

Hartbrich, Jakob; Weidner, Florian; Kunert, Christian; Arévalo Arboleda, Stephanie; Raake, Alexander; Broll, Wolfgang

Eye and face tracking in VR: avatar embodiment and enfacement with realistic and cartoon avatars

Original published in: MUM 2023. - New York, NY : Association for Computing Machinery. - (2023), p. 270-278.

Conference: International Conference on Mobile and Ubiquitous Multimedia; 22 (Vienna) : 2023.12.03-06

Original published: 2023-12-03

ISBN: 979-8-4007-0921-0

DOI: [10.1145/3626705.3627793](https://doi.org/10.1145/3626705.3627793)

[Visited: 2024-08-01]



This work is licensed under a [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>



Eye and Face Tracking in VR: Avatar Embodiment and Enfacement with Realistic and Cartoon Avatars

Jakob Hartbrich

jakob.hartbrich@tu-ilmenau.de
Audiovisual Technology Group,
Technische Universität Ilmenau
Ilmenau, Germany

Florian Weidner

florian.weidner@tu-ilmenau.de
Virtual Worlds and Digital Games
Group, Technische Universität
Ilmenau and Lancaster University,
United Kingdom
Ilmenau, Germany

Christian Kunert

christian.kunert@tu-ilmenau.de
Virtual Worlds and Digital Games
Group, Technische Universität
Ilmenau
Ilmenau, Germany

Stephanie Arévalo Arboleda

stephanie.arevalo@tu-ilmenau.de
Audiovisual Technology Group,
Technische Universität Ilmenau
Ilmenau, Germany

Alexander Raake

alexander.raake@tu-ilmenau.de
Audiovisual Technology Group,
Technische Universität Ilmenau
Ilmenau, Germany

Wolfgang Broll

wolfgang.broll@tu-ilmenau.de
Virtual Worlds and Digital Games
Group, Technische Universität
Ilmenau
Ilmenau, Germany

ABSTRACT

Previous studies have explored the perception of various types of embodied avatars in immersive environments. However, the impact of eye and face tracking with personalized avatars is yet to be explored. In this paper, we investigate the impact of eye and face tracking on embodiment, enfacement, and the uncanny valley with four types of avatars using a VR-based mirroring task. We conducted a study (N=12) and created self-avatars with two rendering styles: a cartoon avatar (created in an avatar generator using a picture of the user's face) and a photorealistic scanned avatar (created using a 3D scanner), each with and without eye and face tracking and respective adaptation of the mirror image. Our results indicate that adding eye and face tracking can be beneficial for certain enfacement scales (*belonged*), and we confirm that compared to a cartoon avatar, a scanned realistic avatar results in higher body ownership and increased enfacement (*own face, belonging, mirror*) – regardless of eye and face tracking. We critically discuss our experiences and outline the limitations of the applied hardware and software with respect to the provided level of control and the applicability for complex tasks such as displaying emotions. We synthesize these findings into a discussion about potential improvements for facial animation in VR and highlight the need for a better level of control, the integration of additional sensing and processing technologies, and an objective metric for comparing facial animation systems.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; • **Computing methodologies** → *Motion capture*.



This work is licensed under a Creative Commons Attribution International 4.0 License.

MUM '23, December 03–06, 2023, Vienna, Austria
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0921-0/23/12.
<https://doi.org/10.1145/3626705.3627793>

KEYWORDS

Virtual reality, avatar, face tracking, eye tracking, embodiment, user study

ACM Reference Format:

Jakob Hartbrich, Florian Weidner, Christian Kunert, Stephanie Arévalo Arboleda, Alexander Raake, and Wolfgang Broll. 2023. Eye and Face Tracking in VR: Avatar Embodiment and Enfacement with Realistic and Cartoon Avatars. In *International Conference on Mobile and Ubiquitous Multimedia (MUM '23)*, December 03–06, 2023, Vienna, Austria. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3626705.3627793>

1 INTRODUCTION

The face is an integral part of human communication. It not only allows us to speak but also enables us to use nonverbal cues to express ourselves. Unfortunately, many of these cues are lost in digitally mediated avatar-based communication. Some studies and applications integrate simple eye and lip movements to reintroduce these cues. For example, Wu et al. [54] concluded that interacting with highly expressive avatars resulted in significantly higher social presence and interpersonal attraction. This becomes more important as today's avatars in immersive environments are rarely realistic but often stylized or cartoon-like. However, as outlined in the review by Weidner et al. [52], a realistic rendering style is often perceived as better than other styles regarding several variables, such as body ownership or social presence. Still, it is unclear if this holds true for the integration of eye and face tracking, as research has also shown that the tracking quality [5], visualization quality [8], and a combination of both [17, 37] may influence a variety of dependent variables such as social presence [4] and body ownership [21]. While a lot of prior research has investigated these variables in studies and compared a variety of avatars, studies on the above-mentioned issue – the impact of eye and face tracking – are sparse, with only a few studies that combined full motion, eye, and face tracking (for an overview, see Weidner et al. [52]). Thus, the interaction between avatar style and eye and face tracking remains unexplored despite eye tracking being in many commercial HMDs, cartoon avatars being the norm in many VR applications,

and realistic avatars (e.g., Lugin et al. [30]) as well as face tracking being a research topic (e.g., Ma and Pan [32]). However, knowing the relationship between eye and face tracking and avatar realism could inform the design of future systems and highlight important aspects of avatar-mediated communication.

In this work, we aim to close this gap by investigating the impact of eye and face tracking on body ownership, enfacement, and the uncanny valley effect with two types of personalized self-avatars. The first type of avatar is a personalized cartoon avatar created with *Ready Player Me* [53], a state-of-the-art commercial avatar generator using a picture of the user's face. The second type of avatar is a personalized photorealistic 3D scanned avatar. Each type of avatar was evaluated with and without eye and face tracking capabilities. We assessed embodiment, the uncanny valley effect, and the enfacement illusion.

The core contributions of this work are the results of our comparative user study as well as a discussion of the current limitations of camera-based eye and face animation when displaying complex emotions. Further, a set of research directions is presented aimed at overcoming these issues, including physically inspired models, data fusion, additional sensing technologies, and objective metrics to compare facial animation systems.

2 RELATED WORK

In the following, relevant related work is briefly discussed that investigates the perception of different types of avatars, considering respective perceptual and cognitive constructs. The section is split into parts addressing embodiment, enfacement, eye and face tracking, and the uncanny valley.

2.1 Embodiment

Embodiment or the body ownership illusion is the feeling of owning an avatar [26, 48]. This important aspect of VR is often separated into the dimensions of ownership, agency, and self-location [26]. In general, avatars and systems that score higher on these dimensions lead to a higher feeling of embodiment [52]. Several other factors contribute to a high feeling of embodiment, such as personalization [51] and appearance [30] but also high tracking fidelity [19], as well as coherence between tracking fidelity and visual appearance [47].

Interestingly and most likely due to the limited availability of proper hardware systems, the majority of prior studies investigating embodiment (even those emphasizing the need for high tracking fidelity) did not integrate eye and face tracking in their setup. However, eye and face tracking arguably increases tracking fidelity and could also improve the feeling of embodiment. We set out to investigate this aspect.

2.2 Enfacement

As a special case of embodiment, *enfacement* describes the phenomenon that the internal self-face representation can be updated with information from the face of another entity [40]. Gonzalez-Franco et al. [20] hypothesize that enfacement and embodiment are closely linked, implying that achieving enfacement may not be feasible unless the avatar is properly embodied. Interestingly, research has also shown that animation realism is not necessarily an essential

factor when it comes to enfacement, whereas agency [22] and synchronicity [50] seem to be. Indeed, in the experiment of Ma et al. [32], participants felt more in control over their face with a cartoon avatar compared to a realistic avatar — however, they used generic and not personalized avatars. Following up on this, Salagean et al. [44] demonstrated the benefits of photorealism and personalization on self-avatars' self-identification in virtual reality (although without face tracking). These seemingly contradicting results show that the interdependency of enfacement and eye- and face animations remain unexplored. We take up this gap and integrate eye and face tracking as well as personalization in our work.

2.3 Eye and Face Tracking in VR for Embodiment and Enfacement

Eye and face tracking allows VR developers to apply the user's facial features to an avatar. Recent advances in hard- and software allowed researchers to integrate eye and face tracking into their research with high-fidelity avatars. It has been shown that both potentially increase embodiment. For example, Gonzalez-Franco et al. [20] highlight facial animation's importance for enfacement and indicate that predefined animations are not necessarily worse than pure lip synchronization (as they focused on the mouth, their experiment did not feature eye tracking). Contrary to that, using only eye tracking and no face tracking, Borland et al. [7] demonstrated that incorporating realistic eye animations in an avatar increases self-identification compared to having no eye animations. In a study similar to ours, Ma and Pan [32] analyzed the effect of visual fidelity (personalized cartoon vs. personalized realistic) on expressive avatars with face tracking (always on). Interestingly, the different avatars did not lead to differences in embodiment. We expand on these studies by investigating if eye and face tracking in combination influence embodiment.

Overall, research highlights the importance of eye and face tracking (individually) for embodiment and enfacement in VR. Still, the question arises of how eye and face tracking (together) interplay with the avatar's degree of realism. This is important as several authors highlight the need for congruency in avatar-mediated applications to perform best [5, 10]: For example, Mottelson et al. [35] emphasize that incongruent stimuli lower the sense of embodiment. This could mean that, for example, an avatar with very realistic facial tracking but an unrealistic look (or vice versa) might lead to worse results in embodiment and enfacement than an avatar with low eye and face tracking capabilities and an unrealistic realistic look. This might be related to the uncanny valley.

2.4 Uncanny Valley

The uncanny valley effect refers to a dip in affinity with increasing humanlikeness before a steep increase [34]. While the uncanny valley has often been shown, a clear explanation of factors is still missing from research [12, 32]. Relevant for this study is that avatar realism seems to be a contributing factor (as the theory states), and a more realistic avatar can lead to a stronger negative effect of the uncanny valley [29, 31, 32]. Further, especially the face's upper part seems to be a critical aspect, as outlined by Tinwell et al. [49]. Ma and Pan [32] showed that there is an interaction between visual and behavioral fidelity that influences the uncanny valley. Our study

includes the uncanny valley questionnaire by Ho and MacDorman [23] to control for this effect.

3 USER STUDY

Overall, the related work shows that both avatar realism, eye tracking, and face tracking individually influence body ownership and enfacement. However, the relationship between avatar realism and eye and face tracking remains unexplored. In this work, we build upon prior findings and investigate the interdependency of both factors on avatar’s rendering style in a multi-factorial experiment.

To investigate the interplay between eye and face tracking as well as embodiment and enfacement, we present a study that evaluates the difference between a personalized cartoon avatar and a scanned realistic avatar with enabled and disabled eye and face tracking. The main research question guiding this study is: “What is the influence of eye and face tracking on embodiment and enfacement when applied to a cartoon and a scanned realistic avatar?” with the following detailed research hypotheses:

- H1: Eye and face tracking will lead to higher scores in embodiment for both avatars.
- H2: Eye and face tracking will lead to higher scores in enfacement for both avatars.
- H3: The realistic avatar will lead to higher embodiment scores than the personalized cartoon avatar.

3.1 Experimental Design

The study followed a within-subject design with avatar representation (cartoon, realistic) and eye and face tracking (enabled, disabled) as independent variables. Each participant tested the two different avatars once with enabled eye and face tracking and once without it, resulting in four conditions:

- Personalized cartoon avatar (*Cartoon.noEFT*)
- Personalized cartoon avatar + eye and face tracking (*Cartoon.EFT*)
- Scanned realistic avatar (*Scanned.noEFT*)
- Scanned realistic avatar + eye and face tracking (*Scanned.EFT*)

3.2 Apparatus

The VR project used for this study was created with *Unity* (2021.2.10f1, c.f. Appendix A). The virtual environment was designed to resemble the laboratory where the study was performed. We added a virtual mirror so that the participants could look at their avatar (see Figure 1).

For both avatars, eye and face tracking was implemented using the *SRanipal SDK* (version 1.3.6.6) software with the *HTC Vive Pro Eye* and the *HTC Vive Facial Tracker. Motive:Body 2.1.2* was used to track the participant’s body movements. Because the motion tracking suit available at the time of this study did not include the tracking of individual fingers, only the whole hand was tracked and presented rigid in a neutral pose. We carried out full-body scans to avoid disrupting self-recognition, Bulthoff et al. [9] showed the benefits of using full-body avatars instead of only heads when investigating face recognition.

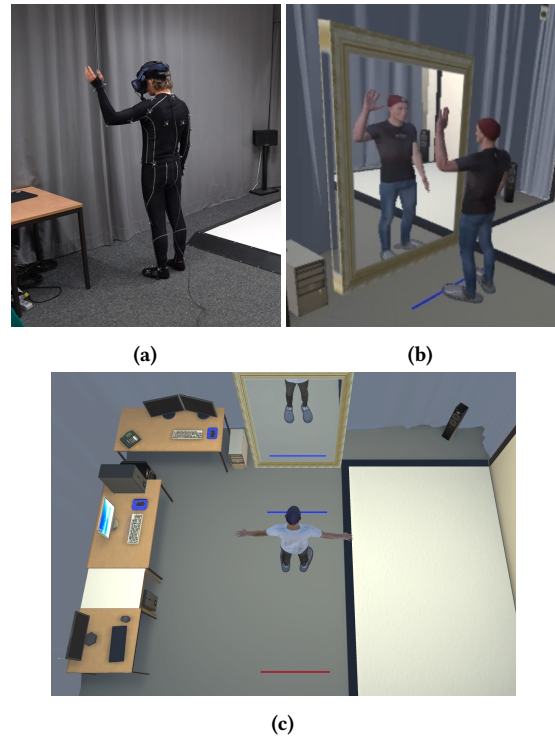


Figure 1: Comparison of the real laboratory (a) and the virtual environment (b) with the added virtual mirror. Overview of the whole virtual laboratory (c).

3.3 Avatars

Figure 2 shows exemplary avatars generated with *Ready Player Me* and the 3D scanning procedure. In the following, both methods are described in more detail.

3.3.1 Scanned Avatars. In order to create scanned realistic avatars of our participants, we used the *Artec Leo* 3D scanner [2]. As the scanner presented problems scanning reflective objects and fine structures such as hair and beards, participants were asked to remove any wristwatches and other shiny jewelry and wore a beanie to cover their hair to prevent scanning errors. Participants were scanned using an A-pose and were instructed to wear tight-fitting clothes for the session. The goal was to represent their body shape in the scan accurately.

The resulting point cloud from the scanned data was processed in the *Artec Studio 16* [3] software for removing scan errors and generating a coherent mesh. Next, to reduce mesh complexity, a retopology of the mesh was performed with *R3DS Wrap* [43]. This led to a reduction from 300,000 vertices to 16,000 vertices. The retopologized version was morphed into a *Genesis8 character* from *DAZ Studio* [41]. By that, we created a full rig for animation and got fully modeled and separated eyes, mouth with tongue, and teeth for animation. The resulting model was also equipped with multiple blendshapes for facial animation. However, we had to create additional blendshapes with the *Blender* plugin *Faceit* [15] (version 2.0.20), resulting in a total of 43 blendshapes. The list of



Figure 2: Two female avatars from our study. Left in both (a) and (b) are scanned avatars. Right in both (a) and (b) are the Ready Player Me avatars.

blendshapes tracked can be found in Appendix A. Compared to state of the art rendering and realism like MetaHumans [18] our scanned avatars offer realistic scale and proportions of the participant with high resolution textures, but lack high quality skin materials as well as detailed hair.

3.3.2 Cartoon Avatars. We used *Ready Player Me*[53] to create the personalized cartoon avatar based on a portrait of the participant. As the avatar generator did not offer enough customization to recreate the outfits of all participants, all cartoon avatars were created with neutral clothing (jeans and a white t-shirt). Next to this difference, the avatars also differed in hairstyle, as the scanned realistic avatar was always depicted with a beanie. The chosen hairstyle was based on the portrait of the participant and adjusted to fit their hair color as closely as possible. To enable the same level of facial animation, we created 43 blendshapes with the *Faceit Blender* plugin [15].

In the end, both the cartoon avatar and the scanned realistic avatar have a full-body skeleton for motion tracking as well as the same blendshapes for eye and face tracking.

3.4 Procedure

We divided the experiment into two sessions: a first session to create participants' avatars and a second session on another day with the laboratory user study. The first session lasted 15 minutes. The second session took around 75 minutes.

3.4.1 Session 1: Avatar Creation. On the first appointment, participants signed the consent form. After that, we created the avatars (c.f. subsection 3.2). Additionally, participants' height was measured to scale their avatars. Participants did not see their avatars until the second appointment.

3.4.2 Session 2: Laboratory Experiment. At the beginning of the second session, participants read the information about the procedure and gave informed consent. Subsequently, they put on the Optitrack suit, and calibration was performed. Next, they completed all four avatar conditions, following a Partial Latin Square format.

When participants put on the VR HMD, they were in a virtual replica of the laboratory (see Figure 1). Inspired by fake mirrors previously used to investigate the perception of self-avatars [28], we placed a virtual mirror in the VR environment. At the start of each condition, the virtual mirror and avatar were turned off to allow for acclimatization to the virtual environment and to control the exposure time of the stimulus. Eye tracking calibration

Enfacement	Questions
Belonged	I felt like the virtual face belonged to me.
(out of) control	I felt like my own face was out of my control.
Mirror	I felt like I was looking at my own face reflected in a mirror.
Ownface	I felt like the virtual face was my face.

Table 1: Additional Enfacement Questions.

was performed beforehand if the condition featured eye and face tracking. Next, the mirror and avatar were turned on, and the investigator started reading movement instructions to the participant. The movement-related instructions were: “Stand on the blue line and look in the direction of the mirror.”, “Raise your right hand and wave at your reflection in a relaxed manner.”, “Repeat with your left hand.”, “Now walk on the spot and lift your legs waist high.”, “Now turn right and walk to the red line.”, “Turn around to the left. Go back to the blue line.”, “Stretch your arms out in front of you and turn your palms.”, and “Stretch your right arm out to the side and move it in a circular motion.”, “Repeat with your left arm.”

Each of the movement instructions was followed by an instruction to look back and forth between the mirror and their own body (“Look at the movement in the mirror... on your own body | in the mirror | on your own body.”). This approach aimed for all participants to perform the same movements and register the coherence between their mirrored avatar and their own body [51].

After performing the above motions, participants were told to show the six core emotions: anger, disgust, enjoyment, fear, sadness, and surprise [13]. While performing these emotions, they were instructed to watch themselves in the mirror. The order of emotions was randomized.

After each condition, the participants were asked to remove the HMD and fill in the questionnaires (uncanny valley, embodiment, and enfacement). After all conditions were concluded, participants filled in a final demographic questionnaire. While taking off the motion capture suit, they were given the possibility to provide verbal feedback.

3.5 Measures

For each avatar representation, we collected subjective impressions through questionnaires. To evaluate *embodiment*, we used the questionnaire from Peck and Gonzalez-Franco [39]. The questionnaire uses a 7-point Likert scale for all questions. Participants answered this questionnaire for each of the four avatar conditions.

Related to embodiment but focused on the face, we asked four additional *enfacement* questions (see Table 1) from Estudillo and Bindemann [14].

The questionnaire uses a 7-point Likert scale from *strongly disagree* (-3) to *strongly agree* (3). These questions were also answered after each avatar condition.

Again, after each condition, participants were asked to answer 18 questions from Ho and MacDorman's *uncanny valley* questionnaire [23]. The questionnaire focuses on four indices: perceived humanness, attractiveness, spine-tingling, and eeriness. Each question is answered on a 7-point Likert scale from -3 to 3.

Finally, the participants were asked demographic questions about their age, education, and how much VR experience they have.

3.6 Participants

Twelve participants were recruited for this study, eight self-identifying as male and four as female, with ages ranging from 22 to 29 years ($M=24, SD=1.83$). All were university students recruited via the public students' mailing list. Ten had used an HMD for up to three hours, one had used an HMD for more than three hours, and one had no prior VR experience. Participants gave their informed consent for participation in the study; participation was voluntary without monetary compensation. The study was executed following the guidelines of the local university, the national research organization, and the declaration of Helsinki. We did not undertake any specific measures for sample diversity.

4 RESULTS

All questionnaires were analyzed using a repeated-measures two-factor ANOVA with avatar (scanned realistic/cartoon) and tracking (EFT/noEFT) as within-factors. Levene's tests were conducted to check the homogeneity of variances. Normality was checked with QQ-plots and Shapiro-Wilk tests. If residuals were not normally distributed, an aligned rank transform (ART) [25] was conducted. A priori significance level was set at $p < 0.05$. We report partial η^2 (η_p^2) as a measure of effect size. We calculated Tukeys HSD posthoc tests (or posthoc pairwise comparisons using ART-C for non-normally distributed data) for posthoc comparisons. We listed all descriptive statistics in Appendix A. Verbal feedback was collected as bullet points and grouped by participant, but was not analyzed further.

4.1 Uncanny Valley

The uncanny valley effect can be divided into four scales: *attractiveness*, *eerie*, *humanness*, and *spine-tingling*. Figure 3 shows an overview of the results.

4.1.1 Attractiveness. Our results point to a significant main effect of the avatar representation ($F(1,44) = 4.414, p = 0.041, \eta_p^2 = 0.09$). The cartoon avatars were considered to be more attractive in both tracking conditions *Cartoon.noEFT* ($M = 1.06, SD = 0.99$) *Cartoon.EFT* ($M = 1.06, SD = 1.06$) compared to the realistic avatars. We did not find two-way interaction effects ($F(1,44) = 0.377, p = 0.542, \eta_p^2 = 0.008$) or significant main effect for the eye and face tracking condition ($F(1,44) = 0.377, p = 0.542, \eta_p^2 = 0.008$).

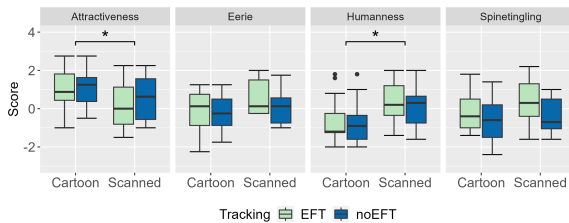


Figure 3: Uncanny valley questionnaire results [-3;+3]. Significant differences were found for the avatar (cartoon/scanned realistic) condition regarding the scales humanness $F(1,44) = 6.208, p = 0.017, \eta_p^2 = 0.12$ and attractiveness $F(1,44) = 4.414, p = 0.041, \eta_p^2 = 0.09$

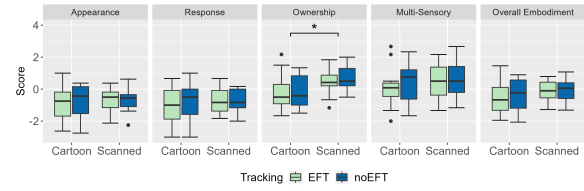


Figure 4: Embodiment questionnaire results [-3;+3]. Significant differences were found for the avatar (cartoon/scanned realistic) condition for the embodiment scale ownership $F(1,44) = 6.613, p = 0.014, \eta_p^2 = 0.13$

4.1.2 Eeriness. No significant differences were found on the eerie scale for either the avatar representations ($F(1,44) = 3.574, p = 0.065, \eta_p^2 = 0.075$) or enabling eye and face tracking ($F(1,44) = 1.496, p = 0.227, \eta_p^2 = 0.033$). There was no significant interaction effect ($F(1,44) = 0.496, p = 0.485, \eta_p^2 = 0.011$).

4.1.3 Humanness. Our results show a significant difference for the avatar representation ($F(1,44) = 6.208, p = 0.017, \eta_p^2 = 0.12$), suggesting that the scanned avatars *Scanned.EFT* ($M = 0.3, SD = 1.13$) and *Scanned.noEFT* ($M = 0.08, SD = 1.12$) are more humanlike than the cartoon one. We did not find two-way interaction effects ($F(1,44) = 0.005, p = 0.941, \eta_p^2 < 0.001$) or a significant main effect of tracking ($F(1,44) = 0.322, p = 0.573, \eta_p^2 = 0.007$) for this scale.

4.1.4 Spine-tingling. We did not detect significant differences for the spine-tingling scale for either the avatar representations ($F(1,44) = 1.375, p = 0.247, \eta_p^2 = 0.03$) or eye and face tracking ($F(1,44) = 3.048, p = 0.088, \eta_p^2 = 0.065$). There was no significant interaction effect ($F(1,44) = 0.153, p = 0.698, \eta_p^2 = 0.003$).

4.2 Embodiment

The embodiment questionnaire has four interrelated scales (*appearance*, *response*, *ownership*, *multi-sensory*) that form an *overall embodiment* score. The results can be seen in Figure 4.

4.2.1 Ownership. Our results show a significant difference in the ownership scale for the avatar representations ($F(1,44) = 6.613, p = 0.014, \eta_p^2 = 0.13$) but not for eye and face tracking ($F(1,44) = 0.206, p = 0.652, \eta_p^2 = 0.004$), indicating that the participants felt higher ownership over the scanned realistic avatar compared to the cartoon avatar. Here, the highest feeling of ownership was reported in the scanned avatar condition ($M = 0.55, SD = 0.84$) condition, followed by the *cartoon* condition ($M = -0.16, SD = 1.05$). There was no significant main effect for eye and face tracking or any other interaction effect ($F(1,44) = 0.092, p = 0.764, \eta_p^2 = 0.002$).

4.2.2 Appearance, Response, Multi-Sensory. The appearance score did not result in a significant difference between the avatar representations ($F(1,44) = 0.226, p = 0.637, \eta_p^2 = 0.005$) or eye and face tracking conditions ($F(1,44) = 0, p = 1, \eta_p^2 < 0.001$). The response score was also not significant for either the avatar representations ($F(1,44) = 0.575, p = 0.452, \eta_p^2 = 0.012$) or eye and face tracking conditions ($F(1,44) = 0.043, p = 0.837, \eta_p^2 < 0.001$). Similarly, the multi-sensory scale did not show significant differences for either

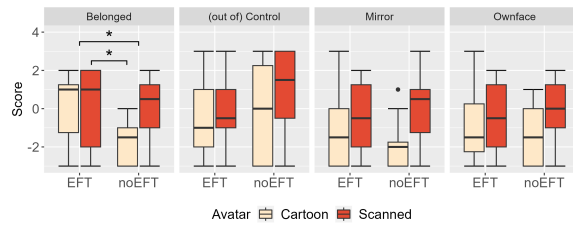


Figure 5: Enfacement questionnaire results [-3;+3]. Eye and face tracking had a significant main effect on the belonged scale ($F(1,44) = 5.063, p = 0.029, \eta_p^2 = 0.08$). Significant differences were also found for the avatar (cartoon/scanned) condition for the enfacement scales ownface ($F(1,44) = 4.52, p = 0.039, \eta_p^2 = 0.09$), belonged ($F(1,44) = 5.751, p = 0.021, \eta_p^2 = 0.1$) and mirror ($F(1,44) = 7.539, p = 0.009, \eta_p^2 = 0.15$).

the avatar representations ($F(1,44) = 0.624, p = 0.434, \eta_p^2 = 0.014$) or eye and face tracking conditions ($F(1,44) = 0.134, p = 0.716, \eta_p^2 = 0.003$). Results did not indicate any interaction effects ($F(1,44) < 0.119, p > 0.732$).

4.2.3 Overall Embodiment Score. We did not find significant differences for the avatar representations ($F(1,44) = 1.61, p = 0.211, \eta_p^2 = 0.04$) or eye and face tracking conditions ($F(1,44) = 0.089, p = 0.767, \eta_p^2 < 0.001$). There was no interaction effect ($F(1,44) = 0.025, p = 0.874, \eta_p^2 < 0.001$).

4.3 Enfacement

The Enfacement questionnaire consists of four scales: *belonged*, *(out of) control*, *mirror*, and *ownface* (see Figure 5).

4.3.1 Belonged. Data did not follow a normal distribution. Thus, we performed an ART procedure. Results indicate an interaction ($F(1,44) = 3.381, p = 0.047, \eta_p^2 = 0.058$) between avatar representation and eye and face tracking. Posthoc pairwise comparisons using ART-C showed that the ratings for the *Cartoon.noEFT* ($M = -1.67, SD = 1.15$) are significantly lower compared to the *Scanned.EFT* ($M = 0.33, SD = 1.97$). Results also point towards two main effects on the *belonged*-scale. Eye and face tracking had a significant effect ($F(1,44) = 5.547, p = 0.025, \eta_p^2 = 0.083$), were avatars with enabled tracking achieved higher scores ($M = 0.21, SD = 1.88$) compared to those without ($M = -0.75, SD = 1.4$). Results also point to a significant main effect of the avatar representations ($F(1,44) = 6.507, p = 0.096, \eta_p^2 = 0.1$), where the realistic avatar representations were rated higher ($M = 0.25, SD = 1.81$) than the cartoon avatars ($M = -0.79, SD = 1.88$). It is important to consider that the interaction effect might explain both main effects.

4.3.2 (Out of) control. Regarding the feeling of control of the avatar's face, we did not find any main for either the avatar ($F(1,44) = 1.148, p = 0.29, \eta_p^2 = 0.025$) or tracking ($F(1,44) = 1.452, p = 0.235, \eta_p^2 = 0.032$) condition. There was also no interaction effect ($F(1,44) = 0.287, p = 0.595, \eta_p^2 = 0.006$).

4.3.3 Mirror. Our results revealed a main effect of the avatar representation ($F(1,44) = 7.539, p = 0.009, \eta_p^2 = 0.15$), suggesting that

the participants' face was felt more as reflected in a mirror with their realistic avatar ($M = -0.17, SD = 1.83$) compared to the cartoon one ($M = -1.55, SD = 1.61$). We did not find a two-way interaction effect ($F(1,44) = 1.558, p = 0.219, \eta_p^2 = 0.034$) or a main effect of face tracking ($F(1,44) = 0.062, p = 0.804, \eta_p^2 = 0.001$).

4.3.4 Own face. The inspection of skewness and kurtosis within the QQ plots revealed a tailed distribution. Hence, the normality assumption was refuted, and an ART procedure was conducted. We did not find a two-way interaction effect ($F(1,44) = 0.368, p = 0.548, \eta_p^2 = 0.008$) or a significant main effect for eye and face tracking ($F(1,44) = 0.02, p < 0.889, \eta_p^2 < 0.001$). However, results show a significant main effect of avatar representation on the perception of their own face ($F(1,44) = 6.008, p = 0.02, \eta_p^2 = 0.09$). Here, participants reported that their virtual face felt like their own the most with the scanned avatar ($M = -0.13, SD = 1.74$) compared to the cartoon avatar ($M = -1.17, SD = 1.74$).

5 DISCUSSION

In this section we discuss our result within the context of our hypothesis. We also contextualize our results by discussing open challenges when it comes to eye and face tracking in VR.

5.1 H1: Eye and face tracking and embodiment

We hypothesized that eye and face tracking would lead to higher scores of embodiment for both avatar representations. The statistical analysis of our results does not support this hypothesis. To better understand these results, we took a closer look at the participants' feedback and found out about technical limitations. Some participants reported that they had fun using eye and face tracking, but they also criticized it. For example, P6 mentioned that *"the tracking is not perfectly accurate"*. In future work, eye and face tracking sensitivity toward facial movements should be carefully considered. If no true 1-to-1 mapping is possible, a potential solution is to exaggerate smaller movements so that the emotions in VR are more accurately displayed by modifying tracking data or increasing the weights of blendshapes.

5.2 H2: Eye and face tracking and enfacement

We hypothesized that eye and face tracking would lead to higher scores of enfacement for both avatar representations. Our results only partially support this hypothesis.

Of the four scales measuring enfacement, only the feeling of the face belonging to the participant was significantly higher with eye and face tracking for both avatar representations compared to the absence of it. However, this main effect may be explained by looking at the interaction effect: our results show that a scanned avatar with eye and face tracking evokes higher enfacement than a cartoon avatar without eye and face tracking. Participants had a significantly stronger feeling that the face of the former type belonged to them compared to the face of the latter type. This indicates that a scanned avatar without eye and face tracking can provide a better enfacement illusion than a cartoon avatar without eye and face tracking. In addition to that, the main effect between no tracking and tracking also points towards the superiority of eye and face tracking over conditions that do not feature this functionality.

It is important to remember that, similar to embodiment, the overall perception of the system was neutral or neutral to negative. Again, participants' comments hint towards a major limitation of the applied eye and especially face tracking technology: participants highlighted difficulties in expressing emotions, P5, "*it is sometimes difficult to express certain emotions*", P4, "*it is difficult to smile correctly*", or P3, "*you have to exaggerate your emotions a lot to see the emotion in the virtual avatar*". Given that the task was to reproduce emotions, the face is a key element in displaying emotions, and the process of displaying and detecting them is nuanced [6, 11], we believe that current technology, which is only based on sensors and predefined standard blendshapes, is not yet mature enough for this fine-grained task. Rather, it needs refinement to improve the feeling of having control over the virtual face, to create and support the mirror illusion, and the feeling of looking at the own face.

5.3 H3: Rendering style and embodiment

We hypothesized that the scanned realistic avatar would lead to higher embodiment scores than the cartoon avatar, regardless of eye and face tracking. Our results only partially support this hypothesis. When analyzing each embodiment category individually, the scanned avatar led to a higher ownership score than the cartoon avatar.

In addition to that, we found significant differences for the *belonged*, *mirror*, and *own face* scale of the enfacement questionnaire. Namely, compared to the cartoon avatar, participants felt that the scanned realistic avatar's face belonged more to them, had a stronger feeling that it was their face, and that it was like looking at their own face in the mirror. This aligns with previous results about the personalization of avatars (the more personalized, the more embodiment; [16, 51]). For future work, Fiedler et al. [16] provided specific scales for self-similarity and self-attribution in terms of self-identification, which could help elucidate embodiment nuances, especially when using personalized avatars.

5.4 Rendering style, eye and face tracking, and the uncanny valley

Regarding the uncanny valley, eye and face tracking did not lead to significant differences in our study. However, for rendering style, our results pinpoint significant differences favoring the scanned realistic avatar regarding humanness, i.e., it was perceived as more humanlike than the personalized cartoon avatar. We attribute this difference to the characteristics of the scanned realistic avatar, which appears with more human-like textures and less exaggerated and smoothed features. Regarding attractiveness, the cartoon avatar was rated significantly more attractive than the scanned avatar, but we cannot say if this was caused by the cartoon look or the visible hairstyle instead of a beanie. These results regarding attractiveness align with those of Ma and Pan [32], McDonnell et al. [33], who also report that their cartoon avatars and agents were often rated as more appealing and pleasant compared to realistic agents. It is possible that the style of many cartoon avatars (size of eyes, nose, mouth, facial width-to-height ratio, slim body shape, which are measures of attractiveness [38, 42, 46]), will often lead to the cartoon version outperforming scanned realistic self-avatars with respect to attractiveness. Regarding eeriness or spine-tingling, our results did

not point to significant differences. This means that participants did not perceive our avatars as particularly eerie and spine-tingling, with average values being close to or below the neutral value of 0. Thus, we believe that our avatars were not perceived as particularly uncanny.

5.5 Open Challenges: Eye and face tracking for VR

At the beginning of our research, we hypothesized that combining eye tracking, face tracking, and scanned realistic avatars would significantly outperform cartoon avatars without any eye and facial animations with regard to a variety of different perceptual constructs. Interestingly, our results were less conclusive than we expected. This poses the question of what research and industry need to pay attention to when integrating these technologies. Several aspects are essential for future researchers to consider and work on: First, the level of control provided by the commercial camera-based face and eye tracking systems employed by us does not seem to be accurate enough to deliver convincing animation fidelity. Accordingly, many participants mentioned that the provided level of control was not realistic, accurate, or responsive enough to provide a good user experience. This is especially important as we asked them to act out certain emotions that require a high level of control over facial muscles to convey the individual nuances. To overcome this issue, research should investigate approaches that operate on a physiological level [45]. Another approach is to investigate various exaggeration strategies to compensate for the lack of the device's responsiveness (e.g., Oh et al. [36]). In addition to that, while the HTC Facial Tracker is based on purely visual information, integrating other sensing approaches and technologies such as electromyography, electrooculography [27], and neural networks for signal fusion and generation [24] could further enhance the overall fidelity of eye and face tracking in avatar-mediated communication. Future head-mounted displays could and should integrate these technologies to provide a high level of control. Here, we also highlight that we only utilized eye direction vectors and eye openness values and did not employ other measures, such as pupil dilation, to enhance facial expressions.

Finally, research should strive to (objectively) quantify the quality of eye and face tracking so that a comparison across studies is possible (this includes, but is not limited to, sharing research prototypes as open-source). Arguably, the eye and face tracking employed by us (based on our pipeline, fixed blendshapes, and the soft- and hardware provided by HTC) did not perform sufficiently well. However, it is currently not easily possible to objectively compare the performance of various components of or even integral facial animation systems regarding avatar-mediated communication. Having the means to do so would potentially help to describe the contributing factors for a positive (or negative) experience with eye and face tracking.

6 LIMITATIONS

One of the limitations of our study is the number of participants ($N = 12$), which allows us to detect only minor effects. The small sample size is mainly due to the high effort necessary for avatar creation. While others have already proposed fast avatar-creation

pipelines [1] that use multi-camera rigs, we opted for a 3D scanning solution to allow for the portability of the pipeline.

Another limiting factor is the appearance of the avatars. The scanner used to create the personalized realistic avatar had problems scanning hair. Therefore, each participant wore a beanie during the scan, which was reflected in the resulting scanned realistic avatar, while their cartoon avatar wore no beanie as no appealing version was available during the creation process of the cartoon avatar. Future work should consider minimizing avatar differences as much as possible. For the cartoon vs. realistic avatar, some participants may have had *Ready Player Me* avatars closely resembling their actual appearance, while others did not. Additionally, the cartoon generator used only allows customization of the participants' faces and hairstyles. That is why all cartoon avatars have the same body shape and clothes. A follow-up study should quantify any potential effects of matching/mismatching clothes, for example, on embodiment and ownership. Next to this, we only used 43 of the 52 blendshapes supported by the *SRanipal* face tracking SDK, because the *Faceit* plugin used to create these blendshapes was not able to properly create these. The left-out blendshapes can be seen in Appendix A and are mostly blendshapes controlling the tongue, which was irrelevant to our tasks. Related to that, the teeth and tongue of the scanned avatar were not customized but rather simply adopted from the standard *Genesis8* character from *DAZ Studio*. One participant noticed this and remarked that "*the inside of his mouth was brighter than his face*" (P11). Future work should, therefore, consider these details.

7 CONCLUSION

We present a study where we examined the effect of eye and face tracking on the embodiment and enfacement with two types of personalized self-avatars. In the study, participants performed predefined movements and facial and bodily gestures associated with predefined emotions in front of a mirror. Overall, having a realistic avatar compared to a cartoon avatar was often beneficial (as in prior work). However, adding eye and face tracking to the avatars did not affect embodiment or enfacement in a relevant way. Thus, the core finding of our study is that the commercial camera-based eye and face tracking technology employed by us, in combination with predefined blendshapes, is not accurate enough for complex emotions and facial movements. Based on that, we also outline relevant research directions, including additional sensor data, better data fusion, physically inspired animation models, and objective metrics for comparing facial animation technologies.

While adding eye and face tracking did not have the hypothesized effect, participants' feedback and our data suggest that exploring the interplay between control fidelity and rendering style is vital. That is because, until AR and VR technology matures to offer all users realistic avatars with precise eye and face tracking, users will have to navigate a mix of rendering styles and tracking versions. Thus, identifying the optimal combinations is crucial for promoting wider AR and VR adoption.

As our results were obtained in a scenario where participants were observing themselves, a follow-up study in a social scenario with another person is crucial to investigate further any relationship between eye and face tracking and rendering style. Next to

that, the required future work in the area of self-avatars in VR is manifold. We believe that open-source pipelines for avatar generation are relevant to allow for comparability, adaptation, and reproduction. Further, developing and integrating better eye and face tracking technologies that provide a better level of control is crucial. Finally, quantifying various eye and face tracking technologies would further help researchers to investigate individual differences and effects.

ACKNOWLEDGMENTS

This research is funded by the Carl-Zeiss-Foundation ("Breakthroughs 2020" program, <https://www.carl-zeiss-stiftung.de/programm/czs-durchbrueche>), CO-HUMANICS project.

REFERENCES

- [1] Jascha Achenbach, Thomas Waltemate, Marc Erich Latoschik, and Mario Botsch. 2017. Fast Generation of Realistic Virtual Humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (Gothenburg, Sweden) (VRST '17). Association for Computing Machinery, New York, NY, USA, Article 12, 10 pages. <https://doi.org/10.1145/3139131.3139154>
- [2] Artec3D. 2022. Artec Leo. url: <https://www.artec3d.com/de/portable-3d-scanners/artec-leo> (visited on 12.12.2022).
- [3] Artec3D. 2022. Artec Studio 16. url: <http://docs.artec-group.com/as/16/en/> (visited on 23.12.2022).
- [4] Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (May 2021), 2608–2617. <https://doi.org/10.1109/TVCG.2021.3067783>
- [5] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *Presence: Teleoperators and Virtual Environments* 15, 4 (Aug. 2006), 359–372. <https://doi.org/10.1162/pres.15.4.359>
- [6] John N. Bassili. 1979. Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face. *Journal of Personality and Social Psychology* 37, 11 (1979), 2049–2058. <https://doi.org/10.1037/0022-3514.37.11.2049>
- [7] David Borland, Tabitha Peck, and Mel Slater. 2013. An Evaluation of Self-Avatar Eye Movement for Virtual Embodiment. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (April 2013), 591–596. <https://doi.org/10.1109/TVCG.2013.24>
- [8] Lauren E. Buck, Soumyajit Chakraborty, and Bobby Bodenheimer. 2022. The Impact of Embodiment and Avatar Sizing on Personal Space in Immersive Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (May 2022), 2102–2113. <https://doi.org/10.1109/TVCG.2022.3150483>
- [9] Isabelle Bühlhoff, Betty J Mohler, and Ian M Thornton. 2019. Face recognition of full-bodied avatars by active observers in a virtual environment. *Vision research* 157 (2019), 242–251.
- [10] Tara Collingwoode-Williams, Marco Gillies, Cade McCall, and Xueni Pan. 2017. The Effect of Lip and Arm Synchronization on Embodiment: A Pilot Study. In *2017 IEEE Virtual Reality (VR)*. 253–254. <https://doi.org/10.1109/VR.2017.7892272>
- [11] Alan S. Cowen and Dacher Keltner. 2020. What the Face Displays: Mapping 28 Emotions Conveyed by Naturalistic Expression. *American Psychologist* 75, 3 (2020), 349–364. <https://doi.org/10.1037/amp0000488>
- [12] Alexander Diel, Sarah Weigelt, and Karl F. Macdorman. 2021. A Meta-analysis of the Uncanny Valley's Independent and Dependent Variables. *ACM Transactions on Human-Robot Interaction* 11, 1 (Oct. 2021), 1:1–1:33. <https://doi.org/10.1145/3470742>
- [13] Paul Ekman. 1970. Universal facial expressions of emotion. *California Mental Health Research Digest* 8 (1970), 151–158.
- [14] Alejandro Estudillo and Markus Bindemann. 2016. Can Gaze-Contingent Mirror-Feedback from Unfamiliar Faces alter Self-Recognition? *The Quarterly Journal of Experimental Psychology* 70 (03 2016), 1–39. <https://doi.org/10.1080/17470218.2016.1166253>
- [15] FBra. 2022. Faceit. url: <https://blendermarket.com/products/faceit> (visited on 13.12.2022).
- [16] Marie Luisa Fiedler, Erik Wolf, Nina Döllinger, Mario Botsch, Marc Erich Latoschik, and Carolin Wienrich. 2023. Embodiment and Personalization for Self-Identification with Virtual Humans. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 799–800. <https://doi.org/10.1109/VRW58643.2023.00242>

- [17] Rebecca Fribourg, Ferran Argelaguet, Anatole Lécuyer, and Ludovic Hoyet. 2020. Avatar and Sense of Embodiment: Studying the Relative Preference Between Appearance, Control and Point of View. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2062–2072. <https://doi.org/10.1109/TVCG.2020.2973077>
- [18] Epic Games. 2022. Meta Human. url: <https://www.unrealengine.com/en-US/metahuman> (visited on 03.12.2022).
- [19] BoYu Gao, Joonwoo Lee, Huawei Tu, Wonjun Seong, and HyungSeok Kim. 2020. The Effects of Avatar Visibility on Behavioral Response with or without Mirror-Visual Feedback in Virtual Environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 780–781. <https://doi.org/10.1109/VRW50115.2020.00241>
- [20] Mar Gonzalez-Franco, Anthony Steed, Steve Hoogendyk, and Eyal Ofek. 2020. Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2023–2029. <https://doi.org/10.1109/TVCG.2020.2973075>
- [21] Geoffrey Gorisse, Olivier Christmann, Samory Houzangbe, and Simon Richir. 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. *Frontiers in Robotics and AI* 6 (2019). <https://www.frontiersin.org/articles/10.3389/frobt.2019.00008>
- [22] Darragh Higgins, Rebecca Fribourg, and Rachel McDonnell. 2021. Remotely Perceived: Investigating the Influence of Valence on Self-Perception and Social Experience for Dyadic Video-Conferencing With Personalized Avatars. *Frontiers in Virtual Reality* 2 (2021). <https://www.frontiersin.org/articles/10.3389/frvir.2021.668499>
- [23] Chin-Chang Ho and Karl MacDorman. 2017. Measuring the Uncanny Valley Effect: Refinements to Indices for Perceived Humanness, Attractiveness, and Eeriness. *International Journal of Social Robotics* 9 (01 2017), 129–139. <https://doi.org/10.1007/s12369-016-0380-9>
- [24] Bitu Houshmand and Naimul Mefraz Khan. 2020. Facial Expression Recognition Under Partial Occlusion from Virtual Reality Headsets Based on Transfer Learning. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*. 70–75. <https://doi.org/10.1109/BigMM50055.2020.00020>
- [25] Matthew Kay, Lisa A. Elkin, James J. Higgins, and Jacob O. Wobbrock. 2021. *mjskay/ARTool 0.11.0*. <https://doi.org/10.5281/zenodo.4721941>
- [26] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (11 2012), 373–387. https://doi.org/10.1162/PRES_a_00124 arXiv:https://direct.mit.edu/pvar/article-pdf/21/4/373/1625283/pres_a_00124.pdf
- [27] Chungwan Kim, Ho-Seung Cha, Jungwan Kim, Hwykuen Kwak, Woojin Lee, and Chang-Hwan Im. 2023. Facial Motion Capture System Based on Facial Electromyogram and Electrooculogram for Immersive Social Virtual Reality Applications. *Sensors* 23, 7 (Jan. 2023), 3580. <https://doi.org/10.3390/s23073580>
- [28] Marc Erich Latoschik, Jean-Luc Lugin, and Daniel Roth. 2016. FakeMi: A Fake Mirror System for Avatar Embodiment Studies. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology (Munich, Germany) (VRST '16)*. Association for Computing Machinery, New York, NY, USA, 73–76. <https://doi.org/10.1145/2993369.2993399>
- [29] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The Effect of Avatar Realism in Immersive Social Virtual Realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (Gothenburg, Sweden) (VRST '17)*. Association for Computing Machinery, New York, NY, USA, Article 39, 10 pages. <https://doi.org/10.1145/3139131.3139156>
- [30] Jean-Luc Lugin, Maximilian Landeck, and Marc Erich Latoschik. 2015. Avatar Embodiment Realism and Virtual Fitness Training. In *2015 IEEE Virtual Reality (VR)*. 225–226. <https://doi.org/10.1109/VR.2015.7223377>
- [31] Jean-Luc Lugin, Johanna Latt, and Marc Erich Latoschik. 2015. Avatar anthropomorphism and illusion of body ownership in VR. In *2015 IEEE Virtual Reality (VR)*. 229–230. <https://doi.org/10.1109/VR.2015.7223379>
- [32] Fang Ma and Xueni Pan. 2022. Visual Fidelity Effects on Expressive Self-avatar in Virtual Reality: First Impressions Matter. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 57–65. <https://doi.org/10.1109/VR51125.2022.00023>
- [33] Rachel McDonnell, Martin Breidt, and Heinrich H. Bühlhoff. 2012. Render Me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. *ACM Trans. Graph.* 31, 4, Article 91 (jul 2012), 11 pages. <https://doi.org/10.1145/2185520.2185587>
- [34] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine* 19, 2 (2012), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- [35] Aske Mottelson, Andreea Muresan, Kasper Hornbæk, and Guido Makransky. 2023. A Systematic Review and Meta-analysis of the Effectiveness of Body Ownership Illusions in Virtual Reality. *ACM Transactions on Computer-Human Interaction* (April 2023). <https://doi.org/10.1145/3590767>
- [36] Soo Youn Oh, Jeremy Bailenson, Nicole Krämer, and Benjamin Li. 2016. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLOS ONE* 11, 9 (Sept. 2016), e0161794. <https://doi.org/10.1371/journal.pone.0161794>
- [37] Catherine Oh Kruzic, David Kruzic, Fernanda Herrera, and Jeremy Bailenson. 2020. Facial Expressions Contribute More than Body Movements to Conversational Outcomes in Avatar-Mediated Virtual Environments. *Scientific Reports* 10, 1 (Nov. 2020), 20626. <https://doi.org/10.1038/s41598-020-76672-4>
- [38] Pamela M Pallett, Stephen Link, and Kang Lee. 2010. New “golden” ratios for facial beauty. *Vision research* 50, 2 (2010), 149–154.
- [39] Tabitha C. Peck and Mar Gonzalez-Franco. 2021. Avatar Embodiment: A Standardized Questionnaire. *Frontiers in Virtual Reality* 1 (2021). <https://doi.org/10.3389/frvir.2020.575943>
- [40] Giuseppina Porciello, Ilaria Bufalari, Ilaria Minio-Paluello, Enrico Di Pace, and Salvatore Maria Aglioti. 2018. The ‘Enfacement’ Illusion: A Window on the Plasticity of the Self. *Cortex* 104 (July 2018), 261–275. <https://doi.org/10.1016/j.cortex.2018.01.007>
- [41] Daz Productions. 2022. DAZ Studio. url: <https://www.daz3d.com/> (visited on 23.12.2022).
- [42] Mateusz Przylypiak, Jerzy Przylypiak, Robert Terlikowski, Emilia Lubowicka, Lech Chrostek, and Andrzej Przylypiak. 2018. Impact of face proportions on face attractiveness. *Journal of Cosmetic Dermatology* 17, 6 (2018), 954–959.
- [43] Russian3DScanner. 2022. R3DS Wrap. url: <https://www.russian3dscanner.com/> (visited on 13.12.2022).
- [44] Anca Salagean, Eleanor Crellin, Martin Parsons, Darren Cosker, and Danaë Stanton Fraser. 2023. Meeting Your Virtual Twin: Effects of Photorealism and Personalization on Embodiment, Self-Identification and Perception of Self-Avatars in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 499, 16 pages. <https://doi.org/10.1145/3544548.3581182>
- [45] Gabriel Schwartz, Shih-En Wei, Te-Li Wang, Stephen Lombardi, Tomas Simon, Jason Saragih, and Yaser Sheikh. 2020. The Eyes Have It: An Integrated Eye and Face Model for Photorealistic Facial Animation. *ACM Transactions on Graphics* 39, 4 (Aug. 2020), 91:91:1–91:91:15. <https://doi.org/10.1145/3386569.3392493>
- [46] Valentin Schwind, Katrin Wolf, and Niels Henze. 2018. Avoiding the uncanny valley in virtual character design. *interactions* 25, 5 (2018), 45–49.
- [47] Richard Skarbez, Frederick P Brooks, Jr, and Mary C Whitton. 2017. A survey of presence and related concepts. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–39.
- [48] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V Sanchez-Vives. 2009. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience* (2009), 29.
- [49] Angela Tinwell, Mark Grimshaw-Aagaard, Debbie Nabi, and Andrew Williams. 2011. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. *Computers in Human Behavior* 27 (03 2011), 741–749. <https://doi.org/10.1016/j.chb.2010.10.018>
- [50] Ana Tjadura-Jimenez, Matthew R. Longo, and Rosie Coleman. 2012. The Person in the Mirror: Using the Enfacement Illusion to Investigate the Experiential Structure of Self-Identification. *Consciousness and Cognition* 21, 4 (Dec. 2012), 1725–1738. <https://doi.org/10.1016/j.concog.2012.10.004>
- [51] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1643–1652. <https://doi.org/10.1109/TVCG.2018.2794629>
- [52] F. Weidner, G. Boettcher, S. Arboleda, C. Diao, L. Sinani, C. Kunert, C. Gerhardt, W. Broll, and A. Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR and VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 05 (may 2023), 2596–2606. <https://doi.org/10.1109/TVCG.2023.3247072>
- [53] Wolf3D. 2022. Ready Player Me. url: <https://readyplayer.me/> (visited on 23.12.2022).
- [54] Yuanjie Wu, Yu Wang, Sungchul Jung, Simon Hoermann, and Robert W. Lindeman. 2021. Using a Fully Expressive Avatar to Collaborate in Virtual Reality: Evaluation of Task Performance, Presence, and Attraction. *Frontiers in Virtual Reality* 2 (2021). <https://doi.org/10.3389/frvir.2021.641296>

A DIGITAL APPENDIX

The Unity project, information on the blendshapes, video figures, descriptive statistics and further graphs are available at <https://zenodo.org/doi/10.5281/zenodo.10022323>.