Coverage of Facial Expressions and Its Effects on Avatar Embodiment, Self-Identification, and Uncanniness

Peter Kullmann, Theresa Schell, Timo Menzel, Mario Botsch, Marc Erich Latoschik

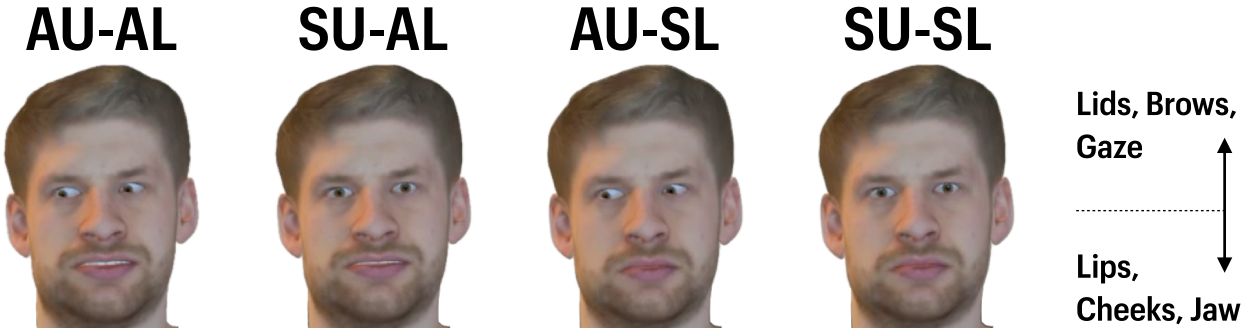

Fig. 1: **Experimental condition illustration.** Participants faced their virtual mirror image in one of four conditions: animated upper and animated lower (AU-AL), static upper and animated lower (SU-AL), animated upper and static lower (AU-SL), or static upper and static lower (SU-SL) face. We derived combinations from corresponding sensor coverage and role of the areas in nonverbal behavior.

Abstract—Facial expressions are crucial for many eXtended Reality (XR) use cases, from mirrored self exposures to social XR, where users interact via their avatars as digital alter egos. However, current XR devices differ in sensor coverage of the face region. Hence, a faithful reconstruction of facial expressions either has to exclude these areas or synthesize missing animation data with model-based approaches, potentially leading to perceivable mismatches between executed and perceived expression. This paper investigates potential effects of the coverage of facial animations (none, partial, or whole) on important factors of self-perception. We exposed 83 participants to their mirrored personalized avatar. They were shown their mirrored avatar face with upper and lower face animation, upper face animation only, lower face animation only, or no face animation. Whole animations were rated higher in virtual embodiment and slightly lower in uncanniness. Missing animations did not differ from partial ones in terms of virtual embodiment. Contrasts showed significantly lower humanness, lower eeriness, and lower attractiveness for the partial conditions. For questions related to self-identification, effects were mixed. We discuss participants' shift in body part attention across conditions. Qualitative results show participants perceived their virtual representation as fascinating yet uncanny.

Index Terms—Facial Animation, Self-Avatar, Uncanny Valley, Augmented Reality, Plausibility, Embodiment

1 Introduction

Nonverbal behavior is crucial for diverse social phenomena in interpersonal communication [4, 57] and self-perception [12]. Accordingly, it has been studied by scholars from all disciplines for a long time, dating back at least to Darwin's seminal work "The expression of the emotions in man and animals" two centuries ago.

Given the overall importance of facial expressions for interpersonal communication and self-perception, it has also motivated notable research in the area of XR (encompassing virtual, augmented and mixed reality) and virtual humans [56, 81]. In XR, avatars, the digital representations of the users in the virtual environments, can be of almost any conceivable form or shape [9]. However, human-like avatars seem particularly appropriate to satisfy the affordances of social XR [46, 50], self-perception [89], and body image intervention [18]. When given the option, people tend to represent themselves similar to their actual or idealized self [40]. Here, in analogy to the term virtual embodiment [26] describing the illusion of virtual body ownership [53], some authors specifically describe the illusion of owning a virtual face as *enfacement* [71, 85]. Previous work has reported that availability of facial expressions improved interaction in social XR [25, 68] and a lack of facial expressions was commented on as deficient by study participants in shared XR scenarios [60, 83]. As reported by Kokkinara and McDonnell [42], an increase in animation realism can make a virtual human more appealing. Two findings are particularly important for the motivation of the work reported here. First, there is evidence of media-induced differences in faithful reconstruction of realistic facial expression, e.g., caused by principle technological or implementation-specific deficits [84]. Second, an increase of realism not reaching a point of a non-perceivable quality can have diminishing or even negative returns, hypothesized as the *Uncanny Valley*. It describes an increase in eeriness, discomfort and revulsion in perceiving almost human-like characters that imperfectly resemble actual humans [59].

3D reconstruction of the shape and appearance of virtual humans has recently seen remarkable progress and now allows to achieve impressive visual fidelity [41, 72]. Some approaches operate automatically in less than 10 minutes [1], others require minimal hardware (just a smartphone) [92]. The resulting rigged and textured mesh models typically provide control over facial animation by mixing pre-computed mesh deformations, referred to as morph targets, blend shapes, or shape keys. Hence, they provide all the necessary means to be fully animated even in real-time given sufficient and accurate body and face tracking of the users controlling the avatars [78].

- *Peter Kullmann, Theresa Schell, and Marc Erich Latoschik are with Julius-Maximilians-Universität (JMU) Würzburg. E-mail: theresa.schell@stud-mail.uni-wuerzburg.de, {peter.kullmann | marc.latoschik}@uni-wuerzburg.de.*
- *Timo Menzel and Mario Botsch are with TU Dortmund University. E-mail: {timo.menzel | mario.botsch}@tu-dortmund.de.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

However, reliable face tracking poses unique challenges in XR when using Head-Mounted Displays (HMDs). The inherent occlusion of parts of the face while wearing HMDs limits sensor coverage and accuracy of tracking approaches. Some devices only track the pose of the headset and a pair of hand controllers, while others capture intricate finger motion and detailed facial expressions. Lack of sensors for full face coverage can be compensated for with add-on sensors, for example, for eye tracking[1] or lower face expressions[2]. Alternatively, missing face tracking data can be synthesized from other sources like eyes or voice [32, 66] or even using predefined models and rules [70]. Still, these alternatives compensating the lack of reliable full face tracking that covers all face areas with the same precision, accuracy, and low latency, often fall short in terms of their faithfulness. In addition, users might cover or disable sensors, e.g., accidentally or intentionally to protect their privacy [44] using obfuscating middleware [62]. Overall, we conclude that reliable and high-quality tracking of the entire face is generally not guaranteed with current HMDs, and that even technological advancements would still not overcome user preferences concerning sensor data disclosure. As a consequence, we need to understand how these boundary conditions impact self-perception and interpersonal communication in XR. This work starts to investigate the effect on self-perception. We pose the following research question:

RQ: *How do facial animations in the lower and/or upper face influence important factors of self-perception in XR, i.e., avatar embodiment, self-identification, and uncanniness?*

The upper and lower regions as delimiters were selected because typical sensor devices particularly target these areas and evidence for the variable salience these areas have.

2 Related Work

This section discusses previous work beginning with a general introduction to the significance of nonverbal behavior and facial expressions. This is followed by work investigating the perception of facial animation specifically in XR, and concluded by previous work on tracking facial expressions with integrated or attached sensors, or by synthesizing facial expressions either from other signals or using model- or rule-based approaches.

2.1 Nonverbal Behavior and Facial Expressions

Nonverbal behavior is typically separated into different categories depending on the role, significance, or performing body parts of the resulting expressions [19, 80]. While posture and gesture are expressed by our larger skeletal body configurations and movements, facial expressions are expressed by the fine-grained activation of our facial muscles and eye movements. Facial expressions are specifically prominent in human interaction [30]. Different expressions can rely on joint motion in several parts of the face or on a single facial region [6, 8, 65]. Eyes are often perceived as most salient in the face [34], movement in the larger mouth area aids in speech comprehension [63].

Seminal work by Ekman and Friesen [20] has popularized the idea of universal recognizability of some facial expressions, implying a common biological underpinning. They proposed the Facial Action Coding System (FACS) — a taxonomy of facial expressions deconstructed into action units (AU) and action descriptors (AD). It is based on the facial muscles that make up emotional display. Recent work has argued that emotional display and recognition can differ based on culture, familiarity, or context [7]. Overall, humans are suggested to have a shared "language" of facial expressions, albeit with some "dialects" [21]. Occluding parts of the face impedes facial identity and expression processing, as with sunglasses over the eyes or face masks over mouth and nose [74, 75]. Facial inexpressiveness, such as from severe facial paralysis, can be perceived as less favorable, but less so when compensated by other nonverbal cues like arm gestures [10].

2.2 Perception of Virtual Humans' Face Animation

Several scientific articles have examined the influences of face animation with a focus on the lower face, upper face, or both. For the lower face, enhancing animation parameters has shown beneficial: exaggerating lip motion can facilitate speech perception [3], while exaggerating smiles can render interaction partners more positive [67]. Earlier, Makarainen et al. [54] found that overly exaggerated facial expressions increased perceived strangeness.

With a focus on upper face animation, Borland et al. [11] varied gaze behavior in a virtual reality mirror exposure: eyes were either static, animated from eye tracking data, or animated as constant self-gaze. Their study results suggest an increase in self-identification with the avatar when adding eye movement. For dyadic interactions, Roth et al. [76] showed that being gazed at by one's stylized interlocutor avatar while speaking led to better dyadic interactions than interacting with an avatar with randomized gaze behavior. Tinwell et al. [87] explained the uncanny perception of virtual characters lacking upper facial expressions as hinting to an attempt to hide or mask unpleasant features. They refer to diagnosed psychopaths' lack of a startle reflex which inanimate behavior in virtual humans could remind us of, resulting in uncanny responses thereof.

Considering both upper and lower face animation, Murcia-Lopez et al. [61] let participants observe a stylized virtual presenter. They could incrementally increase one factor out of eye gaze, eye blinking, mouth animation, and microexpressions. When instructed to configure the best form of presentation with the least amount of changes, they reported to aim for most "life-like", "human", or "real". Gonzalez-Franco et al. [29] let participants give a pep talk to their virtual selves with a static face, audio-driven lip movement, or lip-sync and keyframed idle animations including eye blinks. They found that increasing face animation levels increased ratings for embodiment, enfacement, and self-identification. Kullmann et al. [45] let participants rate an observed avatar's naturalness and plausibility. Animated faces were rated as more natural and plausible than static faces — interestingly more so for synthesized than for tracked animations. Kimmel et al. [39] report that displaying tracked lower face expressions increased co-presence in spoken and wordless conversation, whereas displaying tracked eye movements increased co-presence. Their behavioral measures show an increase in looking at the partner avatar face when displaying all tracked movements. In their virtual mirror exposure, Hartbich et al. [31] showed participant's mirrored faces either static or with eye and face tracking data. Adding face animation partially increased enfacement. Looser and Wheatley [52] observed that eyes contribute most to perceiving animacy in a face.

2.3 XR Face Tracking and Animation

Many facial tracking solutions combine cameras covering the lower face with cameras targeting the eyes, for example [69, 86]. Some derive movement of occluded mimetic muscles via strain gauges (e.g., Li et al. [47]), electromyography like [13], or deploy acoustic sensing, as did Li et al. [49]. For broader overviews we point to reviews on general facial performance capture [91] and on gaze-specific works [2, 79].

To synthesize face animation without specific tracking data for face regions, many approaches refer to speech: When conversing, movement in the mouth region is highly correlated with speech output. Humans tolerate audio-visual asynchronies of up to several hundred milliseconds, likely also due to regular exposure to delay between seen and heard speech in media [14]. State-of-the-art lip synchronization is generated quickly enough for us to accept their asynchrony with visual speech. Hence, co-speech (audio-driven) facial expressions can be synthesized plausibly, as shown by [15, 36, 51]. Similarly, a listener's facial motion can be synthesized from speaker motion and audio [64]. Less common, Hickson et al. [32] derived facial expressions from eye tracker data. For brevity, we acknowledge rule-based approaches [70] and approaches using artist-generated keyframes. Interested readers consult a recent review on co-speech gesture generation including facial animation [66].

[1] https://pupil-labs.com/products/vr-ar

[2] https://www.vive.com/us/accessory/facial-tracker/

2.4 Contribution

In sum, virtual humans tend to be perceived more positively when (face) animation is more dynamic. Also, higher visual fidelity seems to increase expectations of behavioral fidelity [55]. This might conflict with tracking coverage and/or tracking quality in common consumer headset face tracking because increasing noise levels in facial animation can decrease observed communication experience [93].

Proprioception could interfere with low behavioral fidelity when, for example, looking at a mirror in mind-body therapy, or at a self-view in telepresence scenarios. Seeing yourself in a mirror with animation from an external source can reduce body ownership [28] and might be more akin to recognizing oneself in a virtual doppelganger [5]. While such use cases might call for more veridical behavior display, facial expression data is not available in all XR embodiment systems. XR face sensors typically cover the lower face, the upper face, or both.

We expand on the presented studies by investigating nonconversational face animation for photorealistic, truthful avatars. It is not clear how face animation coverage influences the embodiment of and self-identification with such avatars. From prior work presented above, we derive the following hypotheses:

H1: Self-identification and embodiment are higher when the whole face is animated. More specifically, we expect highest levels for whole face animation (group AU-AL), and higher levels for partial face animation (groups SU-AL, AU-SL) than no face animation (group SU-SL).

H2: Uncanniness is higher when face animation is inconsistent across both regions (groups AU-SL, SU-AL) than when it is consistent across regions (AU-AL, SU-SL).

3 Method

We investigate how congruent and incongruent animation of the lower and upper face affects the perception of one's self-avatar. To test this, we designed a study in a 2x2 between-subjects factorial design, varying the display of both upper face region (static vs. animated) and lower face region (static vs. animated). In an XR mirror exposure, participants reacted nonverbally to hypothetical scenario prompts and freely explored their personalized avatar as virtual mirror image. Written approval for the study was obtained from the local ethics committee.

3.1 Measures

We inquire about virtual embodiment, self-location, self-identification, uncanniness, most/least liked avatar aspects, simulator sickness, and mandatory free-text comments.

As control variable, we assessed symptoms related to simulator sickness [37] right before and right after the XR exposure.

Ownership of and agency over the body in the virtual mirror was measured with the Virtual Embodiment Questionnaire (VEQ) by Roth et al. [77] because the component structure and individual items match our research goals. Also, due to its early validation state, it facilitates comparison across studies. Some prior work measured self-identification with a face-morphing test [29]. Since our digital reconstructions closely resemble the participant, this does not fit our approach. Instead, we follow Fiedler et al. [23] in also assessing *self-location*. It is a common factor of investigations into the sense of virtual embodiment [38].

We query self-identification, as proposed by Fiedler et al. as extended Virtual Embodiment Questionnaire (VEQ+) [23], with statements grouped into the factors *self-similarity*, and *self-attribution*.

We assess uncanniness with the Uncanny Valley Index by Ho and McDorman [33], covering the factors humanness, eeriness, and attractiveness.

To inform future improvements to our approach of reconstructing and animating virtual humans, we prompt participants to select their most liked and disliked aspects of the avatar. We allow to pick one or more of the following options: face, hair, skin or texture, eyes, mouth, movements, skin tone, hands, clothing, upper torso, neck and shoulders, nothing, or a custom answer provided by participants themselves. We then aggregate mentions of face-related factors (face, eye, mouth) and others as either most or least liked avatar aspect. The resulting relative occurrence of face-related factors shows how salient the face region was to participants. We extend our hypotheses as follows:

H3: Face-related factors are mentioned more often as most/ least liked avatar aspects when any face animation is available (groups AU-AL, SU-AL, AU-SL).

3.2 Procedure

Our study took about 90 minutes and proceeded in five blocks (depicted in Figure 2).

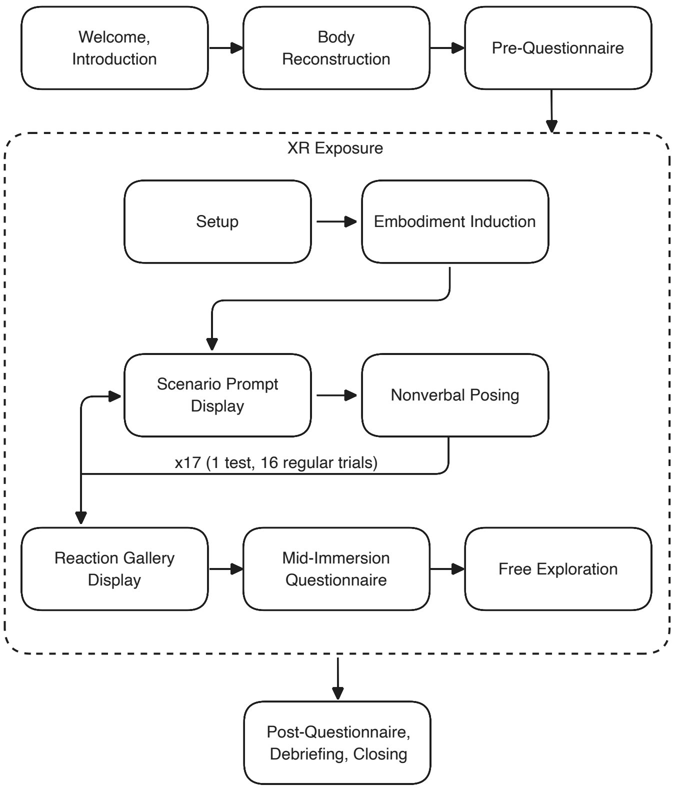

Fig. 2: *Experiment Procedure.*

First, participants read our briefing, data privacy policy, study participation consent form, and image recording consent form. After the experimenter answered questions, participants gave informed written consent to their participation and use of their data.

Second, a personalized mesh model of the participants was created (cf. subsubsection 3.3.2). Multi-view images of the participants' full bodies in a standing pose were captured, as were RGB-D images of their faces displaying different facial expressions. Gathered data was then processed into a skinned mesh model with personalized facial expression blend shapes.

Third, participants filled out our pre-questionnaire on a dedicated workstation. It inquired about previous XR and gaming experience, demographics, impairments in vision and hearing, and symptoms related to simulator sickness. Reported gender, previous XR experience, and regular game consumption were used to assign participants quasi-randomly to an experimental group by covariate-adaptive randomization [35]. This was done to evenly balance covariate levels across experimental groups.

Fourth, participants sat at another table and proceeded with the XR exposure using a Meta Quest Pro. Initially, the experimenter guided participants through device setup and explained symbolic input via direct touch or using a raycast pointer while viewing their egocentric view mirrored to a laptop. Participants performed the headset's fit adjustment procedure for optimal comfort and display clarity. It guides its wearer through balancing and centering the headset and adjusting distance between the lenses. Afterwards, participants went through the built-in 9-point grid calibration procedure for the eye trackers. In our experimental application, the experimenter registered the virtual with the physical environment. The previously blackened headset view faded to the XR view of the passthrough environment, allowing participants to acclimatize themselves. Subsequently, the end effector offsets for the body inverse kinematics solver were calibrated and screen-mirroring to the laptop was deactivated. The virtual mirror was positioned so participants saw the mirrored room and their mirrored avatar's head and most of their torso (cf. Figure 3).

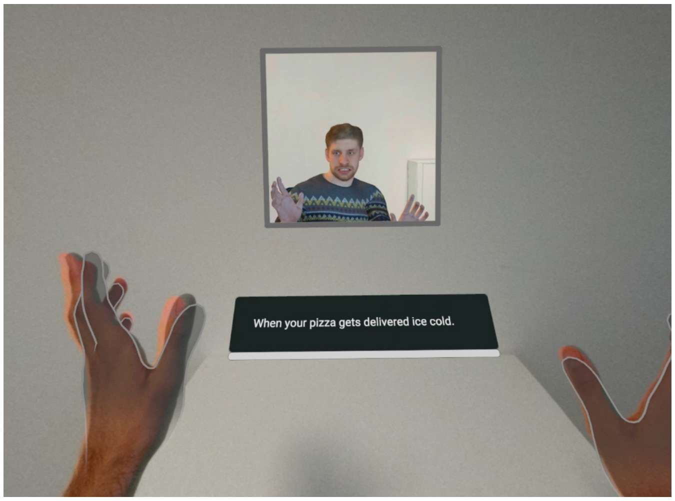

Fig. 3: *Egocentric view during main study task.* Participants were asked to depict prompted scenarios nonverbally.

In the remaining XR exposure, participants were instructed by text and audio instructions. As embodiment induction, they were guided through simple movements for ca. 135s. This included raising the stretched arm forward, hovering it in front of the torso, nodding, and rotating the head, always with pauses and instructions to look at one's mirrored or non-mirrored body. Next, participants were lead through a test trial to familiarize themselves with the trial procedure: after being shown a prompt about a hypothetical scenario, they were asked to spontaneously interpret the scenario by depicting it nonverbally. After a five-second countdown with beeps in the last three seconds, a snapshot was taken, indicated by a camera shutter sound effect. Participants had the option to clarify potential questions with the experimenter before performing the actual trials in two blocks with a break between them. Following the 16 trials, the mirror was hidden and participants were shown a gallery of their past reactions and the mid-immersion questionnaire. It inquired virtual embodiment and uncanniness. To conclude the XR exposure, participants had 60s to freely explore their mirror image. In total, participants spent around 15 minutes in XR. This depended on how quickly they answered the immersive questionnaire and pause duration between the two experimental task blocks. Display time of the virtual mirror was the same for every participant.

Finally, participants doffed the headset and filled out a questionnaire on a dedicated workstation. It contained questions on simulator sickness symptoms, most liked and most disliked aspects of their avatar, the presumed aim of the study, and open comments. Before closing, participants were shown our debriefing.

3.3 Apparatus

We digitally reconstruct participants following the pipeline proposed by Achenbach et al. [1] and blend shape personalization pipeline by Menzel et al. [58].

3.3.1 Software

We implemented our apparatus in Unity v2022.3.20f1 using the Universal Render Pipeline, Meta Movement SDK v4.0.1[3] and the paid plugin FinalIK v2.3 from Root Motion Inc.[4]. We executed the application standalone on a Meta Quest Pro headset (1832×1920 px per eye, 90 Hz refresh rate) running Meta Quest OS v62. Wei et al. [90] report its eye tracker has an average accuracy of 1.652° with a precision of .699°.

3.3.2 Personalized Avatar Mesh Model

For the body mesh, we photograph participants in our custom-built photogrammetry rig while in A-pose (standing upright with shoulders abducted, fingers spread, and neutral face). The 94 photos taken simultaneously are used to generate a dense point cloud. Then, pose and shape parameters of a rigged base model are optimized to best fit this point cloud. The resulting model has 60k triangles and a 4096×4096 px texture.

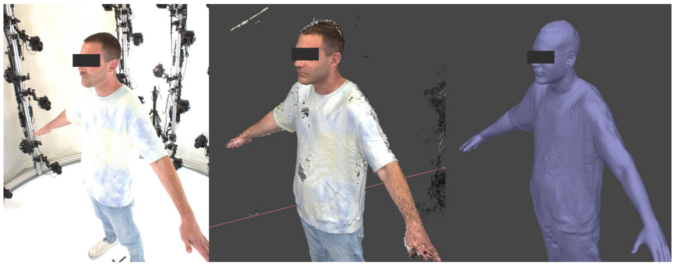

Fig. 4: *Body Reconstruction.* (Left) Images from 94 DLSR cameras are transformed into (center) dense point cloud, then (right) template mesh model is fit into pointcloud.

To personalize blend shapes, we captured face geometry and corresponding blend shape weights for 52 facial expressions detected by Apple ARKit[5]. For that, we ran our custom iOS app on an iPhone 12 Pro. Participant faces were lit evenly with a frontal ring light and captured from the phone mounted on a tripod (cf. Figure 5). To keep the lip seal and the inside of the upper eyelid visible, we positioned it slightly below the captured face at an upward facing angle. For each target expression, the capture app shows a text description, an animated illustration, and its currently detected intensity. Captures can be triggered in automatic or manual mode. Automatic mode is active while pressing and holding a button. It takes a snapshot of the cached peak expression when its coefficient is above 30% and it is either held consistently for several frames or released. Manual mode is used to capture a neutral face pose and target expressions that are not detected above our threshold (either due to tracking or posing difficulties). Blend shapes were then refined using a modification of example-based facial rigging [48] and merged with the body mesh.

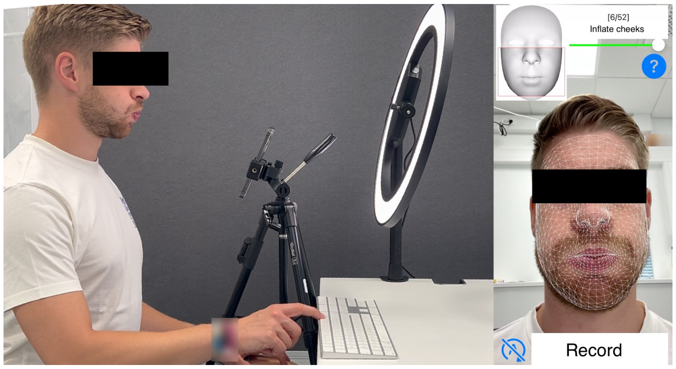

Fig. 5: *Facial Expression Capture.* (Left) Capture rig with tripod-mounted phone and ring light, (right) user interface.

[3] https://github.com/oculus-samples/Unity-Movement

[4] https://assetstore.unity.com/packages/tools/animation/final-ik-14290

[5] https://developer.apple.com/documentation/arkit/arfaceanchor/blendshapelocation

3.3.3 Character Animation

We process tracking data from the Meta Quest Pro HMD to steer the personalized avatar mesh model in real-time: We feed the headset pose and wrist poses to the VRIK body pose solver shipped with Root Motion's FinalIK. The head end effector follows the headset with a fixed offset so the foremost eye vertices are on the display plane. Similarly, wrist end effectors are offset from the wrist pose to align the virtual index finger tips with the tracked index finger tip when extended. We configured the solver to slightly stretch the arm when approaching maximum elbow extension. This avoids elbow snapping artifacts. Additionally, we provide a pelvis target to the solver whenever the headset is above the virtual table. This prevents the avatar from intersecting with the virtual table. The provided finger movements are transferred to our skeletal animation rig as forward kinematics pose.

Face animation weights are retrieved from the HMD's five infrared tracking sensors and mapped to semantically matching blend shapes in our mesh model. In the upper face, this comprises movement of eye lids, eye brows, and gaze direction. In the lower face, expression weights corresponded to movement of lips, jaw, cheeks, and tongue. We discarded the tongue animation coefficient since our blend shape rig did not include tongue motion. For conditions with static upper face animation, there was still occasional movement in the eye region. This was due to the natural co-activation of some blend shapes, e.g. wrinkling of the nose (tracked in lower face) also lowers the inner brows. Hence, the condition still reflects lack of tracking coverage in the upper face.

3.3.4 Digital Room Twin

We created a mesh model of the main experiment room as background for the XR mirror. We scanned and processed it with Niantic Inc.'s Scaniverse app[6] v2.1.9 running on an iPhone 13 Pro Max. To align the virtual with the physical environment, we sample a known landmark on the table with a Meta Quest Pro controller. We mirror the room model by inverting its x-axis and position it with respect to the landmark calibration. The stencil buffer mirror is visible through a framed rectangle positioned at participant eye height.

We used the room texture with its light baked during scanning without additional light. The character was lit with three-point lighting, as suggested for an appealing look [94].

3.3.5 Experimental Tasks

We aimed to bring about a wide spectrum of reactions without triggering unpleasant emotions. Hence, we chose to let participants nonverbally interpret prompts from relatable scenarios. We selected the parlor game What Do You Meme[7] as corpus for our prompts and picked some of them in a pre-study: Four persons unfamiliar with our study were asked to freely tag scenarios with its best fitting emotion. We further considered scenarios with clear inter-rater agreement and discarded explicit scenarios and ones with tags of negative connotations ("disgust" and "pain"). As a result, we selected 17 scenarios, such as "When your pizza gets delivered ice cold", "When you hear a recording of your own voice", or "When you cut wrapping paper and the scissors glide perfectly".

Additionally, we decided to let participants freely interact with their virtual mirror image for one minute.

3.3.6 Latency

To measure our embodiment system performance, we measured motion-to-photon latency for eye gaze and for jaw movement. Therefore, we used two smartphones to capture through-the-lense footage and the person wearing the HMD from the side, both in slow-motion. We then compared twenty-one movement onsets by counting the frames from when the person started a salient movement until their mirrored avatar did initiated the same movement respectively. This yielded an offset of ca. 56ms for lower face motion and about ca. 61ms for eye gaze

3.4 Participants

We recruited 100 participants with our institutional participation management system. Our inclusion criteria were language fluency, full legal age, no gaming addiction, and no pre-existing medical risks (seizure risk, binocular vision abnormalities, psychiatric disorders, heart conditions, or other serious medical conditions). For 34 of them, participation was compensated for with student credit. Others were paid according to the statutory minimum wage. Sessions were scheduled on weekdays during regular working hours. We excluded 17 datasets from participants that we did not reconstruct faithfully in an earlier version of the blend shape personalization.

The 83 included participants (66 female) had a mean age of 24.4 years (range: 19 to 49). Twenty completed the experiment in condition AU-AL, twenty-one in SU-AL, twenty-two in AU-SL, and twenty in SU-SL. They were mostly native speakers (81) or reported language fluency (2).

4 Results

We used R v4.2.2 [73] for analysis. We report descriptive statistics in Table 1 and provide charts in Figure 6. We report effects as significant at p<.05. To select an appropriate model at group size of around twenty, we tested dependent variables for normality (Shapiro-Wilk test) and homogeneity of variance (Levene's test with the median as center). Since group assignment was imbalanced, we computed ANOVAs with the *car* package [24] and type III sums of squares. Group comparisons were conducted with planned orthogonal contrasts, thus not requiring alpha correction.

4.1 Virtual Embodiment

For *ownership*, the Shapiro-Wilk test indicated that residuals were approximately normally distributed, W=.98, p=.1. Levene's test indicated that there was no significant difference in variances across groups, F(3,79)=1.58, p=.20. With both assumptions satisfied, we performed a two-factorial ANOVA. Ratings showed no significant effect of upper face animation, F(1,79)=.21, p=.65, η_p^2=.003. Similarly, the effect of lower face animation was not significant, F(1,79)=1.39, p=.24, η_p^2=.003. The interaction between upper and lower face animation did not have a significant effect, F(1,79)=.01, p=.92, η_p^2<.001. Planned contrasts revealed that having whole face animation (AU-AL) significantly increased ownership compared to the other groups (AU-SL, SU-AL, SU-SL), t(79)=12.20, p<.001. However, inconsistent face animation levels (AU-SL, SU-AL) did not significantly affect ownership compared to no face animation (SU-SL), t(79)=−.44, p=.86.

Agency ratings met assumptions for normal distribution, W=1, p=.1, and for variance homogeneity, F(3,79)=.77, p=.51. Hence, we used a two-factorial ANOVA. There was no significant effect of upper face animation, F(1,79)=.21, p=.65, η_p^2=.04, or of lower face animation, F(1,79)=1.39, p=.24, η_p^2=.04. Interaction between upper and lower face animation was not significant, F(1,79)=.01 , p=.92, η_p^2<.001. Planned contrasts showed a significant increase in agency in group AU-AL when compared to the other groups, t(79)=17.06,p<.001. The second contrast, comparing inconsistent face animation levels (AU-SL, SU-AL) to no face animation (SU-SL), indicated no significant difference, t(79)=−.41, p=.88.

Self-location data satisfied assumptions of homogeneity, F(3,79)=.41, p=.74, but not of normality assumptions, W=1, p=.04, so we used an Aligned Rank Transform (ART) ANOVA. It revealed no significant effect for upper face animation, F(1,79)=.60, p=.4, η_p^2=.004, no significant effect for lower face animation, F(1,79)=.59, p=.4, η_p^2=.004, and no significant interaction effect, F(1,79)=.75, p=.4, η_p^2=.003. Planned contrasts showed significantly higher self-location for whole face animation (AU-AL) when compared to the other groups, t(79)=9.23p=<.001. The second contrast, comparing inconsistent face animation levels (AU-SL, SU-AL) to no face animation (SU-SL), indicated no significant difference, t(79)=−.66, p=.72.

[6] https://scaniverse.com

[7] http://whatdoyoumeme.com/

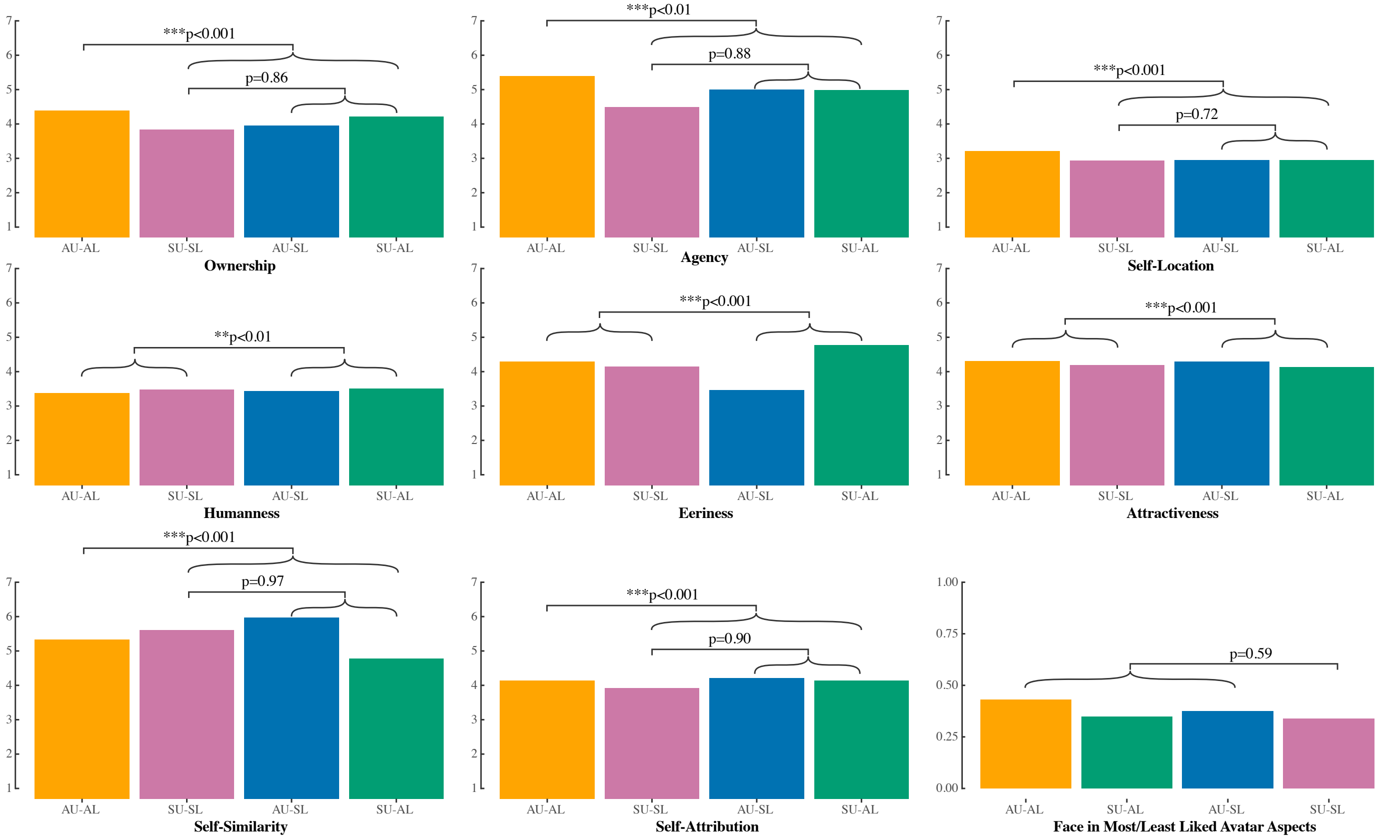

Fig. 6: *Bar charts with mean values and planned contrasts.*

4.2 Self-Identification

The factor *self-similarity* met the assumption of normality, W=1, p=.3, but did not meet the assumption of variance homogeneity, F(3,79)= .0056. Hence, we used an Aligned Rank Transform (ART) ANOVA. The effect of upper face animation was non-significant, F(1,79)=3.78, p=.06, η_p^2=.05. Ratings were significantly higher when the lower face was static rather than animated, F(1,79)=6.47, p=.01, η_p^2=.11. Interaction was not significant, F(1,79)=.34, p=.56, η_p^2=.002. The first planned contrast showed significantly higher self-similarity in group AU-AL when compared to the other groups, t(79)=15.79p=<.001. The planned contrast comparing inconsistent face animation levels (AU-SL, SU-AL) to no face animation (SU-SL), indicated no significant difference, t(79)=−.19, p=.97.

For *self-attribution*, data met assumptions of normal distribution, W=.99,p=.8, and of variance homogeneity, F(3,79)=.45, p=.72. We performed a two-factorial ANOVA. It showed no significant effect of upper face animation, F(1,79)=.00, p=.99, η_p^2=.004, no significant effect of lower face animation, F(1,79)=.03,p=.86, η_p^2<.001, and no significant interaction, F(1,79)=.26,p=.61, η_p^2<.001. Planned contrasts showed significantly higher self-attribution in group AU-AL when compared to the other groups, t(79)=10.79,p<.001. The second contrast, comparing inconsistent face animation levels (AU-SL, SU-AL) to no face animation (SU-SL), indicated no significant difference, t(79)=−.37, p=.9.

4.3 Uncanniness

For *humanness*, data satisfied assumptions of normality, W=.97, p=.05, and of variance homogeneity, F(3,79)=.9, p=.45, so we tested with a two-factorial ANOVA. Upper face animation had no significant effect, F(1,79)=.14, p=.70, η_p^2<.001, neither did lower face animation, F(1,79)=.03, p=.86, η_p^2<.001, nor was there a significant interaction between upper and lower face animation, F(1,79)=.03, p=.86, η_p^2<.001. The planned contrast showed significantly lower humanness for inconsistent face animation levels (AU-SL, SL-AU) than for consistent face animation levels (AU-AL, SU-SL), t(79)=−2.67, p<.01.

Data on *eeriness* met the assumption of normality, W=.98, p=.2, and the assumption of variance homogeneity, F(3,79)=1.58, p=.2. Therefore, we performed a two-factorial ANOVA. It showed no significant effect of upper face animation, F(1,79)=1.86, p=.176, η_p^2=.07, significantly less eeriness if lower face animation was static compared to animated lower faces, F(1,79)=5.72, p=.019, η_p^2=.10, and no significant interaction effect, F(1,79)=.16, p=.688, η_p^2=.002. The planned contrast showed significantly lower eeriness for inconsistent face animation levels (AU-SL, SL-AU) than for consistent face animation levels (AU-AL, SU-SL), t(79)=−3.93, p<.001.

Ratings of *attractiveness* showed normality, W=.99, p=.9, and variance homogeneity, F(3,79)=2.45, p=.069. A two-factorial ANOVA showed no significant effect of upper face animation, F(1,79)=.24, p=.63, η_p^2<.001, no significant effect of lower face animation, F(1,79)=.00, p=.96, η_p^2<.001, and no significant interaction effect, F(1,79)=.02, p=.88, η_p^2<.001. The planned contrast showed significantly lower attractiveness for inconsistent face animation levels (AU-SL, SL-AU) than for consistent face animation levels (AU-AL, SU-SL), t(79)=−4.56, p<.001.

4.4 Face in Most/Least Liked Avatar Aspects

For the *relative occurrence of face-related factors among avatar preferences*, the data showed that the assumption of normality was violated, W=.95, p=.002, and the assumption of variance homogeneity was met, F(3,79)=2.38,p=.076. Hence, we performed an Aligned Rank Transform (ART) ANOVA. There was no significant effect of upper face animation, F(1,79)=1.38, p=.24, η_p^2=.03, no significant effect of lower face animation, F(1,79)=.23,p=.64, η_p^2=.01, and no significant inter-

Measure	Items	Type	Animated Lower		Static Lower	
			Animated Upper	Static Upper	Animated Upper	Static Upper
Ownership [77]	4	Likert (7)	4.39 (0.95)	4.21 (1.16)	3.94 (1.25)	3.83(1.47)
Agency [77]	4		5.39 (0.95)	4.98 (0.94)	5.00 (1.08)	4.49 (1.33)
Self-Location [16,27]	4		3.21 (1.05)	2.94 (1.10)	2.94 (1.24)	2.93 (1.26)
Self-Similarity [23]	4	Likert (7)	5.33 (1.21)	4.77 (1.29)	5.97 (0.59)	5.60 (1.10)
Self-Attribution [23]	4		4.14 (1.27)	4.13 (1.14)	4.20 (1.33)	3.91 (1.33)
Humanness [33]	5	SD	3.37 (1.09)	3.50 (0.82)	3.43 (1.11)	3.47 (1.16)
Eeriness [33]	9		4.29 (1.28)	4.76 (1.03)	3.47 (1.32)	4.14 (0.69)
Attractiveness [33]	4		4.30 (0.66)	4.13 (1.20)	4.28 (1.29)	4.19 (1.16)
Face most/ least liked	13	multiple choice	0.43 (0.24)	0.35 (0.15)	0.37 (0.13)	0.34 (0.13)

Table 1: *Descriptive Statistics*. Mean with standard deviation in parentheses.

action effect, F(1,79)=.00,p=.96, η_p^2=.005. The planned comparison showed no significant difference between groups with any face animation (AU-AL, AU-SL, SU-AL) compared to having no face animation (SU-SL), t(79)=.54, p=.59.

4.5 Participant Comments

Overall, participants mentioned positive and negative aspects about their virtual mirror image, particularly concerning avatar similarity and identification, perceived facial expressions, avatar looks, and realism. Many participants were intrigued by the reconstruction process and XR mirror exposure, as phrased in the comparison to “an animal seeing its mirror image for the first time”. To several, this intrigue was mixed with uncanniness. This was reflected in“unpleasant but fascinating at the same time” or in stressing that their resemblance was “almost faithful”. In all conditions, participants expressed positive feedback about the interaction with the virtual mirror image and their fascination with the avatar’s resemblance to themselves. A few appreciated small details, such as the depiction of tattoos and clothing. Several participants criticized “unnatural” or “mechanical” body poses in the form of hand tracking loss, virtual arms intersecting with the virtual trunk, or “twirling” wrists (so-called candy-wrapper effect of linear blend skinning). Facial features were commented on with stark differences across groups. In groups with static face regions (SU-AL, AU-SL, SU-SL), the lack of facial movement in these regions was criticized. Animated face regions were perceived as both fascinating and uncanny, sometimes as “appearing unnatural”. Across conditions, several described the avatar eyes as “eerie” or “lifeless”. Some participants desired a more aesthetic appeal and realism in the avatar’s appearance (SU-AL), especially regarding features like the eyes and mouth. Various discrepancies in the avatar’s portrayal while smiling, as well as incorrect iris color and the depiction of teeth, were seen as distracting and were reported to partly hinder complete self-recognition, and as less beneficial to overall perceived avatar realism (AU-AL, SU-AL).

5 Discussion

We showed study participants their self-avatar as virtual mirror reflection with different levels of facial animation, varying animation in two regions (upper/ lower face) at two levels (static or animated).

Overall, ratings of embodiment and self-identification were highest for whole face animation (AU-AL) compared to other conditions (AU-SL, SU-AL, SU-SL). This is in line with previous work that found increases in measures related to self-perception when using avatars with more face animation [29,31,45] and our hypothesis H1. However, the data did not show H1’s hypothesized benefits of partial face animation (SU-AL, AU-SL) over no face animation (SU-SL) for embodiment and self-identification factors. This might have to do with the main effect of lower face animation (lower embodiment and lower self-identification with avatar for animated rather than static lower faces). Static lower faces were rated significantly more self-similar than animated ones. This might stem from the reconstruction accuracy of our approach, highlighting that observers are more sensitive to subtle details than designers might expect: While we deform the template model’s mouth cavity to fit the skin mesh without penetration, neither teeth nor tongue are personalized. Mismatches in shape and texture likely deteriorated the mirror exposure, but were only visible for animated lower faces (SU-AL, AU-AL).

We anticipated inconsistent face animation (groups AU-SL, SU-AL) to increase uncanniness (hypothesis H2). Participants’ reports overall showed significant differences between contrasted groups, but not consistently in the direction we expected. Inconsistent animation levels were perceived as less human and less attractive. This fits results by Tinwel et al. [87] that indicated partial animation (static upper face) to contribute to perceived uncanniness. Notably, inconsistent animation levels were perceived as *less* eerie. We suggest this to relate to the main effect of lower face animation (more eeriness for animated than static lower faces): seeing a non-personalized, straight, and symmetrical set of teeth might appear one’s mirror image more attractive and human, but the unfamiliarity of the inner mouth region with one’s own face might lead to perceived eeriness. This is contrary to findings by Looser et al. [52], where eyes reportedly contributed most to perceptions of animacy, though their stimuli were static images.

For preference ratings of avatar aspects, we expected that presence of any face animation (groups AU-SL, SU-AL, AU-SL) to result in more mentions of face-related aspects (hypothesis H3). While the condition with both face regions static (SU-SL) did have fewest mentions of face-related factors, we did not find this difference. Participants might not have paid the most attention to their mirrored avatar faces, but focused on their overall body expression.

6 Limitations and Future Work

Some limitations of our study should be mentioned. We relied on opportunity sampling, with about one third of participants being university students and ca. 80% female participants. We are not aware of systematic effects of education on avatar perception. Hence, we argue that observed effects are likely to generalize to participants with more diverse educational backgrounds. However, prior work has shown gender differences in avatar hand perception, suggesting female participants to“have increased expectations for their representation” [82]. Since evaluation of physical bodies has also shown gender-dependent [88], the gender composition of our sample warrants consideration when interpreting the results. Our predominantly female sample might have had higher body awareness, possibly noticing more discrepancies in appearance and animation, thus experiencing a stronger aversion. Future research should validate these findings with more demographically diverse populations.

Our apparatus preparation might have influenced participants. They spent around 15 minutes capturing over 50 distinct expressions and we calibrated the eye tracker built into the headset. We tried to avoid priming participants by framing the need for expression capture as "required for virtual human reconstruction" and the need for eye tracking calibration as "display optimization". As evident from participants' comments, some expected to see their mirrored avatar also manifest some form of face animation.

Many participants mentioned disliking their avatar's eyes. We did not personalize iris color in our reconstruction pipeline which might have contributed to that impression. Subsequent investigations should investigate whether the negative impression of the eyes can be reduced by correctly coloring irises or whether this stems from animation factors. Adding procedural microsaccadic jitter and/ or pupil unrest has previously been shown to increase naturalness of rendered eyes [43]. Such augmentations might also improve self-avatar eye perception.

Artifacts in our skeletal body animation were occasionally very noticeable. This could be addressed by handling tracking loss more smoothly as proposed by Ferstl et al. [22]. Integrating recent improvements in body pose estimation from sparse tracking data might reduce artifacts so that the arms distract less from perceiving the virtual face.

Future work should also investigate longitudinal effects of being embodied with different facial animation levels, similar to investigations by Dobre et al. [17]. We anticipate the animation to be perceived differently when users need to finish tasks instead of actively examining their virtual mirror reflection.

7 Conclusion

We explored the effect of nonconversational face animation for photorealistic avatars. Prior work suggests higher levels of virtual human animation to increase self-identification and sense of embodiment. Reportedly, inconsistent levels of animation in a virtual human stimulus have shown to make it harder to process depicted nonverbal cues. Face tracking capabilities greatly vary between different XR headsets in terms of their coverage and tracking accuracy of facial regions, mainly separating support for tracking the lower face and/or the upper face or by introducing different tracking or synthesis solutions and hence animation qualities for these regions. In addition, users might cover or disable sensors, e.g., accidentally or intentionally to protect their privacy. We wanted to know about the impact of these different boundary conditions on self-perception in XR. Hence, we systematically varied the scope of facial animation to include all four combinations of upper/lower face animations in a 2x2 between-groups experiment. As experimental task, participants interacted with their mirrored avatar by first reacting to hypothetical scenario prompts and later exploring their digital representation freely.

Quantitatively, we observed an increase in embodiment when face regions were animated rather than static. Displaying all available face animation data resulted in the highest sense of embodiment, showing none in the lowest. Ratings of uncanniness overall were less distinct, but higher animation levels were perceived as slightly less eerie. Effects of face animation on self-identification items were mixed.

Qualitatively, participants showed an ambivalence about the mirror exposure. On the one hand, many found the novel sight of their digital representation fascinating. On the other hand, many described a sensed eeriness in seeing oneself resembled almost faithfully, but not quite.

The negative influence of an animated lower face compared to a static lower face was prominent and surprising. It introduces a perceptual trade-off: while most constructs/factors benefited from face regions being animated rather than static, seeing avatars with an animated rather than a static lower face made participants perceive them as more eerie and less self-similar which we explained with the lower reconstruction accuracy of details like teeth and tongue of the lower face region. These results highlight the nuanced relationship between behavioral fidelity and self-perception of avatars.

We initially posed the following research question:

RQ: *How do facial animations in the lower and/or upper face influence important factors of self-perception in XR, i.e., avatar embodiment, self-identification, and uncanniness?*

Overall, our results show that the absence or presence of parts of facial animations in XR avatars significantly influence perceived embodiment, self-identification, and uncanniness. While avatar perception did not increase in all facets with more face parts included in the animation, showing faithful full face animations overall provided the most benefits for embodiment, self-identification, and uncanniness. However, even subtle inaccuracies and deficits in reconstruction accuracy can invalidate these assumptions. Notably, such deficits might not be apparent for the reconstructed static 3D models during inspection but might become specifically salient only during animation time. This calls for a thorough and systematic quality check of reconstruction results under various animated (and ideally also perceived) conditions. Since these subtle inaccuracies and their resulting effect of uncanniness were also noted by some but not all users, it shows that the salience of these deficits might also be influenced by a user-specific individual component. The latter would further complicate manual quality checks since it introduces yet another variable to be taken into account during the checks, potentially requiring multiple testers.

Our findings have extensive implications for self-avatar perception and social XR when using commodity face-tracking HMDs: When showing one's own digital representation, we suggest using whole face animations to reach overall high embodiment and self-identification. If self-similarity and eeriness are crucial, we advise to keep the lower face static, though this recommendation may diminish with increasing reconstruction quality of lower faces. Further, we anticipate similar effects for perceiving others' avatars. Collaboration likely benefits most from whole face animation, as reported in prior work discussed above. Still, lower face animation could have a similar negative impact, especially for avatars of well-acquainted people. For co-located scenarios, however, mixing an upper face avatar with the lower face as revealed by optical or video pass-through might render the need for a high-quality reconstruction of the lower face redundant.

We suggest future work to continue advancing reconstruction, animation, and quality testing of virtual humans to mitigate negative effects of subtle reconstruction inaccuracies like non-personalized oral cavities, and investigate long-term effects of different face animation.

Supplemental Materials

All supplemental materials are available on OSF at `http://doi.org/10.17605/OSF.IO/Q89GD`, released under a CC BY 4.0 license. In particular, they include files containing the anonymized data for and analyses for creating Tab. 1 and Fig. 6.

Acknowledgments

The authors thank the anonymous reviewers for their insightful feedback. This research has been funded by the German Federal Ministry Education and Research (BMBF) in the projects ViLeArn More (Grant 16DHB2214) and VIA-VR (Grant 16SV8446), the German Federal Ministry of Labor and Social Affairs (BMAS) in the project AIL AT WORK (Grant DKI.00.00030.21), and by the Bavarian State Ministry For Digital Affairs in the project XR Hub (Grant A5-3822-2-16).

References

[1] J. Achenbach, T. Waltemate, M. E. Latoschik, and M. Botsch. Fast generation of realistic virtual humans. In *VRST*. ACM Press, 2017. doi: 10.1145/3139131.3139154 1, 4

[2] I. B. Adhanom, P. MacNeilage, and E. Folmer. Eye Tracking in Virtual Reality: A Broad Review of Applications and Challenges. *Virtual Reality*, 27(2):1481–1505, 2023. doi: 10.1007/s10055-022-00738-z 2

[3] Z. Aldeneh, M. Fedzechkina, S. Seto, K. Metcalf, M. Sarabia, N. Apostoloff, and B.-J. Theobald. On the Role of LIP Articulation in Visual Speech Perception. In *ICASSP '23*. doi: 10.1109/ICASSP49357.2023.10096012 2

[4] M. Argyle. *Bodily communication*. Routledge, 2013. 1

[5] J. N. Bailenson and K. Y. Segovia. Virtual doppelgangers: Psychological effects of avatars who ignore their owners. *Online worlds: Convergence of the real and the virtual*, pp. 175–186, 2010. 3

[6] S. Baron-Cohen, S. Wheelwright, and T. Jolliffe. Is There a "Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. *Vis. Cogn.*, 1997. doi: 10.1080/713756761 2

[7] L. Barrett, R. Adolphs, S. Marsella, A. Martinez, and S. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *PSPI '19*. doi: 10.1177/1529100619832930 2

[8] O. Beaudry, A. Roy-Charland, M. Perron, I. Cormier, and R. Tapp. Featural processing in recognition of emotional facial expressions. *Cogn. Emot.*, 2013. doi: 10.1080/02699931.2013.833500 2

[9] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *PI*, 2002. 1

[10] K. Bogart, L. Tickle-Degnen, and N. Ambady. Communicating without the face: Holistic perception of emotions of people with facial paralysis. *BASP*, 2014. doi: 10.1080/01973533.2014.917973 2

[11] D. Borland, T. Peck, and M. Slater. An Evaluation of Self-Avatar Eye Movement for Virtual Embodiment. *IEEE TVCG*, 19(4):591–596, 2013. doi: 10.1109/TVCG.2013.24 2

[12] D. R. Carney, A. J. Cuddy, and A. J. Yap. Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological science*, 21(10):1363–1368, 2010. 1

[13] Y. Chen, Z. Yang, and J. Wang. Eyebrow emotional expression recognition using surface EMG signals. *Neurocomputing*, 168:871–879, 2015. doi: 10.1016/j.neucom.2015.05.037 2

[14] B. Conrey and D. B. Pisoni. Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J. Acoustical Soc. America*, 119(6):4065–4073, 2006. doi: 10.1121/1.2195091 2

[15] D. Cudeiro, T. Bolkart, C. Laidlaw, A. Ranjan, and M. J. Black. Capture, Learning, and Synthesis of 3D Speaking Styles. In *IEEE/CVF CVPR*, pp. 10093–10103, 2019. doi: 10.1109/CVPR.2019.01034 2

[16] H. G. Debarba, E. Molla, B. Herbelin, and R. Boulic. Characterizing embodied interaction in First and Third Person Perspective viewpoints. In *IEEE 3DUI*, 2015. doi: 10.1109/3DUI.2015.7131728 7

[17] G. C. Dobre, M. Wilczkowiak, M. Gillies, X. Pan, and S. Rintel. Nice is Different than Good: Longitudinal Communicative Effects of Realistic and Cartoon Avatars in Real Mixed Reality Work Meetings. In *CHI EA*, pp. 1–7. ACM, 2022. doi: 10.1145/3491101.3519628 8

[18] N. Döllinger, E. Wolf, D. Mal, S. Wenninger, M. Botsch, M. E. Latoschik, and C. Wienrich. Resize me! exploring the user experience of embodied realistic modulatable avatars for body image intervention in virtual reality. *FRIVR*, 2022. doi: 10.3389/frvir.2022.935449 1

[19] P. Ekman and W. V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, 1969. 2

[20] P. Ekman and W. V. Friesen. Facial action coding system, 1978. doi: 10.1037/t27734-000 2

[21] H. A. Elfenbein and N. Ambady. On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol. Bull.*, 128(2):203–235, 2002. doi: 10.1037/0033-2909.128.2.203 2

[22] Y. Ferstl, R. McDonnell, and M. Neff. Evaluating Study Design and Strategies for Mitigating the Impact of Hand Tracking Loss. In *ACM SAP*, pp. 1–12. ACM, 2021. doi: 10.1145/3474451.3476235 8

[23] M. Fiedler, E. Wolf, N. Döllinger, M. Botsch, M. E. Latoschik, and C. Wienrich. Embodiment and Personalization for Self-Identification with Virtual Humans. In *IEEE VRW '23*. doi: 10.1109/VRW58643.2023.00242 3, 7

[24] J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third ed., 2019. 5

[25] A. Fraser, I. Branson, R. Hollett, C. Speelman, and S. Rogers. Expressiveness of real-time motion captured avatars influences perceived animation realism and perceived quality of social interaction in virtual reality. *Front. Virtual Real.*, 3, 2022. doi: 10.3389/frvir.2022.981400 1

[26] A. Genay, A. Lecuyer, and M. Hachet. Being an avatar "for real": A survey on virtual embodiment in augmented reality. *IEEE TVCG*, 2022. doi: 10.1109/tvcg.2021.3099290 1

[27] M. Gonzalez-Franco and T. C. Peck. Avatar Embodiment. Towards a Standardized Questionnaire. *Front Robot AI*, 2018. doi: 10.3389/frobt.2018.00074 7

[28] M. Gonzalez-Franco, D. Perez-Marcos, B. Spanlang, and M. Slater. The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In *IEEE VR*, pp. 111–114, 2010. doi: 10.1109/VR.2010.5444805 3

[29] M. Gonzalez-Franco, A. Steed, S. Hoogendyk, and E. Ofek. Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification. *TVCG*, 2020. doi: 10.1109/TVCG.2020.2973075 2, 3, 7

[30] R. Groner, F. Walder, and M. Groner. Looking at faces: Local and global aspects of scanpaths. In *Adv. Psychol.*, vol. 22, pp. 523–533. Elsevier, 1984. doi: 10.1016/S0166-4115(08)61874-9 2

[31] J. Hartbrich, F. Weidner, C. Kunert, A. Raake, W. Broll, and S. Arévalo Arboleda. Eye and Face Tracking in VR: Avatar Embodiment and Enfacement with Realistic and Cartoon Avatars. In *MUM*, pp. 270–278. ACM, 2023. doi: 10.1145/3626705.3627793 2, 7

[32] S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. Essa. Eyemotion: Classifying Facial Expressions in VR Using Eye-Tracking Cameras. In *IEEE WACV*, 2019. doi: 10.1109/WACV.2019.00178 2

[33] C.-C. Ho and K. MacDorman. Measuring the Uncanny Valley Effect: Refinements to Indices for Perceived Humanness, Attractiveness, and Eeriness. *Int J Soc Rob*, 2017. doi: 10.1007/s12369-016-0380-9 3, 7

[34] S. Janik, A. Wellens, M. Goldberg, and L. Dell'Osso. Eyes as the Center of Focus in the Visual Examination of Human Faces. *Percept Mot Skills*, 47(3):857–858, 1978. doi: 10.2466/pms.1978.47.3.857 2

[35] M. Jin, A. Polis, and J. Hartzel. Algorithms for minimization randomization and the implementation with an R package. *Commun. Stat. - Simul. Comput.*, 2021. doi: 10.1080/03610918.2019.1619765 3

[36] T. Karras, T. Aila, S. Laine, A. Herva, and J. Lehtinen. Audio-driven facial animation by joint end-to-end learning of pose and emotion. *ACM TOG*, 36(4):1–12, 2017. doi: 10.1145/3072959.3073658 2

[37] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int. J. Aerosp. Psychol.*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303_3 3

[38] K. Kilteni, R. Groten, and M. Slater. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012. doi: 10.1162/PRES_a_00124 3

[39] S. Kimmel, F. Jung, A. Matviienko, W. Heuten, and S. Boll. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *ACM CHI*, 2023. doi: 10.1145/3544548.3580707 2

[40] D. Ko and J. Park. I am you, you are me: Game character congruence with the ideal self. *INTR*, 2020. doi: 10.1108/INTR-05-2020-0294 1

[41] M. Kocabas, J.-H. R. Chang, J. Gabriel, O. Tuzel, and A. Ranjan. Hugs: Human gaussian splats. In *IEEE/CVF CVPR*, 2024. 1

[42] E. Kokkinara and R. McDonnell. Animation realism affects perceived character appeal of a self-virtual face. In *ACM MIG*, pp. 221–226. ACM, 2015. doi: 10.1145/2822013.2822035 1

[43] K. Krejtz, A. Duchowski, H. Zhou, S. Jörg, and A. Niedzielska. Perceptual evaluation of synthetic gaze jitter. *Computer Animation and Virtual Worlds*, 29(6):e1745, Nov. 2018. doi: 10.1002/cav.1745 8

[44] J. Kröger, O. Lutz, and F. Müller. What does your gaze reveal about you? On the privacy implications of eye tracking. In *Privacy and Identity Management. Data for Better Living: AI and Privacy*. 2020. doi: 10.1007/978-3-030-42504-3_15 2

[45] P. Kullmann, T. Menzel, M. Botsch, and M. E. Latoschik. An Evaluation of Other-Avatar Facial Animation Methods for Social VR. In *CHI EA*, pp. 1–7. ACM, 2023. doi: 10.1145/3544549.3585617 2, 7

[46] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *VRST '17*, pp. 1–10. ACM Press, 2017. doi: 10.1145/3139131.3139156 1

[47] H. Li, L. Trutoiu, K. Olszewski, L. Wei, T. Trutna, P.-L. Hsieh, A. Nicholls, and C. Ma. Facial performance sensing head-mounted display. *ACM TOG*, 34(4):1–9, 2015. doi: 10.1145/2766939 2

[48] H. Li, T. Weise, and M. Pauly. Example-based facial rigging. *ACM TOG*, 29(4), article no. 32, 2010. doi: 10.1145/1778765.1778769 4

[49] K. Li, R. Zhang, S. Chen, B. Chen, M. Sakashita, F. Guimbretiere, and C. Zhang. EyeEcho: Continuous and Low-power Facial Expression Tracking on Glasses. In *ACM CHI*, 2024. doi: 10.1145/3613904.3642613 2

[50] J. Lin and M. E. Latoschik. Digital body, identity and privacy in social virtual reality: A systematic review. *Frontiers in Virtual Reality*, 3, 2022. doi: 10.3389/frvir.2022.974652 1

[51] H. Liu, Z. Zhu, G. Becherini, Y. Peng, M. Su, Y. Zhou, X. Zhe, N. Iwamoto, B. Zheng, and M. J. Black. EMAGE: Towards unified holistic co-speech gesture generation via expressive masked audio gesture modeling. In *IEEE/CVF CVPR*, pp. 1144–1154, 2024. 2

[52] C. E. Looser and T. Wheatley. The Tipping Point of Animacy: How, When, and Where We Perceive Life in a Face. *Psychol Sci*, 21(12):1854–1862, 2010. doi: 10.1177/0956797610388044 2, 7

[53] J.-L. Lugrin, J. Latt, and M. E. Latoschik. Anthropomorphism and illusion of virtual body ownership. In *International Conference on Artificial Reality and Telexistence/Eurographics Symposium on Virtual Environments (ICAT/EGVE)*, pp. 1–8, 2015. 1

[54] M. Mäkäräinen, J. Kätsyri, and T. Takala. Exaggerating Facial Expressions: A Way to Intensify Emotion or a Way to the Uncanny Valley? *Cogn Comput*, 6(4):708–721, 2014. doi: 10.1007/s12559-014-9273-0 2

[55] R. McDonnell, M. Breidt, and H. Bülthoff. Render me real?: Investigating

the effect of render style on the perception of animated virtual humans. *ACM TOG*, 31(4), 2012. doi: 10.1145/2185520.2185587 3

[56] R. McDonnell, M. Larkin, B. Hernández, I. Rudomin, and C. O'Sullivan. Eye-catching crowds: Saliency based selective variation. *ACM TOG*, 28(3):1–10, 2009. doi: 10.1145/1531326.1531361 1

[57] A. Mehrabian. *Nonverbal communication*. Routledge, 2017. 1

[58] T. Menzel, M. Botsch, and M. E. Latoschik. Automated blendshape personalization for faithful face animations using commodity smartphones. In *VRST '22*, article no. 22. ACM, 2022. doi: 10.1145/3562939.3565622 4

[59] M. Mori, K. MacDorman, and N. Kageki. The Uncanny Valley.From the Field. *IEEE Rob & Autom '12*. doi: 10.1109/MRA.2012.2192811 1

[60] F. Moustafa and A. Steed. A longitudinal study of small group interaction in social virtual reality. In *ACM VRST '18*. doi: 10.1145/3281505.3281527 1

[61] M. Murcia-Lopez, T. Collingwoode-Williams, W. Steptoe, R. Schwartz, T. J. Loving, and M. Slater. Evaluating Virtual Reality Experiences Through Participant Choices. In *IEEE VR*, pp. 747–755, 2020. doi: 10.1109/VR46266.2020.00098 2

[62] V. Nair, W. Guo, J. F. O'Brien, L. Rosenberg, and D. Song. Deep Motion Masking for Secure, Usable, and Scalable Real-Time Anonymization of Ecological Virtual Reality Motion Data. In *IEEE VRW*, pp. 493–500, 2024. doi: 10.1109/VRW62533.2024.00096 2

[63] J. Navarra, H. H. Yeung, J. F. Werker, and S. Soto-Faraco. Multisensory Interactions in Speech Perception. In *New Handbook of Multisensory Processing*. 2012. doi: 10.7551/mitpress/8466.003.0038 2

[64] E. Ng, H. Joo, L. Hu, H. Li, T. Darrell, A. Kanazawa, and S. Ginosar. Learning to listen: Modeling non-deterministic dyadic facial motion. In *IEEE/CVF CVPR*, pp. 20395–20405, 2022. 2

[65] M. Nusseck, D. Cunningham, C. Wallraven, and H. Bulthoff. The contribution of different facial regions to the recognition of conversational expressions. *J. Vis.*, 8(8):1–1, 2008. doi: 10.1167/8.8.1 2

[66] S. Nyatsanga, T. Kucherenko, C. Ahuja, G. Henter, and M. Neff. A Comprehensive Review of Data-Driven Co-Speech Gesture Generation. *Comp. Graph. Forum*, 42(2), 2023. doi: 10.1111/cgf.14776 2

[67] S. Oh, J. Bailenson, N. Krämer, and B. Li. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLoS ONE*, 2016. doi: 10.1371/journal.pone.0161794 2

[68] C. Oh Kruzic, D. Kruzic, F. Herrera, and J. Bailenson. Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. *Sci Rep*, 10(1):20626, 2020. doi: 10.1038/s41598-020-76672-4 1

[69] K. Olszewski, J. J. Lim, S. Saito, and H. Li. High-fidelity facial and speech animation for VR HMDs. *ACM TOG*, 2016. doi: 10.1145/2980179.2980252 2

[70] K. Perlin. Layered compositing of facial expression. In *ACM SIGGRAPH*, vol. 3, pp. 226–227, 1997. 2

[71] G. Porciello, I. Bufalari, I. Minio-Paluello, E. Di Pace, and S. M. Aglioti. The 'Enfacement' illusion: A window on the plasticity of the self. *Cortex*, 104:261–275, 2018. doi: 10.1016/j.cortex.2018.01.007 1

[72] S. Qian, T. Kirschstein, L. Schoneveld, D. Davoli, S. Giebenhain, and M. Nießner. GaussianAvatars: Photorealistic head avatars with rigged 3D gaussians. In *IEEE/CVF CVPR*, 2024. 1

[73] R Core Team. *R: A Language and Environment for Statistical Computing*, 2022. 5

[74] M. Rinck, M. Primbs, I. Verpaalen, and G. Bijlstra. Face masks impair facial emotion recognition and induce specific emotion confusions. *Cogn. Res.*, 7(1):83, 2022. doi: 10.1186/s41235-022-00430-5 2

[75] D. Roberson, M. Kikutani, P. Döge, L. Whitaker, and A. Majid. Shades of emotion: What the addition of sunglasses or masks to faces reveals about the development of facial expression processing. *Cognition*, 125(2):195–206, 2012. doi: 10.1016/j.cognition.2012.06.018 2

[76] D. Roth, P. Kullmann, G. Bente, D. Gall, and M. E. Latoschik. Effects of Hybrid and Synthetic Social Gaze in Avatar-Mediated Interactions. In *ISMAR-Adjunct*, 2018. doi: 10.1109/ISMAR-Adjunct.2018.00044 2

[77] D. Roth and M. E. Latoschik. Construction of the Virtual Embodiment Questionnaire (VEQ). *TVCG '20*. doi: 10.1109/TVCG.2020.3023603 3, 7

[78] D. Roth, J.-P. Stauffert, and M. E. Latoschik. *Avatar Embodiment, Behavior Replication, and Kinematics in Virtual Reality*, vol. 1, pp. 321–348. Springer US, 2019. 1

[79] K. Ruhland, S. Andrist, J. B. Badler, C. E. Peters, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. Look me in the Eyes: A Survey of Eye and Gaze Animation for Virtual Agents and Artificial Systems, 2014. doi: 10.2312/EGST.20141036 2

[80] A. E. Scheflen. The significance of posture in communication systems. *Psychiatry*, 27(4):316–331, 1964. 2

[81] V. Schwind and S. Jäger. The Uncanny Valley and the Importance of Eye Contact. *i-com*, 15(1), 2016. doi: 10.1515/icom-2016-0001 1

[82] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. "These are not my hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1577–1582. ACM, Denver Colorado USA, May 2017. doi: 10.1145/3025453.3025602 7

[83] P. Sykownik, S. Karaosmanoglu, K. Emmerich, F. Steinicke, and M. Masuch. VR Almost There: Simulating Co-located Multiplayer Experiences in Social Virtual Reality. In *ACM CHI*, 2023. doi: 10.1145/3544548.3581230 1

[84] A. H. Syrjämäki, P. Isokoski, V. Surakka, T. P. Pasanen, and J. K. Hietanen. Eye contact in virtual reality – A psychophysiological study. *Comput. Hum. Behav.*, 2020. doi: 10.1016/j.chb.2020.106454 1

[85] A. Tajadura-Jiménez, M. R. Longo, R. Coleman, and M. Tsakiris. The person in the mirror: Using the enfacement illusion to investigate the experiential structure of self-identification. *Consciousness and Cognition*, 21(4):1725–1738, 2012. doi: 10.1016/j.concog.2012.10.004 1

[86] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner. FaceVR: Real-Time Gaze-Aware Facial Reenactment in Virtual Reality. *ACM TOG*, 37(2):1–15, 2018. doi: 10.1145/3182644 2

[87] A. Tinwell, D. A. Nabi, and J. P. Charlton. Perception of psychopathy and the Uncanny Valley in virtual characters. *Comput. Hum. Behav.*, 29(4):1617–1625, 2013. doi: 10.1016/j.chb.2013.01.008 2, 7

[88] M. M. Voges, C.-M. Giabbiconi, B. Schöne, M. Waldorf, A. S. Hartmann, and S. Vocks. Gender Differences in Body Evaluation: Do Men Show More Self-Serving Double Standards Than Women? *Frontiers in Psychology*, 10, Mar. 2019. doi: 10.3389/fpsyg.2019.00544 7

[89] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE TVCG*, 24(4):1643–1652, 2018. doi: 10.1109/TVCG.2018.2794629 1

[90] S. Wei, D. Bloemers, and A. Rovira. A Preliminary Study Eye Tracker in the Meta Quest Pro. In *ACM IMX*, pp. 216–221. ACM, 2023. doi: 10.1145/3573381.3596467 4

[91] L. Wen, J. Zhou, W. Huang, and F. Chen. A Survey of Facial Capture for Virtual Reality. *Access'22*. doi: 10.1109/ACCESS.2021.3138200 2

[92] S. Wenninger, J. Achenbach, A. Bartl, M. E. Latoschik, and M. Botsch. Realistic virtual humans from smartphone videos. In R. J. Teather, C. Joslin, W. Stuerzlinger, P. Figueroa, Y. Hu, A. U. Batmaz, W. Lee, and F. Ortega, eds., *VRST*, pp. 29:1–29:11. ACM, 2020. 1

[93] M. Wilczkowiak, K. Jakubzak, J. Clemoes, C. Treptow, M. Porubanova, K. Read, D. McDuff, M. Kuznetsova, S. Rintel, and M. Gonzalez-Franco. Ecological Validity and the Evaluation of Avatar Facial Animation Noise. In *IEEE VRW*, pp. 72–79, 2024. doi: 10.1109/VRW62533.2024.00019 3

[94] P. Wisessing, K. Zibrek, D. Cunningham, J. Dingliana, and R. McDonnell. Enlighten Me: Importance of Brightness and Shadow for Character Emotion and Appeal. *TOG '20*. doi: 10.1145/3383195 5