual and hearing impairments, computer game experience, and virtual reality experience. We also asked them if they have been scanned before. Before the experiment, we measured the participants' height and asked them which of their hands is their dominant one. Additionally, we measured immersive tendency using the Immersive Tendency Questionnaire (Witmer and Singer, 1998).

3.4 Procedure

Figure 4 shows the experimental procedure. Each session took around 90 min: 30 min for the scan preparation, the two scans, and the avatar generation (Phase 0). 30 min for answering questionnaires before and after the experiment as well as for putting on the VR equipment. 30 min for the VR exposure (Phases 1–3). At the beginning of Phase 0, participants received a written introduction and signed consent forms for being scanned, participating in the study, and for COVID-19

related regulations. The video scan to create the low-cost avatars was made first to optimize the schedule. After the two scans, the participant filled in pre-questionnaires while the avatars were generated. Then, the experimenter helped the participant to put on the VR equipment and explained how to operate the controllers. After an initial calibration of the avatar, the experiment started. The participants received audio instructions that guided them through the VR exposure phases. The low- and high-cost avatars' rating order and therefore their display in the left or right mirror, was counterbalanced. Each participant went through Phases 1 to 3 twice. Once for the self-avatar, a second time for the virtual other. Half of the participants started with the self-avatar, the other half started with the virtual other. After repeating the phases, participants left the virtual environment and answered the post-questionnaire on a computer in the experiment room.

3.5 Participants

A total of $N = 51$ people participated in the study. We had to exclude six participants from the analysis. Three were excluded because the quality of the point cloud of the low-cost scan was insufficient. Another three participants were excluded due to errors in the experimental procedure, e.g., wrong height input when generating the avatars. The mean age of the resulting sample was $M = 21.78$, $SD = 1.80$. 75.6% stated to be female, 24.4% stated to be male. They were all students that received credit points necessary for completing their bachelor's degree.

TABLE 2 | Means, standard deviations, and test statistics for the paired samples t-tests for the perception of the virtual humans. For all t-tests: $df = 44$. * $< .05$

Measurement		Phase 1				Phase 2			
		high-cost		low-cost		high-cost		low-cost	
		<i>M(SD)</i>	<i>M(SD)</i>	<i>t</i>	<i>p</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>t</i>	<i>p</i>
Similarity	self	4.82 (1.45)	4.47 (1.52)	1.79	0.08	4.64 (1.45)	4.42 (1.52)	0.87	0.39
	other	5.22 (1.17)	5.51 (1.08)	-1.44	0.16	5.04 (1.30)	5.22 (0.97)	-0.85	0.40
Human-likeness	self	4.42 (1.52)	4.16 (1.49)	1.45	0.15	4.16 (1.41)	3.98 (1.34)	1.02	0.32
	other	4.93 (1.23)	4.91 (1.06)	0.11	0.92	4.73 (1.27)	5.07 (0.94)	-1.56	0.13
Beauty	self	3.69 (1.28)	3.42 (1.47)	1.18	0.24	3.69 (1.51)	3.29 (1.41)	1.46	0.15
	other	4.38 (1.28)	4.67 (1.23)	-1.48	0.15	4.40 (1.39)	4.73 (1.01)	-1.39	0.17
Eeriness	self	3.98 (1.55)	4.36 (1.55)	-1.57	0.12	4.18 (1.81)	4.71 (1.63)	-1.76	0.09
	other	3.29 (1.63)	3.16 (1.35)	0.63	0.53	3.47 (1.78)	3.29 (1.36)	0.57	0.57
VEQ-Owners.	self	4.09 (1.50)	4.06 (1.43)	-	-	4.03 (1.45)	4.12 (1.45)	-	-
VEQ-Agency	self	5.64 (1.03)	5.67 (0.97)	-0.24	0.82	5.43 (1.23)	5.42 (1.01)	0.08	0.94
VEQ-Change	self	3.38 (1.53)	3.55 (1.47)	-0.82	0.42	3.03 (1.55)	3.50 (1.60)	-2.42	*

TABLE 3 | Effect sizes indicated by Cohen's d_z (Cohen, 1977) for the perception measures.

Measurement		Phase 1		Phase 2		
		d_z	d_z	d_z	d_z	
Similarity	self	0.27	0.13	other	-0.22	-0.13
Human-likeness	self	0.22	0.15	other	0.02	-0.23
Beauty	self	0.18	0.22	other	-0.22	-0.20
Eeriness	self	-0.24	-0.26	other	0.09	0.09
VEQ-Agency	self	-0.04	-0.07			
VEQ-Change	self	-0.12	-0.36			

Ten participants had been scanned with the high-cost method before. The sample's VR experience was low, with 84.4% stating that they have 0–5 h of VR experience. Only four participants had no prior VR experience at all.

4 RESULTS

The analysis was performed using IBM SPSS Statistics 26. First, we report on the main analysis of the presented user study, including objective measurements. Then we proceed with the results of our control measures. We performed paired t-tests for all within-subjects comparisons and independent t-tests for between-subjects comparisons. Effect sizes are indicated by Cohen's d_z (Cohen, 1977).

4.1 Perception of the Virtual Humans (RQ₁)

Table 2 shows the dependent variables' descriptive data: similarity, uncanniness, and sense of embodiment. Table 3 shows the effect sizes of the comparisons.

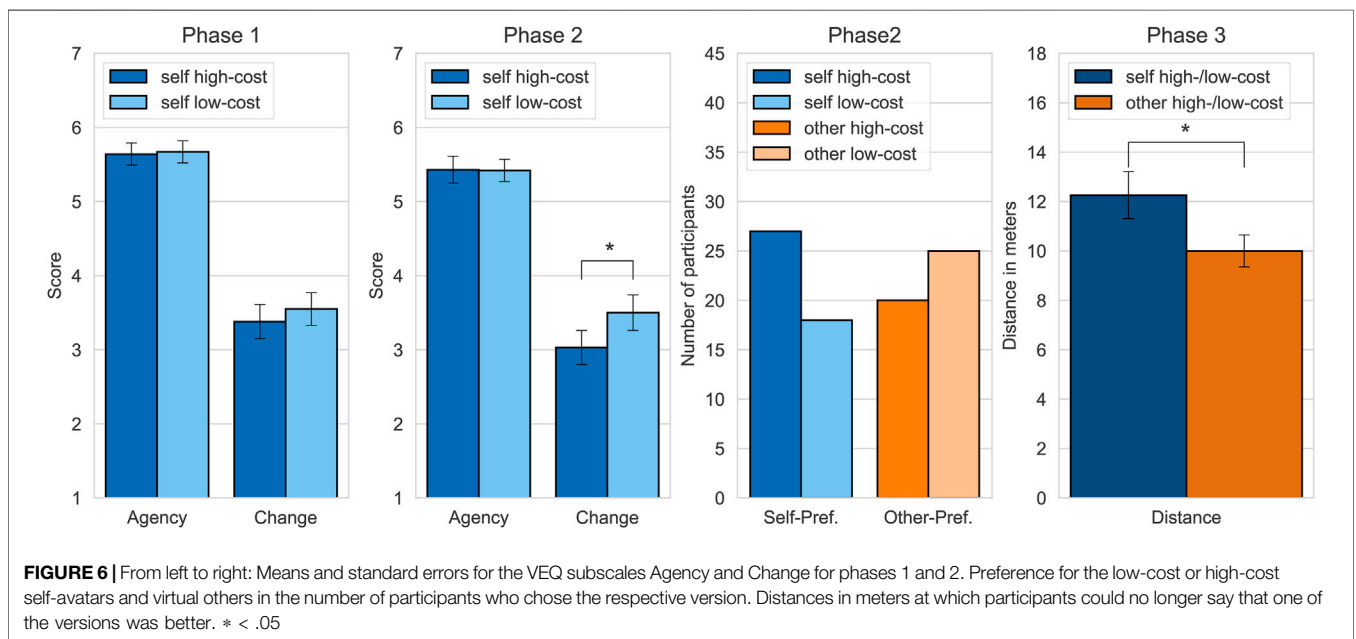
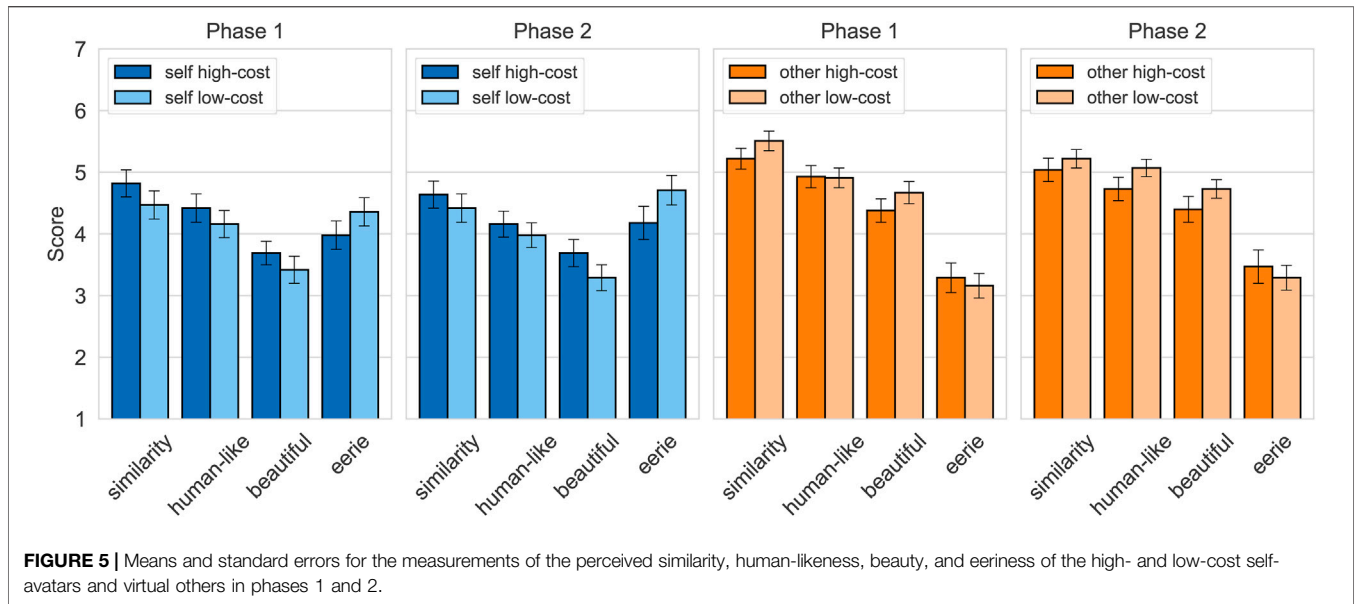
Similarity: Figure 5 shows the results for the perceived similarity. We found no significant difference in the perceived similarity to oneself between the low-cost and the high-cost self-avatar, neither when compared one after the other in Phase 1 nor when compared side-by-side in Phase 2. We also found no significant difference in the perceived similarity to the other

person's picture between the low-cost and the high-cost virtual other neither in Phase 1 nor in Phase 2.

Uncanny Valley: Figure 5 shows the results for the items human-like, beautiful, and eerie associated with the uncanny valley effect. For the self-avatars, we found no significant difference regarding the perceived human-likeness, beauty, and eeriness of the avatars when evaluated one after the other (Phase 1). We also found no significant difference regarding the perceived human-likeness, beauty, and eeriness of the avatars when evaluated side-by-side (Phase 2). For the virtual others, we also found no significant differences in both phases regarding the perceived human-likeness, beauty, and eeriness.

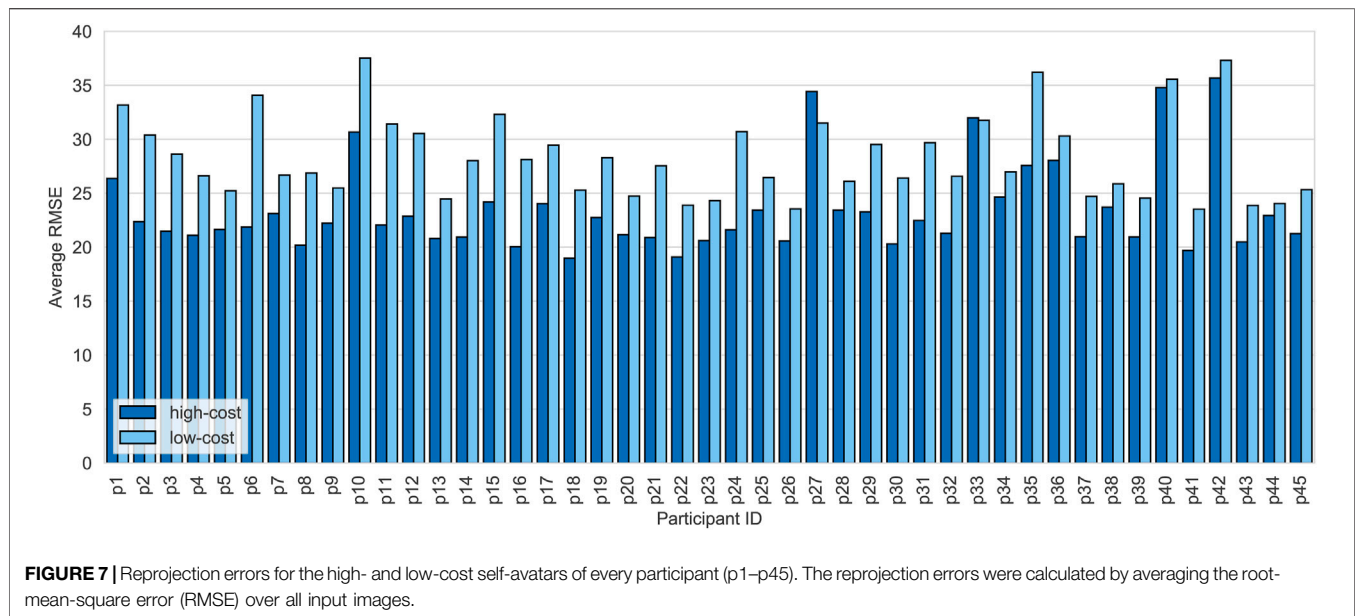
Sense of Embodiment: Data logging failed for one of the four items of the subscale Body Ownership. Therefore, we exclude this subscale from the calculation of the comparisons and only report the descriptive statistic derived from the remaining three items. Table 2 shows the mean scores calculated with three instead of four items which are almost identical between conditions. Agency did not differ between the high-cost and the low-cost self-avatar in both phases. The perceived change did not differ in Phase 1. It did, however, differ in Phase 2 when participants saw the self-avatars side-by-side. The perceived change of the own body was significantly higher for the low-cost self-avatar than for the high-cost self-avatar. The left two diagrams in Figure 6 show these results.

Preference: The third diagram in Figure 6 shows the participants' preferences for the high- and low-cost self-avatars and virtual others. When asked directly, $n = 27$ participants preferred the high-cost self-avatar and $n = 18$ participants preferred the low-cost self-avatar. On a scale ranging from -2 (much worse) to 2 (much better), the participants, on average, found the low-cost self-avatar to be only slightly worse than the high-cost self-avatar, $M = -0.42$, $SD = 1.29$. Regarding the virtual others, $n = 20$ preferred the high-cost version and $n = 25$ preferred the low-cost version. On average, they rated the low-cost virtual other to be slightly better than the high-cost virtual other, $M = 0.24$, $SD = 1.15$.



Qualitative Feedback: Participants described the high-cost scan process as interesting, easy, professional, and quick. They stated the number of cameras to be slightly intimidating, futuristic, and strange because they felt observed. As for the low-cost scan, some participants found it strange (especially that a stranger had to film them rather closely), slightly more complicated, more time-consuming, and more exhausting because they had to stand still for a longer time. At the same time, many others described this scan process as easy, interesting, and pleasant. Feedback regarding their preference focused on some main aspects: 1) The face played a vital role in their judgment. Many stated that the bodies of both virtual humans

were similarly good in quality. However, artifacts in the face of the one virtual human or a perceived higher similarity made them choose the other version. 2) Participants could rather precisely name artifacts, e.g., messy textures under the arms and smaller deformations that deviated from their real body. However, often, they just described an overall feeling that one virtual human was more uncanny or less human-like or more similar to the original. The arguments for the two versions overlapped a lot. However, many of the participants who chose the high-cost avatar as their preference named artifacts on the low-cost avatar as their reason. 3) The lighting and brightness of the virtual human was an important factor. Some stated that the low-cost version looked



more realistic because the lighting looked more natural and that it had more details. Some felt the other way round, that the high-cost version was illuminated better, was more detailed, and looked more realistic.

4.2 Distance (RQ₂)

The right diagram in Figure 6 shows the distances that participants set in Phase 3. For the self-avatars, the distance at which participants could no longer tell which avatar was better was, on average, $M = 12.26$, $SD = 6.36$ m. For the virtual other, this distance was, on average, $M = 10.00$, $SD = 4.34$ m. The distance for the self-avatars was significantly greater than for the virtual other, $t(44) = 2.61$, $p = 0.01$, $d_z = 0.39$.

4.3 Objective Measures

Figure 7 shows the reprojection error for all participants for both the high-cost and the low-cost self-avatar. On average, the high-cost method's reprojection error was $M = 23.40$, $SD = 4.14$, while the reprojection error of the low-cost method was $M = 28.30$, $SD = 3.84$. A paired samples t -test showed, that the difference was significant, $t(44) = -11.52$, $p < 0.001$, $d_z = -1.72$. The modified Hausdorff distance between the two reconstructions was, on average, $M = 7.67$ mm, $SD = 2.43$ mm. The reprojection errors and the modified Hausdorff distance that we found, are in the same range as reported by Wenninger et al. (2020).

4.4 Control Measurements

The experienced VR sickness before, $M = 7.54$, $SD = 7.98$, and after the experiment, $M = 16.37$, $SD = 12.10$, was low. The increase was significant, $t(44) = -5.4$, $p < 0.001$, $d_z = -0.81$. However, we find this to be uncritical because the values are both low, the application's measured latency was low, the experimenters observed no signs of distress, and the participants did not complain of severe symptoms.

Table 4 shows the descriptive data of the control measurements that we took in phases 1 and 2. The subjective

experience of presence did not differ between the moment when participants rated the low-cost avatar and when they rated the high-cost avatar, neither when evaluating the self-avatar nor when evaluating the virtual other in both phases. We also found no significant differences regarding the environment's perceived plausibility and the match between the virtual humans and the environment.

5 DISCUSSION

This work addresses the potential of affordable methods for the 3D-reconstruction of realistic virtual humans for immersive virtual environments. In a user study, we compared the results of a low-cost method (Wenninger et al., 2020) to the results of a high-cost method (Achenbach et al., 2017) used as self-avatars and virtual others. Our research followed two research questions: RQ₁: Whether low-cost approaches for generating realistic virtual humans can keep up with high-cost solutions regarding the perception of the resulting virtual humans by users in VR. RQ₂: Whether the quality difference was more noticeable for the own virtual body than the virtual body of someone else.

For investigating RQ₁, participants evaluated self-avatars and virtual others originating from both reconstruction methods. Users perceived the low-cost virtual humans as similarly human-like, beautiful, and eerie as the high-cost versions for the self-avatars and the virtual others. The perceived similarity between the virtual human and the real counterpart did also not differ between the reconstruction methods. Neither did we find significant differences when evaluating the self-similarity, nor when evaluating the similarity between the virtual others and pictures of the real persons. The participants' qualitative feedback suggests that the self-avatars' perceived eeriness – independent of the reconstruction method – depended heavily on their faces. A possible explanation is the lack of facial animations. We did not track the users' facial expressions, and

TABLE 4 | Means, standard deviations, and test statistics for the paired samples t-tests for the control measures presence, plausibility of the environment, and match of the virtual body to the virtual environment. For all t-tests: $df = 44$.

		Phase 1				Phase 2			
		high-cost	low-cost			high-cost	low-cost		
Measurement		<i>M(SD)</i>	<i>M(SD)</i>	<i>t</i>	<i>p</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>t</i>	<i>p</i>
Presence	self	5.62 (1.17)	5.38 (1.35)	1.57	0.13	5.44 (1.20)	5.47 (1.36)	-0.15	0.88
	other	5.09 (1.35)	5.13 (1.25)	-0.39	0.70	5.22 (1.40)	5.36 (1.21)	-1.29	0.20
VE made sense	self	5.38 (1.34)	5.31 (1.38)	0.52	0.61	5.50 (1.20)	5.38 (1.27)	0.93	0.36
	other	5.31 (1.38)	5.36 (1.32)	-0.39	0.70	5.33 (1.41)	5.29 (1.27)	0.36	0.72
Match Body-VE	self	5.44 (1.14)	5.20 (1.16)	1.53	0.13	5.51 (1.20)	5.16 (1.42)	1.91	0.06
	other	5.51 (1.08)	5.51 (0.90)	0	1.0	5.42 (1.17)	5.42 (1.12)	0	1.0

therefore, the self-avatars' faces remained static. This rigidity was inconsistent with the otherwise realistic-looking and -moving virtual human. Following the mismatch hypothesis for the uncanny valley effect, which states that inconsistencies in a virtual human's human-like and artificial features may increase negative affinity (Kätsyri et al., 2015), this potentially increased the perceived eeriness. The virtual others included basic facial animations and the descriptive data suggests that participants perceived them as less eerie. This is also in line with previous research on the interplay between appearance and behavioral realism, especially regarding the importance of eye movements (Garau et al., 2003; Brenton et al., 2005). In future work, we plan to track the users' eyes for two reasons. Firstly, this would improve the behavioral realism of the self-avatars. Additional sensors like the VIVE face tracker, which entered the market shortly after we conducted our study, would be supplementary improvement options. Secondly, the eye-tracking data could reveal which parts of the virtual humans mostly draw the users' visual attention and, consequently, impact the evaluation the most. However, our study did not focus on the general perception but on the differences in the perception of the high-cost and low-cost virtual humans.

For the two different self-avatars, we additionally measured the users' sense of embodiment. Participants accepted both self-avatar versions as their virtual body (body ownership) and felt that they were the cause of the self-avatar's actions (agency). In the first phase of the evaluation, when the participants saw the self-avatars consecutively, we also found no significant difference in the embodiment questionnaire's change subscale. However, in the second phase, when participants saw both avatar variations next to each other, the change subscale was significantly higher for the low-cost self-avatars than for the high-cost self-avatars. The subscale change measures the perceived change in the users' body schema (Roth and Latoschik, 2020). According to the questionnaire's authors, the perceived change could be a predecessor of the Proteus effect. When embodying an avatar that does not look like the user, the perceived change of the users' body would increase with an increased feeling of embodiment. However, a personalized, realistic-looking self-avatar should not create a massive change in the own body schema since it looks (and ideally behaves) like the real body of the user. There are two possible explanations for the increase in perceived change in the second phase: 1) The low-cost self-avatars have more visible inaccuracies than the high-cost self-avatars, e.g., messy textures

under the arms. These artifacts on the otherwise very realistic and faithfully reconstructed avatars represent deviations from the users' body, which might cause the increased feeling of change of the own body. 2) These deviations may also have surprised the users and drawn their attention to them. The incoherence with the users' expectations could have created an increased interest and focus on the discrepancies. Latoschik et al. (2019) observed a similar effect when participants interacted with a mixed crowd of virtual characters that drew attention because of their diversity and unexpectedness. However, we did not find significant differences in the feeling of presence, which is usually also partly dependent on the users' attention (Skarbez et al., 2017). The increase in the perceived change only occurred in the second phase, when participants saw the low-cost and high-cost self-avatars next to each other. This direct comparison, and the fact that they saw the self-avatars for the second time at this point, may have further increased the focus on the artifacts. It is possible that the increase in perceived change of the own body only occurs when participants spend a longer time with the virtual body and when they look for discrepancies.

Interestingly, the perception did not differ significantly on most of our measures, even though we found a significant difference in our objective quality measures. The medium may be one possible explanation for this. Despite ongoing technological advances in terms of display quality, today's common consumer HMDs are still limited. We used an HMD with standard resolution (1,440 × 1,600) and a wide field of view (130°) that we considered at the upper end of the SteamVR compatible hardware. It would be interesting to see if quality differences between the avatar versions become more apparent using better HMDs, like the HTC Vive Pro 2, that was released after we conducted our study. However, as the user feedback shows, participants were able to spot artifacts quite precisely. Nevertheless, the perceived differences did not manifest themselves in the subjective measurements. This is even more surprising since participants were instructed to really focus on the virtual human. It could mean that other factors, e.g., the movements of the virtual humans, were stronger influences than the visible artifacts. Hence, as a consequence we might want to assume the low-cost smartphone-based version to be an accurate technological match to the available state-of-the-art of VR display devices.

To find out which version was overall preferred, we asked the participants to decide which version of the self-avatars and which version of the virtual others they liked better. Here, the tendency was different between the self-avatars and the virtual others. 60% of participants preferred the high-cost self-avatars over the low-cost ones. Regarding the virtual others, the result was the other way round. Around 56% of participants preferred the low-cost virtual others over the high-cost ones. This is interesting and supports the overall findings for RQ_1 that the low-cost and high-cost virtual humans are very similar regarding the users' perception.

To sum up our findings regarding RQ_1 , we conclude that the low-cost method used in our comparison can indeed keep up with the high-cost method regarding the users' overall perception. The two versions of virtual humans were perceived comparably in respect to their perceived similarity to the original, human-likeness, beauty, and uncanniness. The relatively small effect sizes of the non-significant differences for the self-avatars and the virtual others further support this conclusion.

In our second research question, RQ_2 , we focus on the severity of the quality difference for the own body in comparison to the body of a virtual other. Users increased 1) the distance between themselves and their self-avatars and 2) the distance between themselves and the virtual others until they could no longer tell that one of the virtual humans is better than the other. The distance at which the difference between the low-cost and the high-cost version was no longer noteworthy differed between the self-avatars and the virtual others. For the self-avatars, this point was roughly 2 m further away than for the virtual others. This difference implies that smaller discrepancies between the real body and the reconstructed virtual body seem to be more noticeable for one's own body than for another person's body. This is explainable by the familiarity with one's own body, which is usually higher than for someone else's body, in particular if the person is a stranger to you. Our results regarding the participants' preferences also support this assumption. Here, more than half of the participants preferred the low-cost version for the virtual others. However, to further strengthen this finding by correctly representing the interpersonal quality variance of the respective reconstruction methods, a study that evaluates more than two pairs of virtual others would be necessary.

To summarize the results regarding RQ_2 , we conclude that the quality difference between the low- and high-cost method plays a more important role for one's own virtual body than for virtual others. In future work, we plan to strengthen this finding by evaluating a more diverse group of virtual others.

The objectively measured quality differences for our sample are similar to those reported by Wenninger et al. (2020). The reprojection error was significantly higher for the low-cost self-avatars compared to the high-cost self-avatars. However, the severity of the visible artifacts varied a lot within the sample. For some participants, the reprojection error was even lower for the low-cost self-avatars (p27 and p33). We investigated this within-method variance further by scanning the same persons multiple times with both methods. For the resulting virtual humans, we then measured the geometrical variance produced by both methods. This evaluation did, however, not reveal a correlation between the visible artifacts and the geometrical variance. Although both

methods are photogrammetric approaches, they differ in many ways. The low-cost method, for example, uses a stricter regularization to the base model to handle uncertainties in the input material. Therefore, the resulting virtual humans' geometry is not as detailed as in the high-cost method. For example, the folds in the clothes are more accurately reconstructed in the geometry of the high-cost version than in the geometry of the low-cost version. The lack of small details in the geometry of the low-cost version is compensated by the texture's great detail instead. Additionally, the low-cost texture contains more baked-in lighting, which gives the impression of detailed geometry even if the underlying geometry is flat, e.g., as in **Figure 2** where the folds of the clothing are more visible for the low-cost virtual human. Generally, the lighting in the low-cost method is less controlled, since the experimenter walks around the participant. The controlled lighting setup of the high-cost method leads to a more uniform lighting and weaker shadows, allowing for a more faithful lighting in the virtual scene. However, the qualitative feedback shows that the perception of this difference diverges. While some perceived the baked-in lighting as more detailed and more natural, others felt that the more even lighting of the high-cost virtual humans looked more realistic and overall better. A promising direction for future work is the investigation of causal relations between each method's parameters, their impact on the quality of the reconstruction, and their effect on the users' perception. Our study design can be a helpful basis for conducting these follow-up studies and for guiding the development of similar studies. Ultimately, this allows us to retrieve a set of guidelines for creating and using realistic virtual humans in virtual environments.

Photogrammetric approaches rely heavily on good quality input material. With the described high-cost setting, it is easier to reach a stable quality of the input photos since many factors are well controlled. Camera positions and lighting conditions stay the same, and experimenters have almost no influence on the outcome since they only trigger the cameras. The low-cost method includes more variable factors that can easily lead to a quality loss in the input video material. For example, the camera may lose focus from time to time, the filming person may make mistakes, the environmental conditions are less controlled, and the subjects have to stand still for a longer period. However, in most cases, the solution to these downsides is straightforward: When the input material is not good enough, repeating the scan process using different camera parameters or different environments, e.g., different lighting conditions or backgrounds, can improve the result. Changing parameters in the complex camera rig proves to be more cumbersome and requires recalibration of the whole system. Therefore, the low-cost variant is not only more affordable but also more viable for a broader range of applications in research and industry.

5.1 Limitations

Our study has the following limitations: 1) Our sample was predominantly female. Shafer et al. (2017) found females to be more prone to VR sickness symptoms, which might partly explain the increase in VR sickness after the experiment. 2) Additionally, the perception of the male and female virtual other differed regarding the perceived uncanniness, which might have resulted from the comparably low number of male participants. For better

generalizability of our results, it would be necessary to extend the study by a more balanced sample and more than one female and male pair of virtual others. 3) Our study design is suitable to compare the perception of different versions of virtual humans against each other. However, despite the measurement of the perceived similarity with the real person, we did not include an extensive investigation of the perceived faithfulness of the reconstruction. This was a deliberate decision since it is challenging to find a suitable stimulus for a comparison with reality, e.g., real video material, without changing the medium and therefore impacting the immersion, which in turn can influence the evaluation of a virtual human (Waltemate et al., 2018).

6 CONCLUSION

We compared a high-cost and a low-cost method for 3D-reconstruction of virtual humans that differ heavily in their hardware requirements. Both methods use the same photogrammetric reconstruction and template fitting approach with adaptations to the method-specific input material. In a user study, we scanned participants by both methods. Afterwards, they embodied the resulting self-avatars and also encountered virtual others (created with the same methods) in an immersive virtual environment. We found that even though the reconstructions' quality differed on an objective level, the methods did not differ significantly in most of our measurements regarding the users' perception of the virtual humans. Our results further suggest that the quality difference is of greater importance when it comes to one's own virtual body than to a virtual other's body. Based on our findings, we argue that low-cost reconstruction methods like the method of Wenninger et al. (2020) provide a suitable alternative to high-cost methods, specifically given the current state-of-the-art of available consumer-grade VR displays. In conclusion, the reconstructed virtual humans are affordable but not cheap when it comes to a user's perception, especially when used for virtual others. In future work, we plan to further investigate the causal relations between different quality parameters and their effect on the users' perception of the virtual humans.

DATA AVAILABILITY STATEMENT

The videos, images, and 3D reconstructions presented in this article are not readily available due to participant data privacy.

REFERENCES

- Achenbach, J., Waltemate, T., Latoschik, M. E., and Botsch, M. (2017). "Fast Generation of Realistic Virtual Humans," in Proc. of ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden, November 8–10, 2017, 1–10. doi:10.1145/3139131.3139154
- [Dataset] Agisoft (2020). Metashape Pro. Available at: <http://www.agisoft.com/> (Accessed August 30, 2021).
- Alexandrovsky, D., Putze, S., Bonfert, M., Höffner, S., Michelmann, P., Wenig, D., et al. (2020). "Examining Design Choices of Questionnaires in Vr User Studies," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, United States, April 25–30, 2020, 1–21. doi:10.1145/3313831.3376260
- Alldieck, T., Magnor, M., Bhatnagar, B. L., Theobalt, C., and Pons-Moll, G. (2019a). "Learning to Reconstruct People in Clothing from a Single RGB Camera," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, United States, June 15–20, 2019, 1175–1186. doi:10.1109/cvpr.2019.00127

The other raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access the datasets should be directed to andrea.bartl@uni-wuerzburg.de.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AB, SW, MB, and ML contributed to conception and design of the study. AB performed the statistical analysis. SW and MB provided the avatar reconstruction framework. EW provided the avatar embodiment framework. AB wrote the first draft of the manuscript. EW and SW wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This publication was supported by the Open-Access Publication Fund of the University of Würzburg. This research has been funded by the German Federal Ministry of Education and Research in the project ViTraS (project number 16SV8219), the Bavarian State Ministry for Science and Art within the Innovationlab project, and the Bavarian State Ministry for Digital Affairs within the XR Hub Würzburg.

ACKNOWLEDGMENTS

We thank Larissa Brübach, Franziska Westermeier, and Bianka Weisz who contributed to the development of the virtual environment. In addition, we thank Matthias Schmid, Philipp Schaupp, and Yousof Shehada for their help in conducting the experiment.

- Alldieck, T., Magnor, M., Xu, W., Theobalt, C., and Pons-Mol, G. (2018a). Video Based Reconstruction of 3d People Models. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, United States, June 18–23, 2018. 8387–8397. doi:10.1109/cvpr.2018.00875
- Alldieck, T., Magnor, M., Xu, W., Theobalt, C., and Pons-Moll, G. (2018b). “Detailed Human Avatars from Monocular Video,” in Proc. of International Conference on 3D Vision, 98–109. doi:10.1109/3dv.2018.00022
- Alldieck, T., Pons-Moll, G., Theobalt, C., and Magnor, M. (2019b). “Tex2shape: Detailed Full Human Body Geometry from a Single Image,” in Proc. of IEEE International Conference on Computer Vision, Seoul, South Korea, October 27–November 2, 2019, 2293–2303. doi:10.1109/iccv.2019.00238
- [Dataset] Autodesk (2014). Character Generator. Available at: <https://charactergenerator.autodesk.com/> (Accessed August 30, 2021).
- Aymerich-Franch, L., Kizilcec, R. F., and Bailenson, J. N. (2014). “The Relationship between Virtual Self Similarity and Social Anxiety,” 944. doi:10.3389/fnhum.2014.00944. *Front. Hum. Neurosci.* 8
- Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and Blascovich, J. (2005). The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments. *Presence: Teleoperators & Virtual Environments* 14, 379–393. doi:10.1162/105474605774785235
- Banakou, D., Groten, R., and Slater, M. (2013). Illusory Ownership of a Virtual Child Body Causes Overestimation of Object Sizes and Implicit Attitude Changes. *Proc. Natl. Acad. Sci.* 110, 12846–12851. doi:10.1073/pnas.1306779110
- Banakou, D., Hanumanthu, P. D., and Slater, M. (2016). Virtual Embodiment of white People in a Black Virtual Body Leads to a Sustained Reduction in Their Implicit Racial Bias. *Front. Hum. Neurosci.* 10, 601. doi:10.3389/fnhum.2016.00601
- Bouzaz, S., Tagliasacchi, A., and Pauly, M. (2014). Dynamic 2D/3D Registration. *Eurographics Tutorials* 1–17.
- Bouchard, S., St-Jacques, J., Robillard, G., and Renaud, P. (2008). Anxiety Increases the Feeling of Presence in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 17, 376–391. doi:10.1162/pres.17.4.376
- Brenton, H., Gillies, M., Ballin, D., and Chatting, D. (2005). “The Uncanny valley: Does it Exist,” in Proceedings of conference of human computer interaction, workshop on human animated character interaction, Edinburgh, United Kingdom, September 5–9, 2005 (London: Springer).
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: (Academic Press, Inc).
- de Rooij, A., van der Land, S., and van Erp, S. (2017). “The Creative proteus Effect: How Self-Similarity, Embodiment, and Priming of Creative Stereotypes with Avatars Influences Creative Ideation,” in Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition, Singapore, Singapore, June 27–30, 2017, 232–236.
- Döllinger, N., Wiernich, C., Wolf, E., and Latoschik, M. E. (2019). “Vitrax - Virtual Reality Therapy by Stimulation of Modulated Body Image - Project Outline,” in Mensch und Computer 2019 - Workshopband (Bonn: Gesellschaft für Informatik e.V.), 606–611.
- Dubuisson, M.-P., and Jain, A. K. (1994). “A Modified Hausdorff Distance for Object Matching,” in Proceedings of International Conference on Pattern Recognition, Jerusalem, Israel, October 9–13, 1994, 566–568.
- [Dataset] Epic Games (2021). Metahuman Creator. Available at: <https://www.unrealengine.com/en-US/metahuman-creator> (Accessed August 30, 2021).
- Feng, A., Rosenberg, E. S., and Shapiro, A. (2017). Just-in-time, Viable, 3-d Avatars from Scans. *Computer Animation Virtual Worlds* 28, 3–4. doi:10.1002/cav.1769
- Fox, J., Bailenson, J. N., and Tricase, L. (2013). The Embodiment of Sexualized Virtual Selves: the proteus Effect and Experiences of Self-Objectification via Avatars. *Comput. Hum. Behav.* 29, 930–938. doi:10.1016/j.chb.2012.12.027
- Freeman, G., and Maloney, D. (2021). Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* vol. 4, 1–27. doi:10.1145/3432938
- Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M. A. (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 SIGCHI conference on Human factors in computing systems, Ft. Lauderdale, FL, United States, April 5–10, 2003, 529–536. doi:10.1145/642611.642703
- Gonzalez-Franco, M., and Lanier, J. (2017). Model of Illusions and Virtual Reality. *Front. Psychol.* 8, 1125. doi:10.3389/fpsyg.2017.01125
- Gonzalez-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., et al. (2020). The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Front. Virtual Real.* 1, 20. doi:10.3389/frvir.2020.561558
- Gorisse, G., Christmann, O., Houzangbe, S., and Richir, S. (2019). From Robot to Virtual Doppelgänger: Impact of Visual Fidelity of Avatars Controlled in Third-Person Perspective on Embodiment and Behavior in Immersive Virtual Environments. *Front. Robot. AI* 6, 8. doi:10.3389/frbot.2019.00008
- Guo, K., Lincoln, P., Davidson, P., Busch, J., Yu, X., Whalen, M., et al. (2019). The Relightables. *ACM Trans. Graph.* 38, 1–19. doi:10.1145/3355089.3356571
- He, D., Liu, F., Pape, D., Dawe, G., and Sandin, D. (2000). “Video-based Measurement of System Latency,” in International Immersive Projection Technology Workshop, Ames, IA, United States, June 19–20, 2000.
- Ho, C.-C., and MacDorman, K. F. (2017). Measuring the Uncanny valley Effect. *Int. J. Soc. Robotics* 9, 129–139. doi:10.1007/s12369-016-0380-9
- Ho, C.-C., MacDorman, K. F., and Pramono, Z. D. (2008). “Human Emotion and the Uncanny valley: a Glm, Mds, and Isomap Analysis of Robot Video Ratings,” in Proc. of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Amsterdam, Netherlands, March 12–15, 2008 (IEEE), 169–176.
- Ho, C.-C., and MacDorman, K. F. (2010). Revisiting the Uncanny valley Theory: Developing and Validating an Alternative to the Godspeed Indices. *Comput. Hum. Behav.* 26, 1508–1518. doi:10.1016/j.chb.2010.05.015
- Hudson, I., and Hurter, J. (2016). “Avatar Types Matter: Review of Avatar Literature for Performance Purposes,” in International conference on virtual, augmented and mixed reality, Toronto, Canada, July 17–22, 2016 (Springer), 14–21. doi:10.1007/978-3-319-39907-2_2
- Ichim, A. E., Bouaziz, S., and Pauly, M. (2015). “Dynamic 3d Avatar Creation from Hand-Held Video Input,” in ACM Transactions on Computer Graphics. 34, 45: 1–45:14. doi:10.1145/2766974
- Jo, D., Kim, K.-H., and Kim, G. J. (2017). “Effects of Avatar and Background Types on Users’ Co-presence and Trust for Mixed Reality-Based Teleconference Systems,” in Proceedings the 30th Conference on Computer Animation and Social Agents, Seoul, South Korea, May 22–24, 2017, 27–36.
- Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (2015). A Review of Empirical Evidence on Different Uncanny valley Hypotheses: Support for Perceptual Mismatch as One Road to the valley of Eeriness. *Front. Psychol.* 6, 390. doi:10.3389/fpsyg.2015.00390
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *Int. J. Aviation Psychol.* 3, 203–220. doi:10.1207/s15327108jap0303_3
- Kiltner, K., Groten, R., and Slater, M. (2012). The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 373–387. doi:10.1162/pres_a_00124
- Kim, H. K., Park, J., Choi, Y., and Choe, M. (2018). Virtual Reality Sickness Questionnaire (Vrsq): Motion Sickness Measurement index in a Virtual Reality Environment. *Appl. Ergon.* 69, 66–73. doi:10.1016/j.apergo.2017.12.016
- Kocur, M., Kloss, M., Schwind, V., Wolff, C., and Henze, N. (2020). “Flexing Muscles in Virtual Reality: Effects of Avatars’ Muscular Appearance on Physical Performance,” in Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Virtual Event, Canada, November 2–4, 2020, 193–205. doi:10.1145/3410404.3414261
- Latoschik, M. E., Kern, F., Stauffert, J.-P., Bartl, A., Botsch, M., and Lugrin, J.-L. (2019). Not alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality. *IEEE Trans. Vis. Comput. Graphics* 25, 2134–2144. doi:10.1109/tvcg.2019.2899250
- Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., and Botsch, M. (2017). The Effect of Avatar Realism in Immersive Social Virtual Realities. *Proc. ACM Symp. Virtual Reality Softw. Techn.* 39, 1–39. doi:10.1145/3139131.3139156
- Latoschik, M. E., and Wiernich, C. (2021). Coherence and Plausibility, Not Presence?! Pivotal Conditions for Xr Experiences and Effects, a Novel Model. arXiv

- Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., and Black, M. J. (2015). Smpl: A Skinned Multi-Person Linear Model. *ACM Trans. Comput. Graphics* 34, 248: 1–248:16. doi:10.1145/2816795.2818013
- Lugrin, J.-L., Landeck, M., and Latoschik, M. E. (2015c). “Avatar Embodiment Realism and Virtual Fitness Training,” in 2015 IEEE Virtual Reality (VR), Arles, France, March 23–27, 2015 (IEEE), 225–226. doi:10.1109/vr.2015.7223377
- Lugrin, J.-L., Latt, J., and Latoschik, M. E. (2015a). “Avatar Anthropomorphism and Illusion of Body Ownership in Vr,” in Proceedings of the IEEE Annual International Symposium Virtual Reality, Arles, France, March 23–27, 2015, 229–230. doi:10.1109/VR.2015.7223379
- Lugrin, J.-L., Wiedemann, M., Bieberstein, D., and Latoschik, M. E. (2015b). “Influence of Avatar Realism on Stressful Situation in Vr,” in Proceedings of the IEEE Annual International Symposium Virtual Reality, Arles, France, March 23–27, 2015 (IEEE), 227–228. doi:10.1109/vr.2015.7223378
- Magenat-Thalman, N., and Thalman, D. (2005). Virtual Humans: Thirty Years of Research, what Next?. *Vis. Comput* 21, 997–1015. doi:10.1007/s00371-005-0363-6
- McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render Me Real?. *ACM Trans. Graph.* 31, 1–11. doi:10.1145/2185520.2185587
- Mori, M., MacDorman, K., and Kageki, N. (2012). The Uncanny valley [from the Field]. *IEEE Robot. Automat. Mag.* 19, 98–100. doi:10.1109/mra.2012.2192811
- Nelson, M. G., Mazumdar, A., Jamal, S., Chen, Y., and Mousas, C. (2020). “Walking in a Crowd Full of Virtual Characters: Effects of Virtual Character Appearance on Human Movement Behavior,” in International Symposium on Visual Computing, San Diego, CA, United States, October 5–7, 2020 (Springer), 617–629. doi:10.1007/978-3-030-64556-4_48
- Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting Yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias. *Conscious. Cogn.* 22, 779–787. doi:10.1016/j.concog.2013.04.016
- Peña, J., Khan, S., and Alexopoulos, C. (2016). I Am what I See: How Avatar and Opponent Agent Body Size Affects Physical Activity Among Men Playing Exergames. *J. Comput-mediat Comm.* 21, 195–209. doi:10.1111/jcc4.12151
- Praetorius, A. S., and Görlich, D. (2020). “How Avatars Influence User Behavior: A Review on the proteus Effect in Virtual Environments and Video Games,” in International Conference on the Foundations of Digital Games, Bugibba, Malta, September 15–18, 2020, 1–9.
- Ratan, R., Beyea, D., Li, B. J., and Graciano, L. (2020). Avatar Characteristics Induce Users’ Behavioral Conformity with Small-To-Medium Effect Sizes: a Meta-Analysis of the proteus Effect. *Media Psychol.* 23, 651–675. doi:10.1080/15213269.2019.1623698
- [Dataset] Rootmotion (2020). FinalIK. Available at: <http://root-motion.com> (Accessed August 30, 2021).
- Roth, D., and Latoschik, M. E. (2020). Construction of the Virtual Embodiment Questionnaire (Veq). *IEEE Trans. Vis. Comput. Graphics* 26, 3546–3556. doi:10.1109/tvcg.2020.3023603
- Roth, D., Lugrin, J.-L., Latoschik, M. E., and Huber, S. (2017). “Alpha Ivbo – Construction of a Scale to Measure the Illusion of Virtual Body Ownership,” in Proc. of CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, United States, May 6–11, 2017, 2875–2883.
- Salomoni, P., Prandi, C., Rocchetti, M., Casanova, L., and Marchetti, L. (2016). “Assessing the Efficacy of a Diegetic Game Interface with Oculus Rift,” in Proc. of the 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, United States, January 9–12, 2016 (IEEE), 387–392. doi:10.1109/ccnc.2016.7444811
- Seymour, M., Lingyao, Y., Dennis, A., Riemer, K., et al. (2019). “Crossing the Uncanny valley? Understanding Affinity, Trustworthiness, and Preference for More Realistic Virtual Humans in Immersive Environments,” in Proceedings of the 52nd Hawaii International Conference on System Sciences, Wailea, HI, United States, January 8–11, 2019, 1748–1758.
- Seymour, M., Riemer, K., and Kay, J. (2017). “Interactive Realistic Digital Avatars- Revisiting the Uncanny valley,” in Proceedings of the 50th Hawaii International Conference on System Sciences, Wailea Village, HI, United States, January 4–7, 2017, 547–556. doi:10.24251/hicss.2017.067
- Shafer, D. M., Carbonara, C. P., and Korpi, M. F. (2017). Modern Virtual Reality Technology: Cybersickness, Sense of Presence, and Gender. *Media Psychol. Rev.* 11, 1.
- Skarbez, R., Brooks, F. P., Jr, and Whitton, M. C. (2017). A Survey of Presence and Related Concepts. *ACM Comput. Surv. (Csur)* 50, 1–39. doi:10.1145/3135069
- Slater, M., Gonzalez-Liencre, C., Haggard, P., Vinkers, C., Gregory-Clarke, R., Jelley, S., et al. (2020). The Ethics of Realism in Virtual and Augmented Reality. *Front. Virtual Real.* 1, 1. doi:10.3389/frvir.2020.00001
- Slater, M. (2009). Place Illusion and Plausibility Can lead to Realistic Behaviour in Immersive Virtual Environments. *Phil. Trans. R. Soc. B* 364, 3549–3557. doi:10.1098/rstb.2009.0138
- Slater, M., Spanlang, B., Sanchez-Vives, M. V., and Blanke, O. (2010). First Person Experience of Body Transfer in Virtual Reality. *PLOS ONE* 5, e10564. doi:10.1371/journal.pone.0010564
- Steed, A., and Schroeder, R. (2015). “Collaboration in Immersive and Non-immersive Virtual Environments,” in *Immersed in Media* (Springer), 263–282. doi:10.1007/978-3-319-10190-3_11
- Thaler, A., Geuss, M. N., Mölbert, S. C., Giel, K. E., Streuber, S., Romero, J., et al. (2018). Body Size Estimation of Self and Others in Females Varying in Bmi. *PLoS one* 13, e0192152. doi:10.1371/journal.pone.0192152
- Tinwell, A., and Grimshaw, M. (2009). “Bridging the Uncanny: an Impossible Traverse?,” in Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era, Tampere, Finland, September 30–October 2, 2009, 66–73.
- [Dataset] Unity Technologies (2019). Unity. Available at: <https://unity3d.com> (Accessed August 30, 2021).
- [Dataset] Valve Corporation (2020). Index. Available at: <https://store.steampowered.com/valveindex> (Accessed August 30, 2021).
- [Dataset] Valve (2020). SteamVR. Available at: <https://store.steampowered.com/steamvr> (Accessed August 30, 2021).
- Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Trans. Vis. Comput. Graphics* 24, 1643–1652. doi:10.1109/tvcg.2018.2794629
- Waltemate, T., Senna, I., Hülsmann, F., Rohde, M., Kopp, S., Ernst, M., and Botsch, M. (2016). “The Impact of Latency on Perceptual Judgments and Motor Performance in Closed-Loop Interaction in Virtual Reality,” in Proceedings of the 22nd ACM Symposium on Virtual Reality Software and Technology, Munich, Germany, November 2–4, 2016, 27–35. doi:10.1145/2993369.2993381
- Wang, S., Lilienfeld, S. O., and Rochat, P. (2015). The Uncanny valley: Existence and Explanations. *Rev. Gen. Psychol.* 19, 393–407. doi:10.1037/gpr0000056
- Weng, C.-Y., Curless, B., and Kemelmacher-Shlizerman, I. (2019). “Photo Wake-Up: 3d Character Animation from a Single Photo,” in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, United States, June 15–20, 2019, 5908–5917. doi:10.1109/cvpr.2019.00606
- Weninger, S., Achenbach, J., Bartl, A., Latoschik, M. E., and Botsch, M. (2020). “Realistic Virtual Humans from Smartphone Videos,” in Proc. of the 26th ACM Symposium on Virtual Reality Software and Technology (Association for Computing Machinery), Canada, November 1–4, 2020, 1–11. VRST ’20. doi:10.1145/3385956.3418940
- Witmer, B. G., and Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence* 7, 225–240. doi:10.1162/105474698565686
- Wolf, E., Döllinger, N., Mal, D., Wienrich, C., Botsch, M., and Latoschik, M. E. (2020). “Body Weight Perception of Females Using Photorealistic Avatars in Virtual and Augmented Reality,” in Proc. of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Porto de Galinhas, Brazil, November 9–13, 2020, 462–473. doi:10.1109/ISMAR50242.2020.00071
- Wolf, E., Merdan, N., Döllinger, N., Mal, D., Wienrich, C., Botsch, M., et al. (2021). “The Embodiment of Photorealistic Avatars Influences Female Body Weight Perception in Virtual Reality,” in Proceedings of the 28th IEEE Virtual Reality Conference (VR ’21), Lisboa, Portugal, March 27–April 1, 2021. doi:10.1109/vr50410.2021.00027
- Yee, N., and Bailenson, J. N. (2006). Walk a Mile in Digital Shoes: The Impact of Embodied Perspective-Taking on the Reduction of Negative Stereotyping in Immersive Virtual Environments. *Proc. PRESENCE* 24, 26.

- Yee, N., and Bailenson, J. (2007). The proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Hum. Comm Res* 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x
- Zibrek, K., Kokkinara, E., and McDonnell, R. (2018). The Effect of Realistic Appearance of Virtual Characters in Immersive Environments - Does the Character's Personality Play a Role?. *IEEE Trans. Vis. Comput. Graphics* 24, 1681–1690. doi:10.1109/tvcg.2018.2794638
- Zibrek, K., Martin, S., and McDonnell, R. (2019). Is Photorealism Important for Perception of Expressive Virtual Humans in Virtual Reality?. *ACM Trans. Appl. Percept.* 16, 1–19. doi:10.1145/3349609

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bartl, Wenninger, Wolf, Botsch and Latoschik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.